PaPE 2025

Book of Abstracts

Phonetics and Phonology in a Multilingual World

Irie Institut de Recerca i Innovació Educativa UIB - GOIB

IDELab Investigation Development Education Language

GRESIB Grup de Recerca Sociolingüística de les Illes Balears

Universitat de les Illes Balears — iSLA Interdisciplinary Speech and Language Acquisition Lab

Universitat de les Illes Balears — Facultat d'Educació

Editors: Maria del Mar Vanrell Bosch, Gabriel Amengual Pastor, Eva Aguilar-Mediavilla

First Published: 2025

ISBN: 978-84-09-73926-4

Published by: Institut de Recerca i Innovació Educativa (IRIE)

Content

Invited speakers

Adaptive sound systems: segmental speech patterns in multilingual contexts

Mark Amengual
University of California, Santa Cruz

Multilingual individuals do not entirely suppress the languages they are not currently using; instead, all known language systems remain partially active, creating competition during speech perception and production. This simultaneous activation often results in phonetic deviations, where the pronunciation in the target language is subtly influenced by features from non-target languages. The current project explores this phenomenon by investigating the flexibility of sound systems in multilingual speakers, with a specific focus on differences between bilingual and trilingual contexts. By analyzing the acoustic properties of phonetic segments such as laterals and voiceless stops, this research examines how static and dynamic interlingual phonological influence shapes the acquisition and use of second (L2) and third (L3) languages. Data are presented from L2 and L3 learners as well as Heritage speakers, offering insights into how these groups manage phonetic interference and language interaction. Through this lens, this presentation sheds light on the complex phonological dynamics at play in multilingual speech. The talk concludes with both methodological and theoretical reflections, emphasizing the importance of considering interlingual influence in phonological research and proposing future directions for exploring how multilingual individuals navigate and adapt their sound systems across languages.

Dyslexia, rhythm, language and the developing brain

Usha Goswami
Centre for Neuroscience in Education, University of Cambridge

Recent insights from auditory neuroscience provide a new perspective on how the brain encodes speech. Using these recent insights, I will provide an overview of key factors underpinning individual differences in children's development of language and phonology, providing a context for exploring atypical reading development (dyslexia). Children with dyslexia are relatively insensitive to acoustic cues related to speech rhythm patterns. This lack of rhythmic sensitivity is related to the atypical neural encoding of rhythm patterns in speech by the brain. I will describe our recent data from infants as well as children, demonstrating developmental continuity in the key neural variables.

Variation in the phonology of Icelandic, at home and abroad

Nicole Dehé
University of Konstanz

In this talk I will present recent research on variation in the phonology of Icelandic, including varieties of Modern Icelandic as spoken in Iceland as well as North American Heritage Icelandic spoken in Manitoba, Canada. Both Modern Icelandic and Heritage Icelandic are in close contact with English.

Modern Icelandic (ModIce) has generally been argued to be a language with hardly any dialectal variation, in particular in comparison with other Germanic languages. Nevertheless, regional phonological variation can be found and has been studied since the 1940s, dividing Iceland into the North/North East, South/South East, and the Westfjords. North American Heritage Icelandic (NAmIce) is a heritage language still spoken in Manitoba, Canada, by descendants of Icelanders who emigrated to North America mostly from the North/North East of Iceland between the 1870s and 1914.

I will discuss findings from experiments on intonation, word stress, preaspiration and segmental variation and compare the phenomena in the light of language contact and cross-linguistic influence.

The intonation of Icelandic has been studied using read speech as well as map task dialogues. ModIce and American English have similar nuclear contours in declaratives but different contours in, for example, polar questions (L* H-H% in English, L*+H/L+H L-L% in ModIce). Results show that NAmIce speakers use tonal categories present in the tonal inventories of both ModIce and English, but they employ them in different ways, suggesting maintenance, transfer from English, and also innovative strategies. We are currently analyzing data targeting regional differences in Icelandic intonation. Results will also help interpret the NAmIce findings by relating the differences to the origin of the immigrants.

While English is a free word stress language, ModIce has fixed primary **word stress** on the initial syllable of a word, with very few exceptions. We study word stress in ModIce and NAmIce using a picture-naming task, identifying word stress positions in NAmIce that are deviant from ModIce, suggesting transfer effects from the majority language English.

Preaspiration is a salient phenomenon of ModIce. The fact that it is a late acquired phenomenon in first language acquisition as well as its near absence in English could make it vulnerable to cross-linguistic influence, while its contrastive function could have the opposite effect. We show that NAmIce speakers are aware of the distribution of preaspiration and realize it, but vary in its realization as compared to ModIce speakers. The results suggest that typologically marked structures, if phonemically relevant, may be phonologically retained in extreme language contact settings.

Finally, regarding segmental variation, specific variants have been associated with the regions of Iceland mentioned above. For ModIce, recent research indicates that some of the specific features (e.g., voiced pronunciation of sonorants before voiceless stops in the North as opposed to devoiced sonorants elsewhere) are fading out, while others (e.g., *hard speech*, i.e., post-aspiration of /p t k/ after long vowels, also a feature of the North) are maintained. In NAmIce, similar developments can be observed, despite the lack of contact between ModIce and NAmIce. At least in the case of *hard speech*, this cannot be due to transfer assuming that there is no such devoicing in English. Instead, regional features that immigrated to Manitoba, despite not being phonemically relevant, are retained.

All in all, the results will be relevant to the study of the phonological development of languages and their varieties in contact situations.

Oral Presentations

Oral session 1:Bilingualism 1

Place-dependent stop salience among Spanish-English bilinguals

Ernesto R. Gutiérrez Topete
Pomona College

While bilinguals maintain distinct phonetic systems for each language, research suggests that their languages interact during on-line processing, leading to short-term phonetic interference without affecting long-term phonological representations [2, 12]. This interference is often observed in code-switching (CS), the alternation between languages within a single utterance or conversation [6]. CS studies on voice onset time (VOT) reveal phonetic convergence as a common phenomenon where the VOT of one language approaches the norms of the other [3, 7, 13, 14]. While some scholars emphasize the crucial role of sociolinguistic factors in shaping CS [9], including social identity and characteristics, others advocate for a more inclusive approach that integrates social and cognitive factors, suggesting an interplay between language attitudes, social networks, and cognitive processing demands [10].

The present study analyzes the phonetic convergence of VOT in CS among Spanish- English bilinguals as mediated by social factors, focusing on the place-dependent distribution of VOT across languages (i.e., VOT increases for more posterior places of articulation). To do so, data were collected from 60 Spanish-English bilingual adults in four production tasks: word list reading, passage reading, puzzle completion, and a sociolinguistic interview. Only non-cognate words were analyzed to control for potential cognate effects on VOT [1, 8]. Participants also completed the Bilingual Language Profile (BLP) [4] to assess language dominance and obtain demographic and social information. Data were recorded in a sound booth and processed to obtain VOT measurements for voiceless stop-initial words following a switch site, yielding a total of 9,831 tokens. Initial analysis of the VOT data revealed that /t/ deviated statistically from the expected place-dependent linear increase in VOT in both English and Spanish (i.e., higher than expected in English and lower in Spanish; see fig. 1). This statistical deviation was quantified for each speaking using a new statistical measure called the Place Difference Index (PDI), developed for these data, which quantifies the degree to which the VOT deviates from the expected linear increase in VOT as place of articulation moves from /p/ to /t/ to /k/. The analysis revealed a bimodal distribution of PDI values based on language: English had an average PDI of -1.01, whereas Spanish had an average PDI of 1.02 (see fig. 2).

The statistical deviance of /t/ reported in this study is suggestive of bilinguals' efforts to maintain a distinct /t/ category for each language, emphasizing their proficiency in both. Alveolar /t/ may become salient due to its acoustic characteristics: it is louder and produced at higher frequencies [5, 11]. The analysis using the PDI revealed a bimodal distribution influenced by self-reported identification with a Spanish-speaking culture. Participants who identified more strongly with this community produced /t/ with lower VOT in both English and Spanish, further differentiating it from /p/ and /k/. This aligns with the notion that individuals may use specific features, like /t/ production, to express their linguistic identity. While limited comparative data exist, some studies suggest potential support for the observed /t/ deviation. The spectral distinction of /t/ (i.e., louder burst and higher frequencies, compared to /p/ and /k/) may lead to its phonetic salience, allowing it to play a role as a marker of bilingual proficiency and identity. These results emphasize the PDI algorithm's utility in analyzing phonetic variation, particularly in code-switching research.

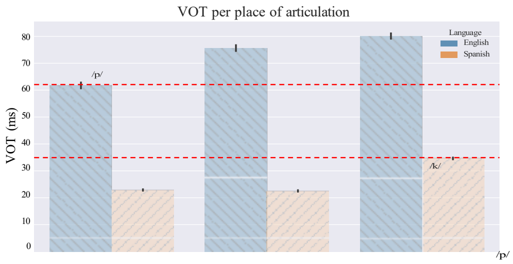


Figure 1: VOT Distributions

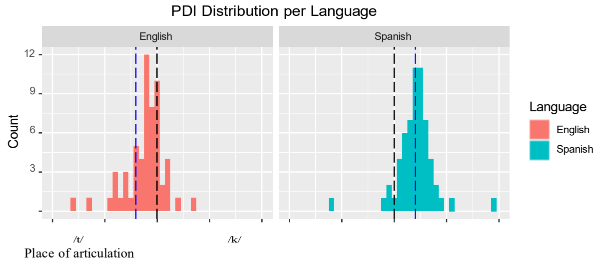


Figure 2: PDI Distributions

[1] Mark Amengual. “Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism”. In: *Bilingualism: Language and Cognition* 15.3 (2012), pp. 517–530.
[2] Mark Amengual. “The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech”. In: *The Journal of the Acoustical Society of America* 149.2 (2021), pp. 1271–1284.
[3] Colleen Balukas and Christian Koops. “Spanish-English bilingual voice onset time in spontaneous code-switching”. In: *International Journal of Bilingualism* 19.4 (2015), pp. 423–443.
[4] David Birdsong, Libby M Gertken, and Mark Amengual. “Bilingual language profile: An easy-to-use instrument to assess bilingualism”. In: *COERLL, University of Texas at Austin* (2012).
[5] Sheila E Blumstein and Kenneth N Stevens. “Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants”. In: *The Journal of the Acoustical Society of America* 66.4 (1979), pp. 1001–1017.
[6] Barbara E Bullock and Almeida Jacqueline Toribio. “Themes in the study of code-switching”. In: *The Cambridge Handbook of Linguistic Code-switching*. Ed. by Barbara E Bullock and Almeida Jacqueline Toribio. Cambridge University Press, 2009, pp. 1–18.
[7] Barbara E Bullock et al. “Language dominance and performance outcomes in bilingual pronunciation”. In: *Proceedings of the 8th generative approaches to second language acquisition conference*. Ed. by Mary Grantham O’Brien, Christine Shea, and John Archibald. Cascadilla Proceedings Project Somerville, MA. 2006, pp. 9–16.
[8] James Emil Flege and Murray J Munro. “The word unit in second language speech production and perception”. In: *Studies in Second Language Acquisition* 16.4 (1994), pp. 381–411.
[9] Penelope Gardner-Chloros. “Sociolinguistic factors in code-switching”. In: *The Cambridge Handbook of Linguistic Code-switching*. Ed. by Barbara E Bullock and Almeida Jacqueline Toribio. Cambridge University Press, 2009, pp. 97–113.
[10] Hong Liu. “A socio-cognitive approach to code-switching: from the perspective of a dynamic usage- based account of language”. In: *International Journal of Bilingual Education and Bilingualism* 23.10 (2020), pp. 1270–1299.
[11] Ralph N Ohde and Kenneth N Stevens. “Effect of burst amplitude on the perception of stop consonant place of articulation”. In: *The Journal of the Acoustical Society of America* 74.3 (1983), pp. 706–714.

[12] Daniel J Olson. "Code-switching and Language Mode Effects in the Phonetics and Phonology of Bilin- guals". In: *The Cambridge Handbook of Bilingual Phonetics and Phonology*. Ed. by Mark Amengual. Cambridge University Press, forthcoming.
[13] Page Piccinini and Amalia Arvaniti. "Voice onset time in Spanish-English spontaneous code-switching". In: *Journal of Phonetics* 52 (2015), pp. 121–137.
[14] Almeida Jacqueline Toribio et al. "Perseverative phonetic effects in bilingual code-switching". In: *The- oretical and experimental approaches to Romance linguistics*. Ed. by R Gess and E Rubin. Philadel- phia/Amsterdam: John Benjamins Publishing Company, 2005, pp. 291–306.

Voice quality, articulation rate, and pitch in different speech tasks produced by Spanish-German bilinguals

Jörn Krantz, Christiane Ulbrich
University of Cologne

Previous research has described differences in voice quality and prosody between languages, e.g. for f0 in Japanese/Dutch (van Bezooijen, 1995), in German/English (Mennen, Schaeffler & Docherty, 2012), in Finnish/Russian (Ullakonoja, 2007), in Mandarin/English (Keating & Kuo, 2012). Wagner and Braun (2003) found differences in mean f0, f0 modulation, HNR, jitter and shimmer in German, Italian and Polish male speakers. However, little is known about if and how an individual's voice changes when speaking another language. Preliminary results revealed greater inclination of the spectral slope for L2 compared to L1 (Järvinen & Laukkanen, 2015) in Finnish and English native speakers. Engelbert (2014) found that native speakers of Brazilian Portuguese (BP) produced their native language with lower f0 compared to their L2 English, whereas H1-H2* frequencies were higher in BP. The current study contributes to the line of research through comparison of vocal parameters related to voice quality and f0 as well as articulation rate in sequential bilinguals. Firstly, we intend to confirm results discussed above, suggesting that languages exhibit specific articulatory settings which can be verified by acoustic measurements. Secondly, we examine whether task type contributes to differences the native and target language of L2 speakers. To do so, we recorded 47 native Spanish learners of L2 German (31 female) and 11 German controls. Materials were taken from recorded *casual conversations* (CC) amongst the participants. Two other tasks were integrated in classroom activities and consisted of a *collaborative task* (CT) and an *individual presentation* (IP) in both native language Spanish and L2 German. Acoustic measurements included in the analysis were Harmonic-Noise-Ratio (HNR), Cepstral Peak Prominence-smoothed (CPPs), spectral tilt operationalized as H1*-H2* as well as f0-range. In addition, we calculated articulation rate (AR) as number of syllables divided by the duration of *inter-pause-stretch* (Trouvain et al., 2001), excluding sequences with background noise, hesitations and pauses. Both linear mixed effects models as well as linear regression models were implemented with the lmer function from the lme4 (Bates et al. 2015) package and the lm function (Chambers, 1992), from the base package in R (R Core Team, 2022). The ANOVA function was implemented using the lmerTest package (Kuznetsova et al., 2017). The emmeans function from the emmeans package (Searle et al.,1980) was used to calculate least square means and to test for significant differences. For the visualization of our data, we used ggplot2 (Wickham, 2016). Overall, our findings suggest differences in multiple dimensions of spoken language including prosodic (f0-range) and temporal (AR) characteristics as well as voice quality (HNR, CPPs, H1*-H2*) between different speech tasks in native and non-native German and native Spanish. More specifically, we found some acoustic parameters, namely, HNR, f0 and AR, to differentiate both speech tasks and languages. CPPs, however, only differentiates between the more spontaneous speech task CC compared to the more elicited speech tasks CT and IP supporting studies reporting task dependency of voice quality. Furthermore, H1*-H2* differs between languages in general, regardless of whether German was produced by native speakers or L2 learners, hence supporting studies reporting spectral changes when languages are switched (Harmegnies & Landercy, 1985, Bruyninckx, et al., 1994, Ng, Chen, and Chan, 2012). To conclude, this awareness of intra- and interspeaker variation (Ulbrich & Werth, 2021) in the production of different languages and speech tasks may therefore lead to a more successful assessment of spoken language in the multilingual classroom. In turn, this will allow for more expectation and acceptance of differences between language users. However, the extent to which the observed task-dependent and speaker-specific differences are perceptually relevant remains open to further investigation.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1--48.

Bruyninckx, M., Harmegnies, B., Llisterri, J., & Poch-Oiivé, D. (1994). Language-induced voice quality variability in bilinguals. *Journal of Phonetics*, 22(1), 19--31. doi:10.1016/S0095-4470(19)30265-7.

Chambers, J. M. (1992). Linear models. Chapter 4. In: J. M. Chambers and T. J. Hastie (Eds.), *Statistical Models in S.*, Wadsworth & Brooks/Cole, Pacific Grove.

Engelbert, A. (2017). Cross-Linguistic Effects on VoiceQuality: A Study on Brazilians' Production of Portuguese and English, Proc. of the Int. Symp. on the Acquisition of Second Language Speech. *Concordia Working Papers in Applied Linguistics*, 5, 157--170.

Harmegnies, B., & Landercy, A. (1985). Language features in the long term average spectrum," *Revue de Phonétique Appliquée,* 73, 69--80.

Järvinen, K., & Laukkanen, A.-M. (2015). Vocal Loading in Speaking a Foreign Language, *Folia Phoniatr Logop,* 67(1), 1--7. doi: 10.1159/000381183.

Keating, P., & Kuo, P. (2012). Comparison of speaking fundamental frequency in English and Mandarin, *J Acoust Soc Am*, 132, 1050--1060. doi: 10.1121/1.4730893.

Ng, M. L., Chen, Y., & Chan, E. Y. K. (2012). Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese- English bilingual speakers: a long-term average spectral analysis. *J. Voice*, 26(4), e171--6. doi: 10.1016/j.jvoice.2011.07.013.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: a comparison of English and German. *J Acoust Soc Am*, 131(3), 2249--60. doi: 10.1121/1.3681950.

R Core Team, (2022). R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Searle, S. R., Speed, F. M., & Milliken, G. A.(1980). Population Marginal Means in the Linear Model: An Alternative to Least Squares Means, *The American Statistician*, 34(4), 216--221.

Trouvain, J., Koreman, J., Erriquez, A., & Braun, B. (2001). Articulation Rate Measures and Their Relation to Phone Classification in Spontaneous and Read German Speech Isca ITR-workshop. In: j.-C. Junqua (Ed.). *Proceedings of the Workshop Adaptation Methods for Speech Recognition: Sophia-Antipolis:* ISCA, 155--158.

Ullakonoja, R. (2007). Comparison of pitch range in Finnish (L1) fluency and Russian (L2). In J. Trouvain, W. J. Barry (Eds.): The Proceedings of the 16th International Congress of Phonetic Sciences, 6-10 August 2007, Saarbrücken, Universität des Saarlandes, 2007, 1701--1704.

Ulbrich, C., & Werth, W. (2021). What is Intra-individual Variation in Language?. In: A. Werth, L. Bülow, S. E. Pfenninger & M. Schiegg (Eds.). Intra-Individual Variation in Language. Trends in Linguistics363), Mouton de Gruyter: Berlin & Boston, 9--43. doi: 10.1515/9783110743036.

van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. Language and Speech, 38(3), 253--265. doi:10.1177/002383099503800303.

Wagner, A., & Braun, A. (2003). Is voice quality language-dependent? Acoustic analyses based on speakers of three different languages. Language, 6(4):2.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Language dominance in bilinguals' prosody

Salvatore Gianninò[1], Alessandro Vietti[1], Mariapaola D'Imperio[2]
[1]Free University of Bozen-Bolzano, [2]Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

In bilingual speakers, elements of the L1 prosody are often found in the L2 speech [10]. In particular, the L1 can influence i) the phonetic realization of intonational categories, ii) the frequency of association between those categories and pragmatic meaning [10]. However, patterns of language use can modify language dominance over lifetime, which is here understood as "the relative strength of the two languages" [4:213] in the bilingual mind. Notably, a change in language dominance is connected to a neutralized or reversed influence between the L1 and the L2 phonological systems [2]. Also, the two languages in the bilingual mind are claimed to interact differently according to the rate of code-mixing in daily life [9]. In particular, a low rate of code-mixing is expected to correlate with a reduced likelihood of transfers from the competing language in monolingual speech [6]. In South Tyrol (ST), speakers are predominantly German-Italian bilinguals [13]. In previous studies, a L+H* pitch accent has been identified in both ST Italian and German, mostly in Corrective Focus (CF) statements [4]. In ST Italian, this accent is produced with a medial alignment of the F0 peak in the tonic vowel [4], while it shows a later alignment in German [7], particularly in Southern varieties [1, 11]. Additionally, Narrow Focus (NF) statements in ST and Standard German are typically associated with rising L+H* nuclear pitch accents [1, 8], whereas in ST Italian a falling H+L* nuclear pitch accent is preferred in the same pragmatic structure [4].

In this study we ask whether, in ST Italian, language dominance and code-mixing frequency in daily life are significant predictors of both the F0 peak alignment and the frequency of L+H* pitch accents in NF condition. 2240 statements were elicited from 28 ST German-Italian bilinguals in a laboratory setting. 8 stimuli were used, with the structure "[They] put the NP1 on the NP2" (e.g., *Ha messo la collana sul denaro* 'He put the necklace on the money'). These were produced during an interactive videogame, aimed at manipulating the information structure. In this way, both NP1 and NP2 could alternatively bear the nuclear accent, being under the scope of Narrow or Corrective Focus. A questionnaire was also administered to assess individual variables. L1 was a categorical variable ("Italian", "German" and "Ita&Ger" for simultaneous bilinguals), while code-mixing rate was understood as the percentage of code-mixing in one typical week. Normalized language dominance was a continuous score derived from the Bilingual Language Profile [3], with Italian absolute dominance on the negative and German absolute dominance on the positive pole. A linear and a logistic mixed effect models were built to predict respectively the F0 peak alignment (as a proportion of the accented vowel [12]) and the probabilities of finding a nuclear L+H* in NF and CF statements. These models reveal that a shift towards the German-dominant pole associates with later alignments of the F0 peak and with a higher probability of L+H* in NF statements (respectively, Figure 1 and 2). Interestingly, a) this latter increase is found on a smaller scale even in CF and b) code-mixing rate shows no significant effect in either model.

Our findings provide evidence of the influence of language dominance on both the substance-to-form (i) and the form-to-meaning (ii) mappings of the bilingual intonational system. This stands in support of hybrid exemplar models, which claim that linguistic features at different levels of abstraction receive an inhibition/activation cost and that this cost is modulated by patterns of language use [5]. In this scenario, the effect in (a) might stem from a similar, yet smaller, cross-linguistic difference in the frequency of L+H* in CF relative to NF statements [4, 8]. Finally, while insufficient statistical power might hide behind (b), further possible explanations of this outcome will be considered.

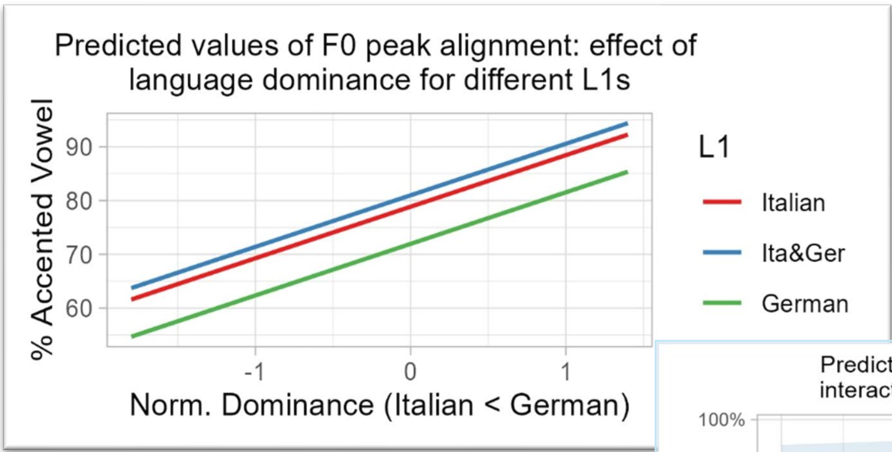


Figure 1

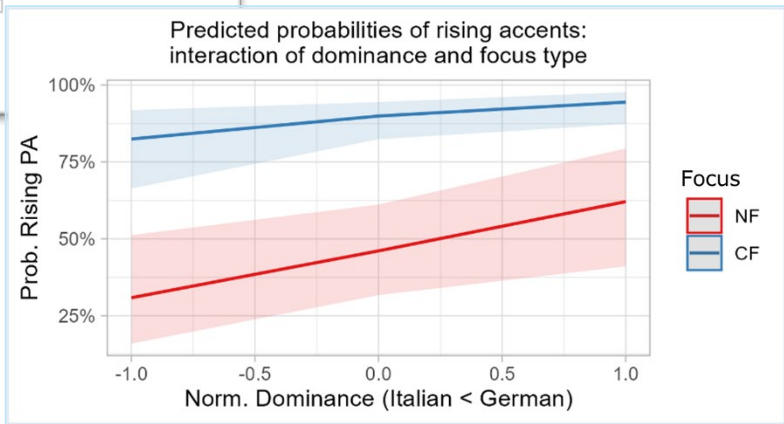


Figure 2

[1] Barker, G. 2005. *Intonation patterns in Tyrolean German an autosegmental-metrical analysis*. Lang.

[2] Flege, J.E. 2007. Language contact in bilingualism: phonetic system interactions. *Laboratory Phonology* (Berlin, 2007), 353–382.

[3] Gertken, L.M. et al. 2014. Assessing Language Dominance with the Bilingual Language Profile. *Measuring L2 Proficiency*. P. Leclercq et al., eds. Multilingual Matters. 208–225.

[4] Gianninò, S. et al. to appear. Phonological and phonetic strategies to focus marking in South Tyrol Italian. Proceedings of the XXI National Conference of the Italian Association for Speech Sciences (AISV) - 2025, February 4-7, Urbino (Italy) (to appear).

[5] Goldrick, M. and Cole, J. 2023. Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics*. 99, (Jul. 2023), 101254. DOI: https://doi.org/10.1016/j.wocn.2023.101254.

[6] Green, D.W. and Abutalebi, J. 2013. Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*. 25, 5 (Aug. 2013), 515–530. DOI: https://doi.org/10.1080/20445911.2013.796377.

[7] Grice, M. et al. 2005. German Intonation in Autosegmental-Metrical Phonology. *Prosodic Typology*. S.-A. Jun, ed. Oxford University Press. 55–83.

[8] Grice, M. et al. 2017. Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*. 64, (Sep. 2017), 90–107. DOI: https://doi.org/10.1016/j.wocn.2017.03.003.

[9] Grosjean, F. 2001. The Bilingual's Language Modes. *One mind, two languages: Bilingual language processing*. Blackwell Publishing. 1–22.

[10] Mennen, I. 2015. Beyond Segments: Towards a L2 Intonation Learning Theory. *Prosody and Language in Contact*. E. Delais-Roussarie et al., eds. Springer Berlin Heidelberg.

[11] Mücke, D. et al. 2006. Articulatory and acoustic correlates of prenuclear and nuclear accents. *Proc. Speech Prosody 2006* (2006).
[12] Silverman, K.E.A. and Pierrehumbert, J. 1990. The timing of prenuclear high accents in English. *Between the Grammar and Physics of Speech*. J. Kingston and M.E. Beckman, eds. 72–106.
[13] Vietti, A. 2017. Italian in Bozen/Bolzano: the formation of a 'new dialect.' *Towards a New Standard*. M. Cerruti et al., eds. De Gruyter. 176–212.

Oral session 2: Bilingualism 2

Cross-language variation of Hong Kong Cantonese n-l merger for early and late trilingual speakers

Grace Wenling Cao[1], Karen Ka Ki So[2], Peggy Mok[2]
[1]*University College Dublin,* [2]*The Chinese University of Hong Kong*

The Cantonese n-l merger refers to the ongoing sound change progress where the word-initial /n/ is realised as [l] in production. The SLM-r (Flege and Bohn 2021) suggests that bilingual speakers' L1 and L2 subsystems exist in a common phonetic space. It would be interesting to explore whether Cantonese-English-Mandarin trilingual speakers who have merged their [n]-[l] in L1 Cantonese develop new [n]-[l] categories formation for L2/L3 or share [n]-[l] categories between L1, L2 and L3. Previous studies by So et al (2021, 2024) found that the n-l merger occurred in Cantonese but not in English for Canadian Cantonese-English bilinguals. The current study expands the scope to Cantonese-English-Mandarin trilinguals in Hong Kong, and asked two research questions: (1) Do trilingual speakers use separate or shared n-l categories in their L1-Cantonese, L2-English and L3-Mandarin? (2) Do early and late trilinguals have different patterns of n-l merger across languages?

Forty Cantonese-English-Mandarin trilingual speakers from Hong Kong (20 young and 20 older participants; mean age: 21, 67 years old) completed a wordlist and semi-spontaneous interview in three separate sessions of Cantonese (CAN), Mandarin (MAN) and English (ENG). Young participants, classified as early trilinguals, learnt ENG and MAN at the age of 3 and 4.5 years old; while older participants, classified as late trilinguals, learnt them at 7 and 26 years old respectively. Due to different language policies, the young group acquired ENG and MAN systematically in school, whereas the older group learnt them through work and other means, resulting in differences in the amount of exposure and the mode of acquisition.

To compare across the three languages, target words were combinations of word-initial n/l with four different rimes (i.e. aan/an, ai, ei, i/in) in CAN, MAN and ENG respectively. All the target words in the wordlists and interviews were manually selected by three research assistants who are native speakers of Cantonese. Both auditory and acoustic analyses were conducted. The RAs listened to each token of the target words and judged whether the word was pronounced with a [n] or [l]. For acoustic analysis, three parameters were chosen based on past literature: F2-F1 (Johnson, 2011), mid-frequency spectral tilt measure (H4-2KHz, So et al. 2021, 2024) and duration of the consonant (Cheng & Jongman, 2019).

For the early trilinguals, the auditory results suggested that the n-l merger ([n]→ [l]) was mainly found in CAN, with a higher percentage in interviews (82%) than in the wordlist (48%). Only a low percentage of the n-l merger was found in ENG (wordlist: 6%; interview: 5%) and MAN (wordlist: 4%, interview: 2%). Figure 1 shows that males had a higher percentage of the n-l merger in CAN than females in both styles, contrary to Cheng et al. (2022) who found no gender effect. Acoustically (see Table 1), CAN [n]-[l] were similar in F2-F1, H4-2KHz and consonant duration, echoing the auditory results where the n-l merger was mainly found in CAN. For ENG and MAN, F2-F1 of [l] was *lower* than that of [n], which is different from Johnson's (2011) but similar to Soo et al.'s (2021) finding on English. Also, male participants had a lower H4-2KHz in [n] than [l], which was not found among females. For the late trilinguals, only the auditory results are ready. Figure 1 shows that the late trilingual speakers have a higher percentage of the n-l merger (i.e. ([n]→ [l]) in MAN and ENG, compared to the young trilinguals. The older male participants showed a strong n-l merger in their MAN interviews, with 74% of the n-tokens pronounced as [l]. The differences in gender and styles were less salient for the late trilingual group. The present study shows that the late trilingual speakers have various degrees of n-l mergers in their CAN, MAN and ENG, whereas the early trilinguals only showed mergers in their CAN. More analyses will be conducted to explore whether these differences were caused by language proficiency and/or a compound factor of age and exposure.

Figure 1. *Percentage of [l]-realisation across languages. The higher the number of [l]-realisation on the column for [n] (right), the stronger the n-l merger represents.*

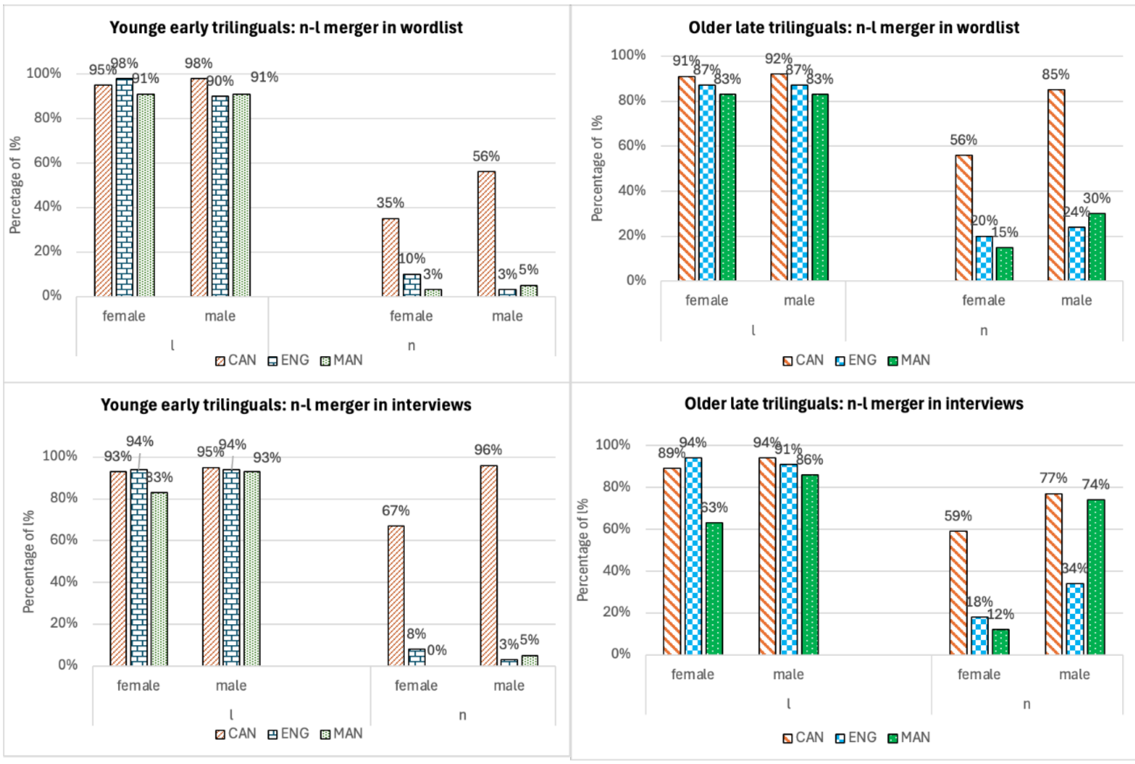


Table 1. *Summary of the p-values for comparisons between [n] and [l] on three chosen features for young early trilingual speakers. 'ns' represents no significance was found.*

			CAN	ENG	MAN
Wordlist	**Female**	**F2-F1**	ns	[n] > [l]; p <.001	ns
		H4-2K	ns	[n] > [l]; p =.004	ns
		Duration	ns	ns	[n] > [l]; p <.001
	Male	**F2-F1**	ns	[n] > [l]; p <.001	[n] > [l]; p <.001
		H4-2K	ns	[n] < [l]; p =.005	[n] < [l]; p <.001
		Duration	ns	ns	ns
Interview	**Female**	**F2-F1**	ns	[n] > [l]; p <.001	[n] > [l]; p <.001
		H4-2K	ns	ns	ns
		Duration	ns	ns	ns
	Male	**F2-F1**	ns	[n] > [l]; p =.03	[n] > [l]; p = .04
		H4-2K	ns	[n] < [l]; p <.001	ns
		Duration	ns	[n] > [l]; p <.001	ns

Cheng, L. S. P., Babel, M., & Yao, Y. (2022). Production and perception across three Hong Kong Cantonese consonant mergers: Community- and individual-level perspectives. *Laboratory Phonology*, 13 (1): 14, 1-54.

Cheng, R, Q., Jongman, A. (2019). Acoustic analysis of nasal and lateral consonants: The merger in Eastern Min. *Journal of the Acoustical Society of America*, 145: 1828.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). chapter, Cambridge: Cambridge University Press.

Johnson, K. (2011). *Acoustic and Auditory Phonetics*. John Wiley & Sons, Incoporated.

Soo, R., Johnson, K. & Babel, M. (2021). Sound change in spontaneous bilingual speech: A corpus study on the Cantonese n-l merger in Cantonese-English bilinguals. *Interspeech 2021*. Brno, Czechia.

Soo, R., Babel, M., & Johnson, K. (OnlineFirst). Cross-linguistic phonetic variation in bilingual speech: Cantonese /n/ > [l] merger in early Cantonese-English bilinguals. *Language and Speech*, 1-29.

"How do you take your /t/?" Investigating /t/-variation in Irish-English bilinguals

Kate Tallon
University of Oxford

The 'slit-t' is a characteristic feature of Irish English (IrE, i.e. English spoken in Ireland). It denotes an apico-alveolar fricative [ṭ], occuring in syllable coda position. Slit-t is geographically widespread in Ireland – particularly in southern IrE – and is considered typical of supraregional IrE [1]. IrE speakers with L1 Irish, however, reportedly do not have this lenited variant, instead using a plosive [t] [2]. A further exception to the widespread patterning of slit-t is found in younger urban speakers of IrE, who have been found to use a wider range of /t/ variants, including glottal and tap variants [3, 4]. Given these findings, it was of interest to (i) investigate whether empirical data supports previous reports that L1 Irish speakers of English show a plosive [t] where a slit-t/fricative [ṭ] is usually found, and (ii) investigate whether younger Irish-English bilinguals display a wider range of /t/ variants than older bilinguals. To contextualise these findings, the speech data of the bilinguals was compared with that of monolinguals from the same area, i.e. the Connemara Gaeltacht, an Irish-speaking region on the west coast of Ireland.

12 bilingual (L1 Irish L2 English) speakers (6 M, 6 F; age range 20-79; median age 38; SD 19.4) and 12 monolingual English speakers (6 M, 6 F, age range 27-66; median age 55; SD 16) were recorded in both read and spontaneous speech tasks. A total of 688 tokens of coda /t/ were acoustically and statistically analysed and the resultant variants are displayed in Table 1. The data is visualised in Figure 1, with speakers further grouped by gender and age. To statistically investigate the variation between bilinguals and monolinguals, a generalised logistic mixed effects model (GLMM) was run in R, using glmer() from lme4 [5,6]. Following acoustic analysis of spectrograms in Praat, the tokens of /t/ were binarily coded as lenited or plosive and were modelled against the factors of linguistic background (monolingual vs. bilingual) and age, with speaker and word as random intercepts. The model results are displayed in Table 2.

The plosive [t] was the most frequent variant for older bilinguals. The older bilingual women in particular, produced [t] in 90% of cases. This contrasts findings for the older monolinguals, who favoured the slit-t/fricative variant [ṭ] overall. In contrast to their older counterparts the younger bilingual and monolingual speakers showed a relatively high rate of tap and glottal variants. Regarding the statistical analysis, the GLMM found that linguistic background and age were found to have significant effects on the likelihood of plosive [t]. The interaction effect between these two factors and /t/- lenition was not significant. The model estimated that a /t/ produced by young bilingual speaker has a 49% likelihood (log odds -0.03) of being a plosive [t], whereas /t/ produced by a monolingual (both age groups) has a 1% probability of being a plosive (log odds -4.43). Regarding the factor of age, the probability of observing a plosive [t] is 97% higher for older speakers compared to younger speakers. This difference is statistically significant ($p = 0.02$). Overall, the results of this GLMM show significant effects of linguistic background and age on the occurrence of plosive [t].

These findings reveal previously unreported /t/-variation within Irish-English bilinguals. They counter previous reports that native Irish speakers do not show lenition of /t/ in their English. Additionally, they indicate ongoing language change in the Connemara Gaeltacht area regarding /t/-variation, with younger speakers showing higher rates of tap and glottal variants. Additional studies of this kind are essential and urgent, given the status of Irish as a critically endangered language and the resulting diminishing opportunities to study language contact effects within Irish-English bilingual speakers.

Table 1. */t/ variants for bilinguals and monolinguals.*

/t/ variants

Ling. background	Variant					
	[t]	**[t̞]**	**[ɾ] ***	**[ʔ]**	**[t̪]**	**Total**
Bilingual	170	97	34	25	3	329
Monolingual	59	201	57	42	0	359

Figure 1. */t/ variants for bilinguals and monolinguals, according to age and gender.*

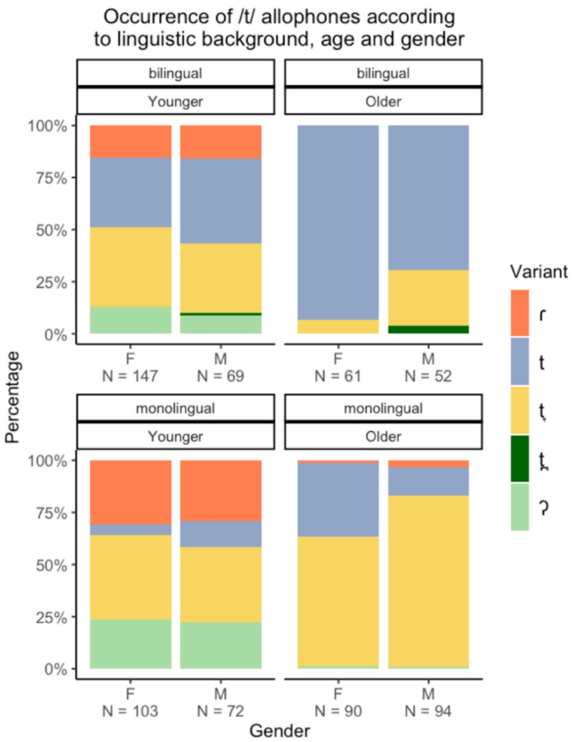


Table 2. *Output of GLMM investigating age and linguistic background on /t/ variation.*

```
glmer(lenition ~ ling*age_group + (1 | word) + (1|speaker))
```

Fixed effects	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-0.02871	0.96796	-0.030	0.97634
lingMonolingual	-4.39666	1.43905	-3.055	0.00225**
age_groupOlder	3.61250	1.53977	2.346	0.01897*
lingMonolingual:age_groupOlder	-1.12744	2.09555	-0.538	0.59056

[1] Hickey, R. 2012. Standard Irish English. In R. Hickey (Ed.), *Standards of English: Codified Varieties around the World* (pp. 96–116). chapter, Cambridge: Cambridge University Press.
[2] Hickey, R. 1984. Coronal segments in Irish English. *Journal of Linguistics*, *20*(2), 233-250.
[3] Schulte, M. (2020). Positive Evaluative Stance and /t/ Frication a Sociophonetic Analysis of /t/ Realisations in Dublin English. de Gruyter Mouton Berlin.

[4] Skarnitzl, R., & Rálišová, D. (2022). Phonetic variation of Irish English/t/in the syllabic coda. *Journal of the International Phonetic Association*, 1-20.
[5] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1 - 48.
[6] R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. <https://www.R-project.org/>.

Mandarin speakers' development of L2 Spanish lexical stress perception and production in a one-year study abroad program

Peng Li[1], Clara Martin[1,2], Xiaotong Xi[3]
[1] *Basque Center on Cognition, Brain and Language (BCBL), Donostia,* [2] *IKERBASQUE, Donostia,* [3] *Shandong University of Finance and Economics, Jinan*

Despite extensive research on novel speech sound learning, less is known about how study abroad (SA) influences the development of second language (L2) suprasegmental features [1]. Lexical stress is one such area, where tonal language speakers, such as Mandarin natives (Mandarin L1ers), overweigh pitch to perceive and produce L2 lexical stress contrasts due to prosody-level L1-L2 transfer, which differs from Spanish natives (duration > pitch) [2], [3]. Despite the L1 background, individual factors like domain-general processing abilities in acoustic cues (e.g., pitch, duration) [4] and L2 use patterns [5] also play critical roles in L2 speech acquisition. The present longitudinal study systematically examines (a) how SA impacts Mandarin L1ers' perception and production of Spanish lexical stress at the group level and (b) how auditory processing abilities and L2 use mediate SA effects at the individual level.

Thirty female Mandarin L1ers, enrolled in one-year master's programs across various disciplines in Spain, participated. They had studied Spanish as their undergraduate major in China with no prior SA experience. The participants were tested upon arrival (T1) and before leaving Spain (T2). At both tests, they were assessed for perception and production of Spanish lexical stress. The perception task involved identifying the first- or third-person conjugation of the verb, which contrasted only in lexical stress (e.g., *paso* 'I pass' vs. *pasó* 'she/he passed'). To create trochaic or iambic lexical stress patterns, the pitch and duration ratios of /a/ and /o/ were manipulated (7 steps each). Each stimulus was presented four times. The production task required participants to read short discourses embedding the five Spanish vowels (/a, e, i, o, u/) under 2 stress (stressed vs. unstressed) × 2 prominence (prominent vs. non-prominent) conditions. For the production task, we acoustically measured the duration and the mean pitch of the target vowels. For both perception and production data, we calculated the extent to which each participant relied on pitch relative to duration to make lexical contrasts (i.e., pitch cue weighting). In addition, 19 female native Spanish speakers completed the perception and production tests once to provide baseline data. As measures of individual differences, we tested the learners' domain-general auditory processing abilities [4] in pitch and duration at T1 and surveyed the participants' Spanish use history during SA at T2.

At the group level, SA led to clearer lexical stress categorization in perception; however, pitch remained the primary cue, which differed from Spanish natives (duration > pitch). (Fig. 1). In production (Fig. 2), learners lowered the pitch of stressed syllables after SA, but their mean pitch in stressed syllables remained higher than those of Spanish natives, and SA had no significant effect on duration differences between stressed and unstressed syllables. Interestingly, there was a positive correlation between pitch cue weighting in perception and production at T1, but this link disappeared at T2 (Fig. 3). At the individual level, auditory processing abilities significantly influenced perceptual cue-weighting changes: learners with poorer auditory processing abilities of pitch reduced their reliance on pitch to perceive lexical stress after SA. Additionally, frequent Spanish use during SA helped learners maintain the positive perception-production link at T2, whereas less frequent use weakened this association.

In conclusion, the study reveals that SA has unbalanced effects on L2 suprasegmental development. The results underscore the role of aptitude (auditory processing abilities) and experiential factors (L2 use) in mediating SA effects on lexical stress learning. These findings highlight the complexity of L2 lexical stress acquisition and suggest the need for tailored approaches that consider individual aptitudes and experiential factors in L2 instruction.

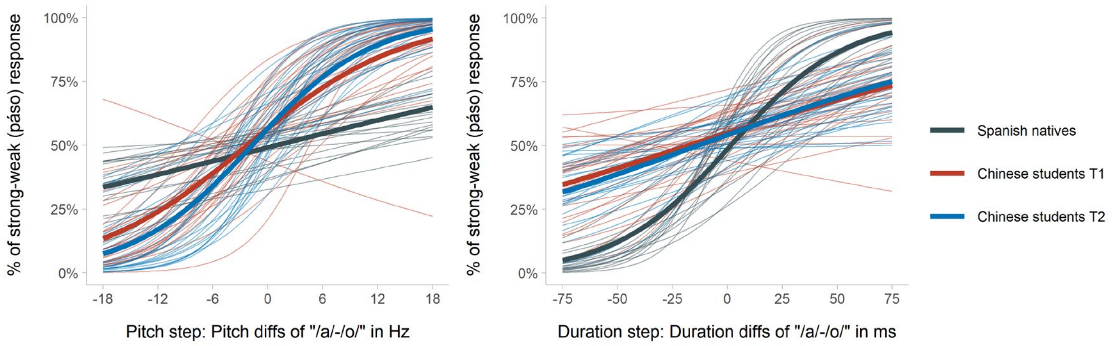


Figure 1. *Perception: Percentage of "páso" response as a function of pitch step (left panel) and duration step (right panel) split by tests (T1 vs. T2), with Spanish natives as a reference. Large lines show group means, and light lines indicate individual means.*

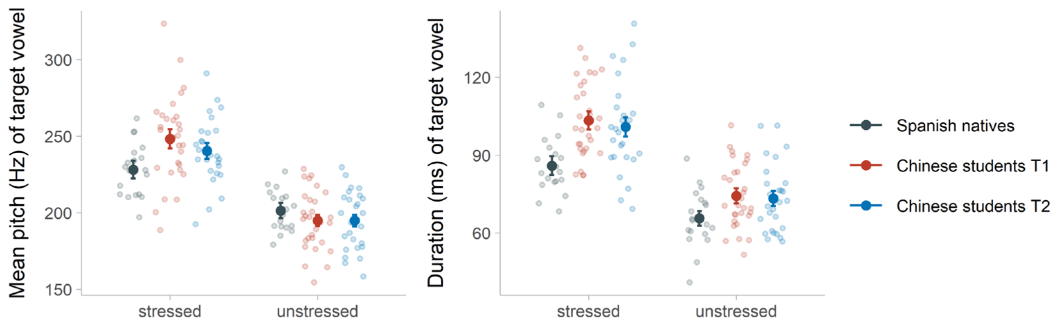


Figure 2. *Production: Learners' mean pitch (left panel) and duration (right panel) of target vowels across tests (T1 vs. T2), with Spanish natives as a reference. Large dots show group means, error bars represent 95% CI, and light dots indicate individual means.*

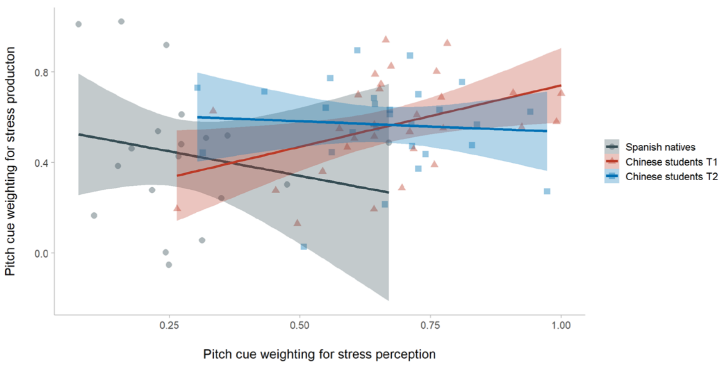


Figure 3. *Perception-production link as measured by pitch cue weighting split by test sessions (T1 vs. T2), with Spanish natives as a reference. Points plot individual data.*

[1] M. Solon and A. Y. Long, "Acquisition of Phonetics and Phonology Abroad: What We Know and How," in *The Routledge Handbook of Study Abroad Research and Practice*, C. Sanz and A. Morales-Font, Eds., Routledge, 2018, pp. 71–85.
[2] P. Li and X. Xi, "Spanish lexical stress produced by proficient Mandarin learners of Spanish," in *Proc.ISAPh4*, 2022, pp. 41–45.

[3] P. Li and X. Xi, “The perception of Spanish lexical stress by proficient Mandarin learners of Spanish,” in *Proc. Speech Prosody*, 2024, pp. 354–358.
[4] K. Saito, “How does having a good ear promote successful second language speech acquisition in adulthood? Introducing Auditory Precision Hypothesis-L2,” *Lang. Teach.*, pp. 1–17, 2023.
[5] H. Sun, K. Saito, and J.-M. Dewaele, “Cognitive and Sociopsychological Individual Differences, Experience, and Naturalistic Second Language Speech Learning: A Longitudinal Study,” *Lang. Learn.*, vol. 74, no. 1, pp. 5–40, 2024.

Phonetic reduction in L2 acquisition: Perception of English reductions by French learners

Marine Mouquet
Aix-Marseille Univ, Laboratoire Parole et Langage, CNRS UMR 7309

One of the key challenges for second language learners is the significant variability present in native speech. Studies indicate that native speakers easily understand and use phonetic reductions (e.g., *police* /pəˈliːs/ produced [pˈliːs]), often remaining unconscious of such phenomena. Although phonetic reductions have been extensively studied in native speech, they remain underexplored in second language acquisition. However, recent research ([1], [2]) suggests that L2 learners may encounter difficulties perceiving these reductions. The present study seeks to expand this field of research by examining L2 learners' challenges in perceiving phonetic reductions.

This research may be particularly relevant for learners with L1s and L2s having different rhythmic typologies such as stress-timed vs syllable-timed languages ([3], [4]), as it is the case with French learners of English. Drawing on the Markedness Differential Hypothesis ([5]), this work predicts that phonetic reductions, resulting from the deletion of unstressed syllables, may be challenging for speakers of syllable-timed languages.

The aims of this study are twofold: (a) to assess whether French learners of English can discriminate between full and reduced forms and (b) to examine whether these forms are stored in the mental lexicon or whether reduced forms are derived from the full forms. We predict that proficiency level and exposure may significantly impact perception. To investigate this, we conducted an ABX discrimination task in which L1 French learners of L2 English with different proficiency levels were presented with three acoustic stimuli spoken by three different speakers: A (e.g., reduced form), B (e.g., full form), and X (e.g., reduced form). Listeners were asked to determine whether X was more similar to A or B. A prior study ([6]) involving 48 British listeners revealed that perceived reduction frequency varied depending on stress position, sonority scale and syllable count. These criteria were used to ensure a balanced selection of stimuli. The results of the experimental group and of a control group of British native speakers are currently being analysed and will be presented.

This research will provide further evidence on whether exposure and word-specific characteristics influence learners' perception of reductions and could ultimately contribute to a greater understanding of phonetic processing for L2 phonological models and L2 teaching practice.

[1] Mulder, K., Brekelmans, G., & Ernestus, M. T. C. (2015). The processing of schwa reduced cognates and noncognates in non-native listeners of English.

[2] Brand, S., & Ernestus, M. (2018). Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: Evidence from native listeners and learners of French. *Quarterly Journal of Experimental Psychology*, *71*(5), 1240-1259.

[3] Pike, K. (1945). The Intonation of American English. Ann Arbor: University of Michigan Press.

[4] Abercrombie, D. (1967). Elements of general phonetics (Vol. 203). Edinburgh: Edinburgh University Press.

[5] Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language learning*, *27*(2), 315-330.

[6] Mouquet, M. (2024). Decoding the interplay between usage and acceptability of phonological reductions: A perceptive study on syncopes in British English listeners. Poster presentation at *ALOES24: Prominence in Spoken English*, 28–29 March 2024,

Université Sorbonne Paris Nord, France. https://aloesfrance.wordpress.com/actualites/programme-du-21e-colloque-davril-sur-langlais-oral-de-villetaneuse/

Oral session 3: Language acquisition and literacy

Articulating while listening? The perception-production link in pre-babbling infants

Sonia Frota[1], Catia Severino[1], Jovana Pejovic[1], Nuno Paulino[1], Marina Vigário[1], Aude Noiray[2]
[1]University of Lisbon, [2]Université Grenoble-Alpes

Several approaches to infant speech perception have proposed that the auditory analysis of speech is coupled with articulatory and motor information involved in speech production ([1]-[3]). Results from three related domains have provided converging evidence to this proposal. Brain imaging studies showed that speech perception activates infants' motor brain areas early in the 1st year of life, in response to all speech sounds (including non-native sounds, [1]). Studies on multimodal audio-visual speech perception showed that, before 6 months of age, infants are sensitive to auditory-visual vowel and consonant matches/mismatches ([2, 3]). Research on sensorimotor influences on speech perception demonstrated that experimentally induced inhibitions to tongue movement selectively affected (non)native phoneme discrimination between 3 and 6 months of age ([3]]. Recently, sensorimotor constraints have also been shown to affect prosody perception in 6-month-olds ([4]). Thus, the link between speech perception and production emerges very early in infancy, and crucially before infants have had some experience with speech production. This raises the fundamental question of the mechanisms behind the development of this link. Current views suggest that developing articulatory-auditory mappings are rooted in spontaneous non-speech movements in utero and in the first months of life ([1, 3]). However, a critical piece is missing, as research has not yet addressed whether pre-babbling infants produce linguistically-driven articulatory movements when perceiving speech. To address this key question, we employed ultrasound imaging, a method used to capture tongue movement in young children ([5)].

We recorded the tongue shapes of ten 6-month-old infants (data collection/analysis ongoing) while listening to VCV sequences where only C varies: [apa], [ata], [aka]. A block design with 4 repetitions of each token x 2 was implemented (Fig.1). Infants sat in a baby lounge chair, fixed and slightly reclined, while a researcher gently and firmly held the probe below the infant's chin. The audio stimuli were delivered through loudspeakers positioned on either side of the chair. The Micro Speech Research ultrasound system was used, with a microconvex short 10mm radius probe with large field of view suitable for children. Tongue image analysis was performed with AAA software ([6]). The Curve Position (CP) measure ([7]) was used to determine differences in tongue shape (lower values for back, higher for front) across token types (Fig.2). Specifically, we extracted CP in 16 equally spaced intervals within the 1s interval after the presentation of an VCV token. If tongue shape varies as a function of the token heard, we expect some kind of tongue to sound coordination with lower CP values triggered by listening to [k] as opposed to higher values for [t]. By contrast, an overall arousal response to speech would be more unconstrained. In a linear mixed-models analysis, token type ([apa], [ata], [aka]) and token number (1,2,3,4) were set as fixed effects (with interaction term), and participant as random effect. There was an effect of token type [$F(2,3355)=9,02$, $p <.001$], with [aka] showing the lowest CP (Fig.3). Pairwise contrasts showed that [aka] differed from [ata] [$t=-4,23$, $p < .001$] and [apa] [$t=-2,44$, $p=.029$]. There was also an effect of token number [$F(3,3355)=6,61$, $p <.001$], and an interaction token type*token number [$F(6,3355)=2,94$, $p=.007$]. The differences in CP triggered by token type only start to emerge after token 2 is heard, suggesting that at least two repetitions are needed for the articulating while listening effect to obtain. The results not only suggest that pre-babbling infants produce articulatory movements by moving the tongue more to the front or back in the direction of the consonant they heard, but also contribute to establish the temporal unfolding of the effect.

Our findings are the first to suggest that infants engage in speech driven oral-motor responses modulated by the specific sounds perceived, before producing speech, advancing our understanding of the mechanisms for the early emerging perception-production link.

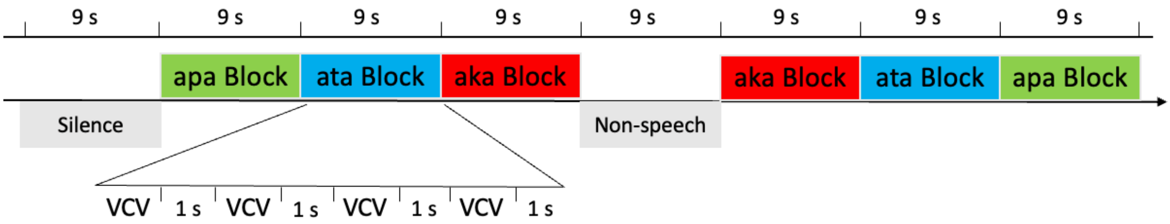


Figure 1. *Experimental design.*

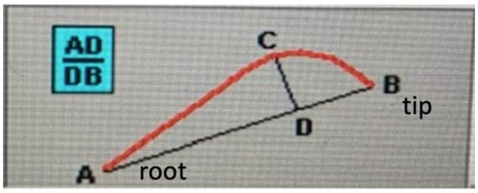


Figure 2. *Curve position measure in AAA.*

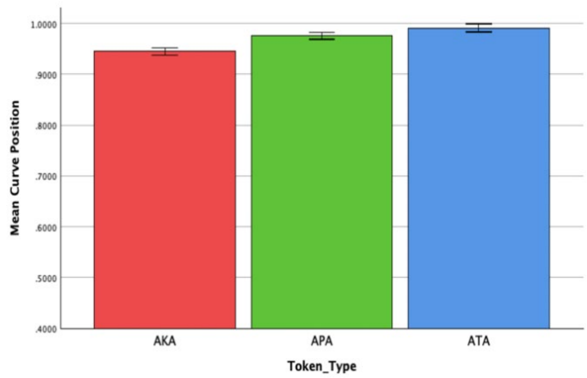


Figure 3. *Mean Curve Position by token type. Error bars indicate +/- 1 standard error of the mean*

[1] Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J. F. L., & Imada, T. 2014. Infants' brain responses to speech suggest Analysis by Synthesis. *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 31, 11238–11245. doi: 10.1073/pnas.1410963111.

[2] Best, C. T., Goldstein, L. M., Nam, H., & Tyler, M. D. 2016. Articulating What Infants Attune to in Native Speech. *Ecological Psychology* 28:4, 216-261.

[3] Choi, D., Yeung, H. H., & Werker, J. F. 2023. Sensorimotor foundations of speech perception in infancy. *Trends Cogn. Sci.*, vol. 27, no. 8, 773–784. doi: 10.1016/j.tics.2023.05.007.

[4] Frota, S., Severino, C., Pejovic, J., & Vigário, M. 2024. Sensorimotor influences on infant speech perception also target prosody. *Proceedings of Speech Prosody 2024* (Lisbon, Portugal), 1140-1144. doi: 10.21437/SpeechProsody.2024-230

[5] Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. 2018. How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research* 61:6, 1355-1368.

[6] Articulate Instruments Ltd. 2022. Articulate Assistant Advanced (Version 220.5.0) [Computer software]. Edinburgh: Articulate Instruments Ltd. Retrieved May 19, 2023 from http://www.articulateinstruments.com/downloads/

[7] Ménard, L., Aubin, J., Thibeault, M., & Richard, G. 2012. Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model. *Folia Phoniatrica et Logopaedica* 64:2, 64-72.

Phonological components of literacy in a shallow orthography. A longitudinal study on Italian mono- and bilingual children

Emanuele Casani
Universitat Pompeu Fabra (UPF), Barcelona

Phonological and meta-phonological skills are prerequisites for reading accuracy, yet the specific roles of different subcomponents of (meta)phonology remain under-researched. Likewise, Rapid Automatized Naming (RAN) is a longitudinal predictor of reading speed; however, the reasons remain unclear as RAN has been classified as either a subcomponent of phonological processing [1] or an executive function that may indicate reading automaticity [2]. Additionally, most research on these topics has focused on English, overlooking shallow orthographies, where phonological processes play a more prominent role in early reading. The high presence of bilingual children in most EU schools adds further complexity to this intricate framework. In this regard, the ongoing debate about a potential bilingual advantage in metalinguistic skills and executive functions raises questions about whether bilingual children may benefit from enhanced meta-phonological and RAN skills.

The present study examined these questions by analyzing the longitudinal development of (meta)phonological skills and their impact on literacy development in a shallow orthography during the early stages of literacy acquisition. The objectives were multifaceted: to clarify the effects of different (meta)phonological skills on typical and atypical literacy development, investigate the role of RAN in literacy acquisition, and explore potential differences in (meta)phonology acquisition between monolingual and bilingual individuals.

A total of 172 Italian children (76 mono-literate bilinguals with various heritage languages+96 monolinguals matched for age, socio-economic status, and educational background) participated from the last year of kindergarten to the 2nd/3rd grade of primary school in the same school setting. A comprehensive set of (meta)phonological skills (including phonological memory, syllable awareness in kindergarten, and phonemic awareness in primary school) and RAN categories (pictures, colors, and digits) were tracked besides literacy skills with standardized tests through 5 testing sessions. A meta-phonological (sound recognition and manipulation activities of increasing difficulty) and a RAN (tachistoscope rapid denomination activities) intervention were administered to different subgroups and compared with a graphic intervention (frame-matrices copying) and no intervention.

A three-stage developmental pattern emerged revealing *(1)*an initial contribution of syllabic awareness to literacy besides a prominent load of working memory, which progressively relieved as long as *(2)*automaticity skills and phonemic awareness would be intervening; *(3)*in half 2nd grade, language plus complex manipulatory meta-phonological skills were the main predictors. Diverse subcomponents of (meta)phonology and RAN predicted specific literacy portions at different developmental stages. The effects of (meta)phonology and RAN extended to accuracy and speed, highlighting a bias in existing literature due to partial analyses. Two meta-phonological measures (initial syllable and rhyme recognition) predicted specific literacy impairments. Double (kindergarten+1st grade) meta-phonological and RAN interventions positively affected particular aspects of reading and writing. Notably, mono and bilinguals exhibited similar skills since kindergarten in two measures (pseudoword repetition and syllable segmentation) that predicted specific literacy skills in 1st grade.

Overall, these findings interpret RAN as a measure of phonological automaticity, identify specific testing measures for Italian (meta)phonology and reading regardless of children's language dominance, highlight effective structural interventions on early literacy, and suggest that the proposed 'bilingual advantage' might either have no cognitive nature or develop at a later stage.

[1] Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. 1999. *Comprehensive Test of Phonological Processing* (*CTOPP*). Austin, TX: Pro-Ed.
[2] Norton, E.S., & Wolf, M. 2012. Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*, 427-452.

Perception of plain-ejective contrast by English- and Spanish-speaking listeners: L1 and orthography effects

Jack Mahlmann[1], Yoonjung Kang[1,2]
[1]*University of Toronto,* [2]*University of Toronto Scarborough*

Introduction: A recent study has shown that a period of silence vs. aspiration after a stop burst (RELEASE_TYPE) is a primary perceptual cue for distinguishing ejectives from non-ejectives, whereas Voice Onset Time (VOT) plays a relatively minor role [4]. Recent studies also highlight the significant effects of orthography in L2 acquisition [1]. We conducted a perceptual training study to investigate English- and Spanish-speaking listeners' acquisition of this novel RELEASE_TYPE cue after brief exposure and distributional training [3] in order to explore the influence of L1 and orthography in non-native ejective perception. Specifically, Spanish listeners are hypothesized to be better than English listeners, because English distinguishes long-lag from short-lag VOT stops and L1 transfer would predict reliance on VOT for non-native perception while VOT doesn't have the same role in Spanish.

Methods: (1) *Stimuli:* Stimuli were created from a token of naturally produced Q'anjob'al [Mayan] /tu/ and /t'u/ by manipulating the content (RELEASE_TYPE: aspiration vs. silence) and duration (VOT: 12 steps of 5ms intervals from 10ms ~ 65ms) of the post-burst release. (2) *Tasks* (see Table 1)*:* In the *Exposure* phase, the participants were presented with a subset of the stimuli for which plain and ejective stops are differentiated by both cue dimensions (Fig.1). As the two cues covary in the exposure data, participants could distinguish the two stops relying on VOT (a cue already available in L1) and/or RELEASE_TYPE (a novel cue). In subsequent phases, participants performed identification tasks on the entire stimuli space, including the stimuli where the two cues conflict (silence with short VOT and aspiration with long VOT). Feedback on their answers was provided only in *Training*, where the category label depended on RELEASE_TYPE inhibiting VOT cue use (Fig. 2). (3) *Participants* – 60 Canadian English speakers and 60 Mexican Spanish speakers were recruited over *Prolific*. They were each divided into three GROUPS, and saw either picture labels (picture), or pseudo-orthographic forms where plain and ejective stops were indicated by <t> and <t*> respectively (ortho) or the reverse (ortho-reverse) (Table 2). (4) *Statistical analysis*: A generalized mixed-effects model was run to estimate the response (ejective = 1, plain = 0) probability with LANGUAGE (English, Spanish), CONDITION (picture, ortho, ortho-reverse), RELEASE_TYPE (aspiration, silence), VOT (short: 10-35ms, long:40-65ms), PHASE (pre-test, post-test), and their interactions as fixed effects. Maximal random effects for ITEM and PARTICIPANT were included.

Results & Discussion: Ejective response percentages are summarized in Fig. 3 and significant cues determined by statistical tests is summarized in Table 3. Our hypothesis that Spanish listeners would fare better than English listeners turned out to hold for the pre-test PHASE of the picture GROUP, where RELEASE_TYPE was a significant cue for Spanish but not English listeners. In other words, after the exposure phase alone, Spanish listeners were already using RELEASE_TYPE to distinguish stops, while English listeners were using neither cue reliably. When orthography was introduced, the situation reverses: the English listeners reliably label any stimuli with aspiration, extending the <t> label of short VOT (per exposure) to long VOT (appropriate for English /t/). For the Spanish listeners, orthography hinders the RELEASE_TYPE cue use because aspirated stops with long VOT is not typical of Spanish voiceless stops (short-lag). This effect of orthography is not found in the ortho-reverse condition, indicating that it is these particular orthographic correspondences, not the general use of orthography, that aids English listeners but impedes Spanish listeners [2]. In post-test, all participant groups successfully used RELEASE_TYPE regardless of any earlier difficulties, showing the effectiveness of inhibition training. The results show that a combination of L1 category distribution, orthography, and training can aid the acquisition of a novel cue for ejective stops.

Table 1: Stimuli, tasks and experimental phases

	Exposure	Pre-test	Training	Post-test
Stimuli	Auditory + Label	Auditory	Auditory	Auditory
Response	NA	2AFC	2AFC	2AFC
Feedback	NA	None	Feedback	None
Trials	3 VOT * 2 Release * 8 reps = 48	12 VOT * 2 Release * 2 reps = 48	12 VOT * 2 Release * 2 reps = 48	12 VOT * 2 Release * 2 reps = 48

Table 2: Groups & labels

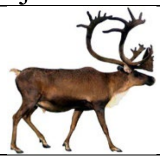
group	plain /tu/	ejective /t'u/
Picture		
Ortho	tu	t*u
Ortho-Reverse	t*u	tu

Fig. 1: Exposure stimuli Fig. 2: Training stimuli

Table 3: Cue use by GROUP, PHASE and LANGUAGE

group	Pre-test		Post-test	
	English	Spanish	English	Spanish
Picture	-	release	release	release
Ortho	release	vot	release	release
Ortho-Reverse	-	release	release	release

Figure. 3. Percent of ejective response separated by VOT, RELEASE TYPE, PHASE, GROUP & LANGUAGE

[1] Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, *54*(3), 297-326.
[2] Holliday, J. J. (2016). Second language experience can hinder the discrimination of nonnative phonological contrasts. *Phonetica*, *73*(1), 33-51.
[3] Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of phonetics*, *38*(4), 569-587.
[4] Percival, M. 2014. *Production and perception of ejective stops across languages.* University of Toronto Doctoral Dissertation.

Oral session 4: Multimodality

Audiovisual perception of vowels in native and non-native speech

Sinéad Marie Rankin, Maria-Josep Solé
Universitat Autónoma de Barcelona

This paper examines the integration of visual features of non-native and native speech. Specifically, it addresses how visual vowel gestures in L1 and L2 speakers affect the intelligibility of those vowels for L1 listeners. While most studies have investigated visual cues in L2 speech for Asian languages (e.g., [1, 2]), European languages have received less attention. Furthermore, much of the research on audiovisual L2 perception has focused on consonant visemes ([2, 3, 4] or word recognition [5, 6, 7], with relatively little attention given to vowel intelligibility. This study examines French productions of English vowels, focusing on visual features such as lip rounding, and investigates how L2-specific cues influence vowel intelligibility in noisy environments. Previous research shows that non-native speakers produce distinct visemes influenced by their native language's articulatory patterns [8], which can affect speech recognition, particularly in noise.

English CVC words with vowels involving salient (e.g., lip spreading for /iː/) and non-salient visual cues (neutral lips for /ɪ/), produced by two French speakers and a native English control, were presented to 24 native English listeners who identified the word heard. Tokens were presented in both auditory-only (A) and audiovisual (AV) mode in cafeteria noise at -15dB SNR. The visual cues studied were lip spreading, lip rounding, jaw opening and tongue frontness in vowels, as well as lip rounding in schwa. The distribution of errors and the audiovisual benefit (over audio-only listening) in background noise were analysed.

Overall, audiovisual benefits were found for both native and non-native speakers (see Figure 1); however, this aggregate similarity masked feature-wise differences for vowel pairs designed to isolate different visual features. Visually salient cues improved vowel intelligibility, compared to non-visual cues (Figure 2), but the audiovisual benefit varied across vowel features and speaker groups. The presence of lip-spreading for /iː/ (vs /ɪ/) and jaw-opening for /æ/ (vs /ɪ/) enhanced intelligibility (i.e., larger AV benefit) for both speaker groups. However, compared to the English speaker, lip rounding in /ɔː/ and /ə/ produced by French speakers (likely accompanied by lip protrusion) had a smaller, or negative AV benefit (Figure 3). These results suggest that the influence of L1 gestures on L2 production may reduce or negatively affect intelligibility. Furthermore, French productions of /ɑː/ exhibited unusually high AV benefits, suggesting an extreme jaw-opening for this vowel in an attempt to distinguish between L2 contrasts (/æ ʌ ɑː/) not present in the L1. These results illustrate that visual information is integrated in perception in complex, feature-specific ways that depend on the precise visual cue expectations of the listener that are not always met in L2 speech. Thus the study suggests that mismatches between L2 speakers' visual gestures and L1 listeners' long-term visual representations for speech sounds result in a smaller visual benefit for L2 speech as compared to L1 speech.

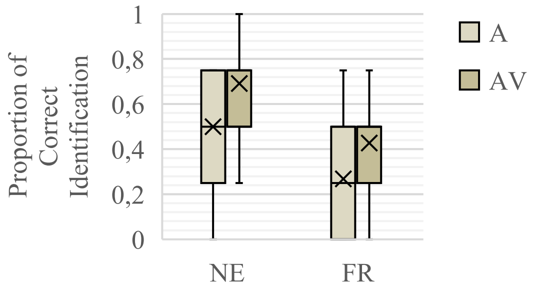


Figure 1. *The increase in intelligibility in the Audio-Visual condition was overall similar in both speaker groups. Mean proportion of vowel intelligibility in noise for native English (left) and French speakers (right) in the Audio-only (lighter fill) and Audiovisual condition (darker fill). The mean is indicated by × in the boxes.*

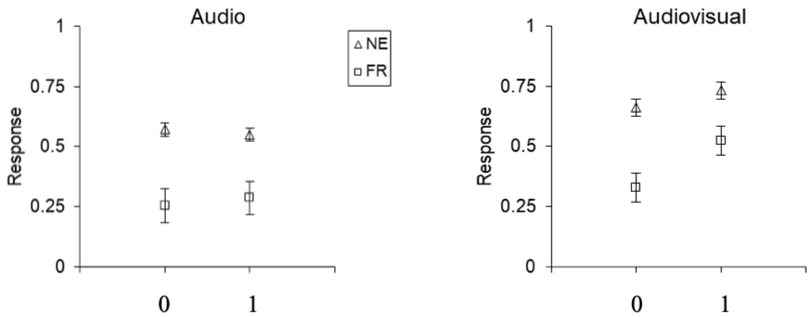


Figure 2. *This chart shows an increase in correct identification in the AV condition in vowels with a visual cue (1) compared to those with no visual cue (0), and that the effect of the visual cue (right panel) is larger for the French than the English speakers. Proportion of correct responses (y-axis) for non-visual (0) and visual cues (1) (on the x-axis) for the native English and French speakers in the Audio and Audiovisual conditions. Vowel features pooled.*

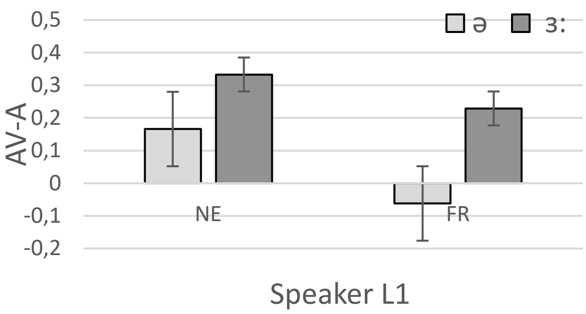


Figure 3. *Mean Visual benefit for schwa and /ɜː/ for native English and French speakers. This chart illustrates that differences between listener expectations of the nature and degree of lip rounding and the visual characteristics of L2 speech caused either weakened AV benefits or detriments in AV perception relative to audio-only.*

[1] Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. Journal of the Acoustical Society of America, 137, 2823–2833.

[2] Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America, 119*(3), 1740–1751.

[3] Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speaker, and listener effects. *Speech Communication, 52*(11-12), 996–1009.
[4] Kawase, S., Hannah, B., & Wang, Y. (2014). The influence of visual speech information on the intelligibility of English consonants produced by non-native speakers. *Journal of the Acoustical Society of America, 136*, 1352–1362.
[5] Yi, H. G., Phelps, J. E., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America, 134*, EL387–EL393.
[6] Xie, Z., Yi, H. G., & Chandrasekaran, B. (2014). Nonnative audiovisual speech perception in noise: Dissociable effects of the speaker and listener. *PLOS ONE, 9*(12), e114139.
[7] McLaughlin, D. J., Brown, V. A., Carraturo, S., & Van Engen, K. J. (2022). Revisiting the relationship between implicit racial bias and audiovisual benefit for nonnative-accented speech. *Attention, Perception, & Psychophysics, 84*, 2074–2086.
[8] Zerling, J. (1992). Frontal lip shape for French and English vowels. *Journal of Phonetics, 20*, 3–14.

Gestural rhythm mirrors rhythmic patterns of speech in New Zealand English and Turkish

Olcay Türk[1] and Wim Pouw[2]
[1] Phonetics Workgroup, Faculty of Linguistics and Literary Studies, Bielefeld University,
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University

Research on gesture-prosody interaction has focused primarily on intonation and stress, especially on their temporal and functional coordination with gestural units [1]. A potential interaction of speech rhythm with gesture, on the other hand, has received less attention. This is partly because the speech rhythm literature has been the subject of considerable debate, largely due to the complexity of defining, measuring, and categorizing rhythmic patterns across languages [2]. On the gesture side, rhythm has been studied within the context of co-speech gestures [3], signed languages [4], and more pure oscillatory movements [e.g., 5, 6], but not in a way that is directly relatable to linguistically constrained speech rhythms.

The present study analyses speech rhythm using envelope-based metrics [7] and adapts them to analyse gesture rhythm in New Zealand English (NZE) and Turkish (TR). The aim is to determine whether these languages show significantly different rhythmic patterns and whether (and which) rhythmic patterns in gesture mirror the patterns in speech.

The analysis is conducted on 1.2–1.6 second clips (*N* TR clips=1088, *N* NZE clips=420) containing both uninterrupted speech and co-speech gesture, extracted from recordings where participants (N = 5 TR, M=1, F=4; N = 8 NZE, M=2, F=6) recount videos they have watched [8]. Envelope-based metrics are calculated based on the decomposition of the amplitude envelope and the distribution of power across different frequency bands of the envelope. They capture (1) the relative influence, (2) the frequency, and (3) the variability of syllable-time and stress-time oscillations. In speech, these metrics are subject to various language-specific linguistic constraints (e.g., phonotactic complexity and vowel reduction), so cross-linguistic differences are expected. The gesture metrics are calculated based on the amplitude envelope of aggregated acceleration of the index finger, wrist, and elbow, forming a complex signal. Parallel to those in speech, the metrics are conceptualised to capture the relative influence, the frequency, and the variability of within-hand and whole-arm oscillations.

The results show statistically significant cross-linguistic differences in several speech rhythm metrics. These differences align with the traditional rhythm classifications of TR (syllable-timed) and NZE (stress-timed). For instance, the relative influence of stress-time oscillations on the envelope is higher in NZE than in TR, and syllable-time oscillations are more stable in TR than in NZE. Interestingly, most of these differences are mirrored in gesture metrics, with trends also reflecting cross-linguistic differences; for example, the relative influence of whole-arm oscillations dominates the signal in NZE but not in TR. A partial least squares canonical correlation analysis shows no significant correlations between speech and gesture metrics, with coefficients less than 0.2. Despite this finding, redundancy indices reveal that one third of the variance in gesture rhythm metrics can be explained by the variance in speech rhythm metrics. Additionally, metrics of the same type in speech and gesture tend to pair up and contribute robustly to the common rhythm structure.

Overall, the cross-modal relationship of rhythm metrics may be weak, but the datasets still coherently explain a notable amount of variability, suggesting a more complex, possibly non-linear relationships. Further analyses will focus on this possibility.

This research was funded by the German Research Foundation (DFG): TRR 318/1 2021–438445824

[1] Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview [Editorial]. *Speech Communication, 57,* 209–232. https://doi.org/10.1016/j.specom.2013.09.008

[2] Arvaniti, A. (2021). Measuring Speech Rhythm. In R.-A. Knight & J. Setter (Eds.), *The Cambridge Handbook of Phonetics* (pp. 312–335). chapter, Cambridge: Cambridge University Press.

[3] Pouw, W., Jaramillo, J., Dixon, J.A., Ozyurek, A. (2020). Quasi-rhythmic features of hand gestures show unique modulations within languages: Evidence from bilingual speakers. *Proceedings of the 7th meeting of Gesture and Speech in Interaction - GESPIN2020*

[4] Brookshire, G., Lu, J., Nusbaum, H. C., Goldin-Meadow, S., & Casasanto, D. (2017). Visual cortex entrains to sign language. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(24), 6352–6357. https://doi.org/10.1073/pnas.1620350114

[5] Rathcke, T., Lin, C., Falk, S. & Dalla Bella, S., (2021) "Tapping into linguistic rhythm", *Laboratory Phonology* 12(1): 11. doi: https://doi.org/10.5334/labphon.248

[6] Weston, H., Pouw, W., & Fuchs, S. (2024). On the Relation Between Leg Motion Rate and Speech Tempo During Submaximal Cycling Exercise. *Journal of Speech, Language, and Hearing Research*, *67*(10S), 3931–3946. https//doi.org/10.1044/2023_JSLHR-23-00178

[7] Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, *134*(1), 628–639. https://doi.org/10.1121/1.4807565

[8] Turk, O. (2020). Gesture, Prosody and Information Structure Synchronisation in Turkish. Open Access Te Herenga Waka-Victoria University of Wellington. Thesis. https://doi.org/10.26686/wgtn.17149058.v1

How beat gestures and written stress marks affect lexical stress production in Spanish

Doenja Hustin[1], Hans Rutger Bosker[2], Marieke Hoetjes[1], Wim Pouw[2], Lieke Van Maastricht[1]
[1]Centre for Language Studies, Radboud University, [2]Donders Institute for Brain, Cognition and Behaviour, Radboud University

Lexical stress is an important linguistic feature in communication, affecting the overall perceived rhythm of speech, as well as playing a crucial role in word recognition ([1], [2]). Prior studies have shown that beat gestures align prosodically with speech ([3]) and, more recently, that they also boost prosodic stress cues through biomechanical coupling, enhancing intensity and/or F0 in English and Dutch ([1], [4], [5]). However, it remains unclear how beat gestures affect lexical stress realization in other, non-Germanic, languages, like Spanish. Moreover, little is known about how biomechanical coupling effects interact with other factors influencing lexical stress production, such as orthographic stress marks. Spanish lexical stress placement follows consistent rules in which these written stress marks play an important role ([6]). Moreover, Spanish is especially interesting to investigate due to the abundant presence of minimal lexical stress pairs and the critical role that lexical stress plays in conveying semantic meaning in the regular verb conjugation paradigm (e.g., 'HAblo' means *I speak*, while 'haBLÓ' means *(s)he spoke*, capital letters representing lexical stress). Hence, our study aims to determine what the effect is of orthographic stress mark presence and the production of a beat gesture on the acoustic realization of lexical stress in Spanish.

To answer this question, we replicated [7] with 26 native speakers of Castilian Spanish, who were video recorded producing 96 Spanish multi-syllable words (e.g., 'miCRÓfono, 'profeSOR') that were counterbalanced for the presence of a written stress mark. Standing up and looking at a screen, the participants read aloud all stimuli twice, once with and once without a beat gesture (see our experimental set-up in Figures 1 and 2). Acoustic analyses using Praat ([8]) were performed to assess whether producing a beat gesture and/or the presence of a written stress mark impacted the acoustic realization of lexical stress, measured as F0, intensity and duration. Kinematic analyses using motion-tracking analyses to assess the temporal relationship between the acoustic realization of stress and the beat gestures are underway.

Statistical analyses are ongoing, though preliminary findings reveal that beat gestures enhance intensity but not F0 and duration, replicating earlier work on biomechanical coupling of gesture and speech in English and Dutch ([9], [10]). Additionally, the presence of a written stress mark seems to enhance pitch and intensity peaks in the stressed syllable indicating that it also boosts the acoustic realization of lexical stress. However, orthographic stress marks also seem to reduce the duration of the relevant syllable in comparison to the stressed syllables of words without a written stress mark. We hypothesize that this reduction of duration might be because a written stress mark aids the word recognition process and/or reduces cognitive processing demands as suggested by [11], but more detailed data analysis is needed to investigate this further. Moreover, initial analysis did not reveal an interaction effect between beat gesture production and the presence of a written stress mark. More detailed analyses, including kinematic analyses to assess gesture-speech synchrony and a comparison with an existing L2 dataset from [7] are underway and will offer insights into how gesture and prosody interact in both L1 and L2 speakers of Spanish. These findings will also allow us to investigate whether producing beat gestures benefits L2 learners (of Spanish), who often struggle with lexical stress placement ([12]). As such, this study contributes to our understanding of gesture-speech interaction in general and serves as a foundation for research on lexical stress acquisition by L2 learners.

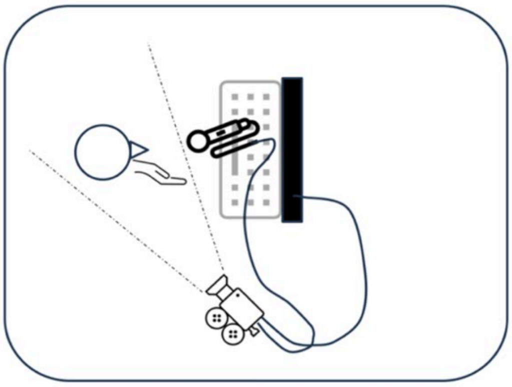


Figure 1: Graphical overview of experiment set-up.

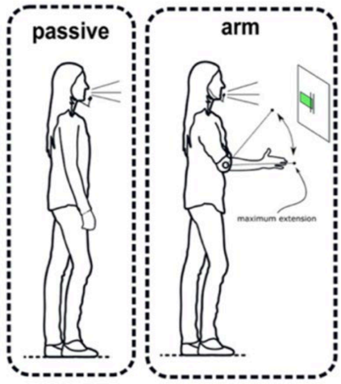


Figure 2: Graphical overview of gesture condition [5].

[1] Bosker, H. R., & Peeters, D. (2021). Beat gestures influence which speech sounds you hear. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1943).

[2] Morett, L. M., Fraundorf, S. H., & McPartland, J. C. (2021). Eye see what you're saying: Contrastive use of beat gesture and pitch accent affects online interpretation of spoken discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(9), 1494–1526.

[3] Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes, 26*(10), 1457–1471.

[4] Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*(3), 396–414.

[5] Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., & Dixon, J. A. (2020). Gesture–speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, *1491*(1), 89–105.

[6] Morales-Front, A. (2014). El acento [The accent]. In R. A. Núñez Cedeño, S. Colina, & T. G. Bradley (Eds.), *Fonología generativa contemporánea de la lengua española [Contemporary generative phonology of the Spanish language]* (pp. 235–265). Georgetown University Press.

[7] Bosker, H. R., Hoetjes, M., Pouw, W., & Van Maastricht, L. (2021). Gesture-speech coupling in L2 lexical stress production: A pre-registration of a speech acoustic and gesture kinematic study.

[8] Boersma, P., & Weenink, D. (2024). *Praat: doing phonetics by computer* (6.4.07) [Computer program].

[9] Pouw, W., Harrison, S. J., Esteve-Gibert, N., & Dixon, J. A. (2020). Energy flows in gesture- speech physics: The respiratory-vocal system and its coupling with hand gestures. *The Journal of the Acoustical Society of America*, *148*(3), 1231–1247.

[10] Werner, R., Selen, L., & Pouw, W. (2024). Arm movements increase acoustic markers of expiratory flow. *Proceedings of the International Conference of Speech Prosody (Speech Prosody 2024)*.

[11] Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*(3), 412–432.

[12] Da Silveira, A. P., van Heuven, V. J., Caspers, J., & Schiller, N. O. (2014). Dual activation of word stress from orthography. *Dutch Journal of Applied Linguistics*, 3(2), 170–196.

Body gestures in pronunciation learning: A linguistic universal?

Josefina Carrera-Sabaté,[1] Patricia Lopez-Garcia,[2] Joan Borràs-Comes [1]
[1] Universitat de Barcelona, [2] Universitat de València

Gesture, speech, and language are related both developmentally and neurologically [1, 2]. Neural control for speech and gesture overlaps, allowing neural activation to spread from one brain region to another, which may explain why they often co-occur [3], and it seems that this overlap is the result of an initial form of communication based on arm gestures to mouth articulation gestures [4]. Embodied cognition further emphasizes the interconnectedness of physiological experiences and psychological states, positing cognition as a combined physiological and psychological interpretation of the environment [5]. In line with this, there exists a category of sound symbolism effects that associates a vocal sound with a particular body action. This sound-action symbolism operates within the levels of motor, perceptual, and conceptual representations and provides the mechanism for the grounding of language in sensorimotor systems [6]. However, limited studies have explored the role of gestures in vowel production, particularly in languages other than English (e.g., [7, 8]). Moreover, in second language acquisition and language therapy, the role of gestures has been successfully used [9, 10, 11]. In this study, we analyze 1,080 French vowels and 1,080 Catalan vowels to examine the effect of non-representational body gestures that either facilitate or inhibit vowel articulation.

Hypothesis: Following previous literature, we expect facilitating gestures (open arms and hands) to correlate with more open vowels, while absence of gestures or inhibitory gestures will be associated with more closed vowels.

Methodology: Participants (female students, aged 25–50) were asked to pronounce three French and Catalan vowels in consonant-vowel-consonant logatoms: [pa ˈpa] / [ta ˈta] / [ka ˈka], [pɛ ˈpɛ] / [tɛ ˈtɛ] / [kɛ ˈkɛ], [pɔ ˈpɔ] / [tɔ ˈtɔ] / [kɔ ˈkɔ]. These vowels were produced in three postures (sitting, lying, standing) with three conditions: facilitating gesture, inhibitory gesture, and no gesture. We acoustically analyzed stressed vowels, focusing on F0, F1, F2, F3, F4, duration, and intensity. Statistical analysis was conducted using generalized mixed models with interactions between condition and vowel type as fixed effects and additional random effects. Reported p-values were derived from likelihood ratio tests.

Results: Findings indicate that stressed vowels [ɛ] and [ɔ] in both Catalan and French were modified by gesture, with F0 and F1 showing higher values for facilitating gestures and lower values for inhibitory gestures. Additionally, inhibitory gestures significantly reduced both intensity and duration across the two languages.

Conclusions: Our study demonstrates a relationship between facilitating gestures and more open, more intensity medium vowels in Catalan and French. These findings suggest that facilitating gestures may universally correspond to mid-vowel openness, supporting further exploration of gestures as tools in pronunciation learning.

[1] Iverson, J. M. & Thelen, E. 1999. Hand, mouth and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studi*es 6(11–12), 19–40.
[2] Bates, E. & Dick, F. 2002. Language, gesture, and the developing brain. *Developmental Psychobiology* 40(3), 293–310. https://doi.org/10.1002/dev.10034
[3] Capone, N. C. & McGregor, K. K. 2004. Gesture development: A review for clinical and research practices. *Journal of Speech, Language, and Hearing Research* 47(1), 173–186. https://doi.org/10.1044/1092-4388(2004/015)
[4] Gentilucci, M. & Corballis, M. C. 2006. From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews* 30(7), 949–960.

[5] Ye, H. 2010. Embodied cognition: A new approach in cognitive psychology. *Advances in Psychological Science* 18(5), 705–710.
[6] Perniss, P. & Vigliocco, G. 2014. The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B* 369(1651), 20130300. https://doi.org/10.1098/rstb.2013.0300
[7] Krahmer, E. & Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57(3), 396–414. https://doi.org/10.1016/j.jml.2007.06.005
[8] Pouw, W., Harrison, S. J. & Dixon, J. A. 2019. Gesture–speech physics: The biomechanical basis for the emergence of gesture-speech synchrony. *Journal of Experimental Psychology: General* 149(2), 391–404. https://psycnet.apa.org/doi/10.1037/xge0000646
[9] Guberina, P. 1985. The role of the body in learning foreign languages. *Revue de Phonétique Appliquée* 73–74–75, 37–50.
[10] Ghio, A. & Tellier, M. 2011. Les mouvements de la parole: De l'articulation à la gestuelle. *HAL Science Ouverte*, hal-01491788, 5-10. https://hal.science/hal-01491788
[11] Hoetjes, M.W., Maastricht, L.J. van & Heijden, L. van der. 2019. Gestural training benefits L2 phoneme acquisition: Findings from a production and perception perspective. In Grimminger, A. (Ed.), *Proceedings of the 6th Gesture and Speech in Interaction conference (GeSpIn)*. 50–55.

Oral Session 5: Perception

Perceptual sensitivity to stress-based f0 in Singapore English

Jasper Sim[1], Julia Schwarz[2], Katrina Kechun Li[3], Adam Chong[4]
[1]Nanyang Technological University, Singapore; [2]University of Cambridge
[3]Leiden University; [4]Queen Mary University of London

Native speakers of stabilised hybrid/contact varieties like Singapore English (SgE) may rely on different acoustic cues to distinguish words compared to speakers of the ancestral varieties [1], [2]. Pitch as a cue for lexical stress is one such example. In mainstream English varieties, a lexically stressed syllable has higher f0 than an unstressed syllable. Constrastingly, in SgE, f0 rises across single words regardless of the underlying lexical stress, and stress-initial words are distinguished from non-stress-initial words through a higher mean f0 at the word onset and possibly an earlier f0 elbow [3], [4] (Figure 1). Whether SgE speakers can also perceive these stress-based f0 patterns observed in their production, and consequently, whether these are used as discriminative cues in spoken word recognition and lexical access remains unclear [5].

In this pilot study, we asked (i) whether SgE-speaking listeners are perceptually sensitive to the f0 differences that distinguish SgE words with different underlying stress, and (ii) whether their ability to discriminate is modulated by individual bilingualism. We employed an approach that involves swapping the pitch contours of SgE disyllabic word pairs that have different underlying lexical stress but similar segmental properties (e.g. carTOON ↔ CARpet). We expected that SgE-speaking listeners would judge the stimuli with the original/expected pitch contour (Matching) as more typically Singaporean than those with the Mismatched pitch contour (Task 1). We also expected that, in a paired comparison task, listeners would show a preference for the Matching stimuli over the Mismatched ones as being more typical (Task 2).

Participants were 16 native Singaporeans who were bilinguals of SgE and an ethnic mother tongue [EMT; Mandarin (n=6), Malay (n=8), or Tamil (n=2)]. They were between 21 and 36 years old (M_{age}=28, SD_{age}=4.8; males=10), with no reported hearing or language impairments. The stimuli were 15 SgE word pairs, produced by a male 32-year-old native English-Mandarin bilingual Singaporean. The smoothed f0 contours of the pre-synthesised stimuli were overall consistent with the previously reported production patterns (Figure 1; [3]). Participants completed the tasks on Gorilla.sc following a headphone screening test [6].

In Task 1, all stimuli (i.e., Matching and Mismatched) were rated as highly typically Singaporean on a seven-point likert scale, and the difference in ratings between Matching, $M(SD)$=5.47(1.42), Mdn=6, and Mismatched stimuli, $M(SD)$=5.53(1.30), Mdn=6, was small. Mixed-effects ordinal regression did not reveal significant difference in ratings between conditions (Matching, Mismatched) nor an effect of listener's EMT (Mandarin, Malay, Tamil). In Task 2 (paired comparisons), Matching stimuli were regarded as more typical on average 58% of the time. Mixed-effects logistic regression revealed that an interaction between lexical stress and listener's EMT significantly predicted outcomes (Figure 2): For stress-initial words only, Mandarin-speaking listeners were more likely to prefer the Matching stimuli compared to Malay-speaking (b=1.02, p=0.005) and Tamil-speaking listeners (b=1.29, p=0.02). Mandarin-speaking listeners were also more likely to prefer the Matching stimuli of stress-initial words than stress-final words (76% v. 54%, b=0.97, p=0.004; Figure 2).

Our preliminary findings suggest that stress-based f0 manipulations may not be particularly salient to some listeners, even when directly comparing Matching–Mismatched pairs, which indicate that stress-related f0 changes may not significantly influence naturalness, and/or that SgE speakers are highly tolerant of small pitch modulations. Interestingly, Mandarin bilinguals were much more sensitive to f0 manipulations, but only for stress-initial words (e.g. CARpet with carTOON's f0), which suggests a potential influence of tonal language background in prosodic processing in SgE. In our full-scale study, we will consider other social factors, their production patterns, and other correlates of stress in perception and also in online processing.

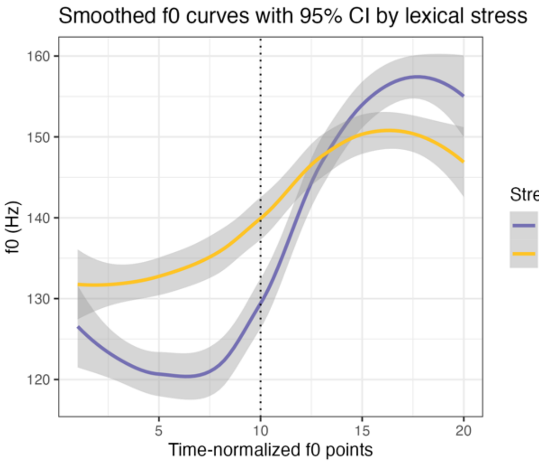


Figure 1: Time normalised plots of f0 averaged over each stress condition (Su: stress-initial; uS: stress-final) for the disyllabic stimuli. Dotted line represents syllable boundary.

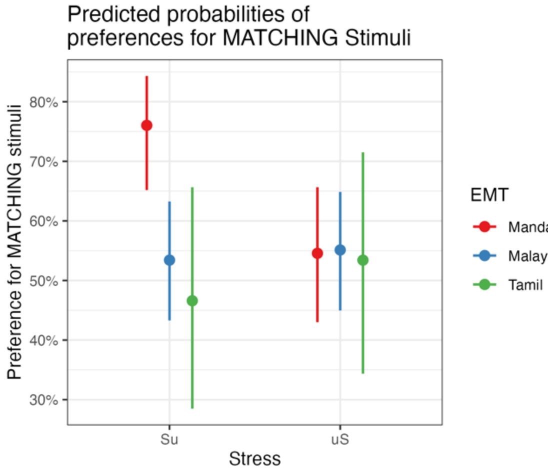


Figure 2: Predicted effects of stress condition and ethnic mother tongue (EMT) on choice in the paired comparison task (Task 2).

[1] J. H. Sim, "Variation in VOT in English child-directed speech of English-Mandarin and English-Malay early bilinguals in Singapore," *Asian Englishes*, pp. 1–17, Sep. 2024, doi: 10.1080/13488678.2024.2391234.
[2] J. H. Sim and B. Post, "Early phonological acquisition in multi-accent contexts," in *Multilingual acquisition and learning: Towards an ecosystemic view to diversity*, E. Babatsouli, Ed., John Benjamins Publishing Company, 2024. [Online]. Available: https://doi.org/10.1075/sibil.67.07sim
[3] A. J. Chong and J. S. German, "Prominence and intonation in Singapore English," *Journal of Phonetics*, vol. 98, p. 101240, May 2023, doi: 10.1016/j.wocn.2023.101240.
[4] J. S. German and A. J. Chong, "Stress, tonal alignment, and phrasal position in Singapore English," in *TAL2018, Sixth International Symposium on Tonal Aspects of Languages*, ISCA, Jun. 2018, pp. 150–154. doi: 10.21437/TAL.2018-30.
[5] K. Connell *et al.*, "English Learners' Use of Segmental and Suprasegmental Cues to Stress in Lexical Access: An Eye-Tracking Study," *Language Learning*, vol. 68, no. 3, pp. 635–668, 2018, doi: 10.1111/lang.12288.
[6] K. J. P. Woods, M. H. Siegel, J. Traer, and J. H. McDermott, "Headphone screening to facilitate web-based auditory experiments," *Atten Percept Psychophys*, vol. 79, no. 7, pp. 2064–2072, Oct. 2017, doi: 10.3758/s13414-017-1361-2.

Flexible vowel categorisation within and across Standard Southern British English and New Zealand English dialects

Stephanie Cooper, Brechtje Post
University of Cambridge

Listeners are highly capable of interpreting variation in speech. This has been attributed to the flexibility of phonetic category boundaries, especially in vowels, which appear to be more accepting of variation than consonants [1], [2]. Whether a variant incoming vowel is classified as a member of a given category may depend on acoustic parameters, such as spectral distance from the category, or the featural specifications of the category itself [3]. Classification may also be influenced by top-down factors, such as lexical bias (preference of perceiving a real word) [4], or a listener's predictive model of how the speaker produces sounds [5].

It is unlikely that these factors apply uniformly across all vowel categorisation, but their limitations are not well understood. Across two experiments, we aimed to establish how some of these factors interact, how they become evident across different combinations of listener and speaker dialects, and how they can vary across the vowel space.

Participants spoke either Standard Southern British English (SSBE) or New Zealand English (NZE). To investigate perceptual boundaries of vowels, eight continua (Table 1) of VC and CVC-structured sequences were used in a forced choice identification task. One experiment used materials produced by a male SSBE speaker, and another used materials produced by a male NZE speaker. We therefore had instances of both SSBE and NZE participants listening to speech from dialects that were familiar and unfamiliar to them.

We were interested in dialectal variation and flexibility around the DRESS and THOUGHT vowels, as DRESS differs between SSBE and NZE, while THOUGHT does not as much [6], [7]. Each continuum contained vowels re-synthesised in steps along either a FLEECE-TRAP front vowel or a GOOSE-LOT back vowel continuum. Each continuum maintained the same consonant context, e.g. ranging from *heed* through *hid* and *head* to *had,* while also containing steps between these points to allow indication of where boundaries would shift. The continua also varied in their lexical status: some contained only words, some only nonwords, and some both. In each trial of the forced choice task, participants heard a sound from one of the continua and were asked to select what they heard from a list of options (Figure 1).

Generalised additive mixed models examined differences in categorisation curves (% classification as a given vowel) for each point along a vowel continuum between groups, vowels, and lexical contexts. Preliminary results indicate that across front and back vowel continua for both dialect groups, vowels with less distinct feature specification were more susceptible to the effects of lexical bias. Additionally, SSBE listeners categorised NZE vowel continua in ways consistent with their own native dialect, while NZE listeners categorised SSBE vowel continua in ways consistent with the speaker dialect. NZE listeners may have been able to identify SSBE features in the continua, like the presence of lowered TRAP.

These findings have implications for multiple theories of speech perception. Findings that featurally underspecified mid vowels are more accepting of variation [8] are supported, and it is also revealed that top-down processes such as lexical bias are more powerful when applied to boundaries between categories that are underspecified and ambiguous. Additionally, experience with a dialect or type of speech seems to enable listeners to select a speaker model based on relatively little, yet reliable, information from speech, such as SSBE TRAP, that can then inform deduction of other more ambiguous segments in the signal, such as SSBE DRESS. This is in line with findings that assumptions about a speaker can have substantial influence on processing of their speech [5], [9]. This study reveals more about the many factors that influence vowel categorisation, and how they act in conjunction with each other. It provides avenues for further research into the neural underpinnings of flexible vowel categorisation.

Frontness	Lexicality	Contents
Front	All	**heed** - - - - - - - **hid** - - - - - - - **head** - - - - - - - **had**
	None	heeb - - - - - - - hib - - - - - - - heb - - - - - - - hab
	Medial	eeg - - - - - - - igg - - - - - - - **egg** - - - - - - - agg
	Peripheral	**feet** - - - - - - - **fit** - - - - - - - fet - - - - - - - **fat**
Back	All	**shoot** - - - - - - **short** - - - - - - - **shot**
	None	foop - - - - - - - forp - - - - - - - fop
	Medial	oob - - - - - - - **orb** - - - - - - - obb
	Peripheral	**hoop** - - - - - - - horp - - - - - - - **hop**

Table 1. *Continua used in experiment. Words are in bold. Dashes indicate steps synthesised between canonical productions.*

Figure 1. *Example screen from forced choice identification task.*

[1] J. Shaw *et al.*, "Effects of short-term exposure to unfamiliar regional accents: Australians' categorization of London and Yorkshire English consonants," in *Proceedings of the 15th Australasian International Speech Science and Technology Conference*, Dec. 2014, pp. 71–74.

[2] J. Shaw *et al.*, "Resilience of English vowel perception across regional accent variation," *Lab. Phonol.*, vol. 9, no. 1, pp. 1–36, Jul. 2018.

[3] A. Lahiri and H. Reetz, "Distinctive features: Phonological underspecification in representation and processing," *J. Phon.*, vol. 38, no. 1, pp. 44–59, Jan. 2010, doi: 10.1016/j.wocn.2010.01.002.

[4] R. A. Fox, "Effect of lexical status on phonetic categorization," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 10, no. 4, pp. 526–540, 1984, doi: 10.1037/0096-1523.10.4.526.

[5] D. F. Kleinschmidt and T. F. Jaeger, "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," *Psychol. Rev.*, vol. 122, no. 2, pp. 148–203, 2015, doi: 10.1037/a0038695.

[6] L. Bauer, P. Warren, D. Bardsley, M. Kennedy, and G. Major, "New Zealand English," *J. Int. Phon. Assoc.*, vol. 37, no. 1, pp. 97–102, 2007.

[7] A. Hughes, P. Trudgill, and D. Watt, English accents and dialects: An introduction to social and regional varieties of English in the British Isles, 5th ed. Routledge, 2013.

[8] M. Scharinger, P. J. Monahan, and W. J. Idsardi, "Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations," *NeuroImage*, vol. 128, pp. 293–301, Mar. 2016, doi: 10.1016/j.neuroimage.2016.01.003.

[9] J. Hay, A. Nolan, and K. Drager, "From fush to feesh: Exemplar priming in speech perception," vol. 23, no. 3, pp. 351–379, Oct. 2006, doi: 10.1515/TLR.2006.014.

The role of cognitive and perceptual abilities in mastering native-like rhythm in second language learners

Linda Bakkouche, Brechtje Post, John Williams
University of Cambridge

Rhythm in language arises from a combination of phonological and prosodic features, including syllable structure, vowel reduction, and accentual lengthening, resulting in distinct acoustic patterns. English, being a stress-timed language, poses unique challenges for Polish learners who come from a "mixed-rhythm" language background. While language experience can support the development of rhythmic accuracy in a second language (L2) [1], it does not fully explain higher levels of attainment in adult learners [2]. Recent research has suggested that specific cognitive and perceptual abilities play a critical role in L2 speech proficiency [3], but the influence of these abilities may vary depending on the learner's stage of acquisition [4]. This study aimed to identify the cognitive and perceptual factors that contribute to achieving native-like rhythm among Polish learners of English at intermediate and advanced proficiency levels.

The study explored how the proficiency level of Polish learners affected their ability to produce rhythm patterns characteristic of native English speakers, specifically focusing on rhythm metrics and accentual lengthening. Additionally, it examined the extent to which perceptual (e.g., auditory acuity) and cognitive abilities (e.g., selective attention, phonological short-term memory, declarative and procedural memory), along with experiential factors (e.g., length of residence in an English-speaking environment and previous foreign language learning experience), influenced rhythm production accuracy in these learners.

A total of 90 adult participants were recruited, divided into three groups: intermediate Polish L2 learners of English ($n = 33$), advanced Polish L2 learners ($n = 32$), and native Southern Standard British English (SSBE) speakers as the control group ($n = 25$). Rhythm accuracy in a sentence reading task was assessed using acoustic analysis with Praat software [5], and rhythm metrics such as %V, VarcoV, VarcoC, rPVI-C, rPVI-V, nPVI-C, and nPVI-V were calculated [1]. The detailed experiment design is shown in ***Figure 1***.

The results revealed significant differences in rhythm accuracy based on proficiency level. Intermediate learners showed significantly lower values for accentual lengthening (65.4% increase, $t = 4.5$, $p = .002$) and final lengthening (129.6% increase, $t = 3.45$, $p = 0.011$) compared to native English speakers. In contrast, advanced learners demonstrated a more native-like differentiation among unaccented, accented, and final syllable durations, aligning more closely with the rhythm patterns of native English speakers (***Figure 2***). Additionally, vocalic variability increased with proficiency (VarcoV, $F(2, 330) = 25.47$, $p < .001$; nPVI-V, $F(2, 330) = 28.92$, $p < .001$) (***Figure 3***), reflecting the adaptation to English's varied syllable structure.

Mixed-effects model analyses (***Figure 4***) revealed that **temporal acuity** ($\beta = .876$, $SE = .789$, $t = 3.540$, $p < .005$) and **length of residence** (LOR) ($\beta = .432$, $SE = .123$, $t = 4.444$, $p < .001$) were significant for **intermediate learners**, with **PSTM** also playing a minor role ($\beta = .456$, $SE = 1.678$, $t = 2.060$, $p = .039$). For **advanced learners**, **procedural memory** ($\beta = .567$, $SE = .123$, $t = 3.384$, $p < .005$) and **selective attention** ($\beta = .345$, $SE = .567$, $t = 3.460$, $p < .003$) became crucial, while **temporal acuity** remained important ($\beta = .654$, $SE = .789$, $t = 2.740$, $p < .005$). **LOR** continued to have a marginal effect ($\beta = .432$, $SE = .234$, $t = 1.971$, $p = .049$), emphasizing the varying influences of cognitive abilities and experience at different proficiency levels.

This study's findings align with existing literature on L2 rhythm acquisition [1] highlighting the importance of proficiency level in rhythm accuracy. Despite advanced learners demonstrating more target-like rhythm patterns, significant individual variation persisted, indicating that certain learners had yet to fully master native-like rhythm. This variability underscores the influence of cognitive and perceptual abilities at different acquisition stages. These insights suggest the potential for developing targeted interventions to support adult L2 learners at different stages of acquisition, thereby facilitating faster and more native-like rhythm production.

Figure 1: *Experiment design.*

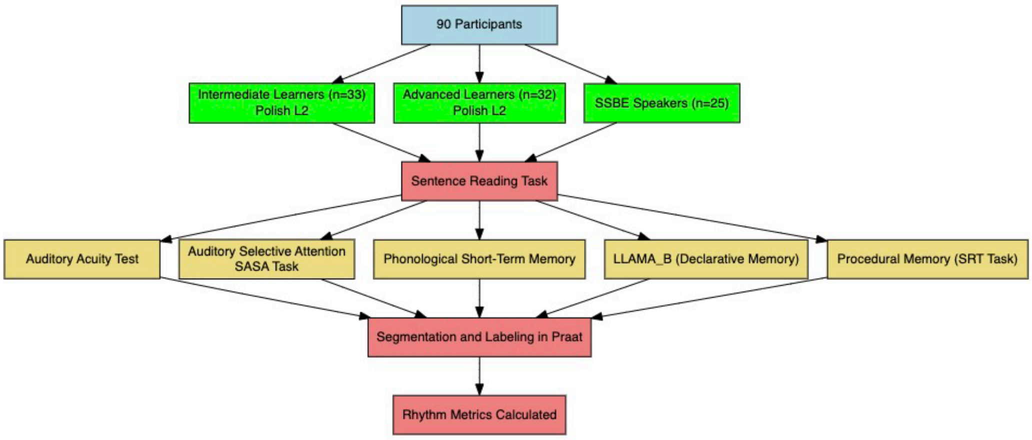


Figure 2: *Prosodic lengthening in SSBE (L1EN) and L2 Groups: Intermediate and Advanced*

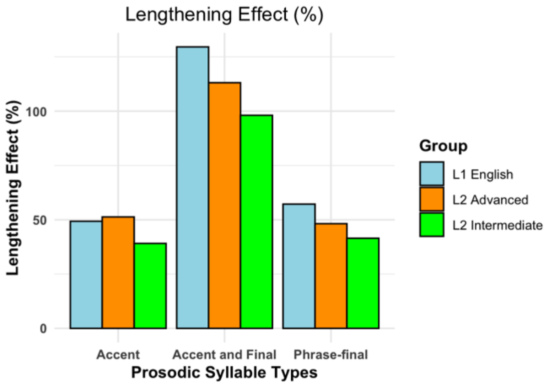


Figure 3

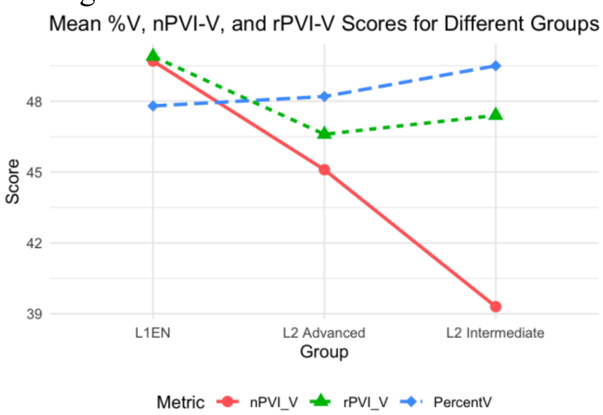
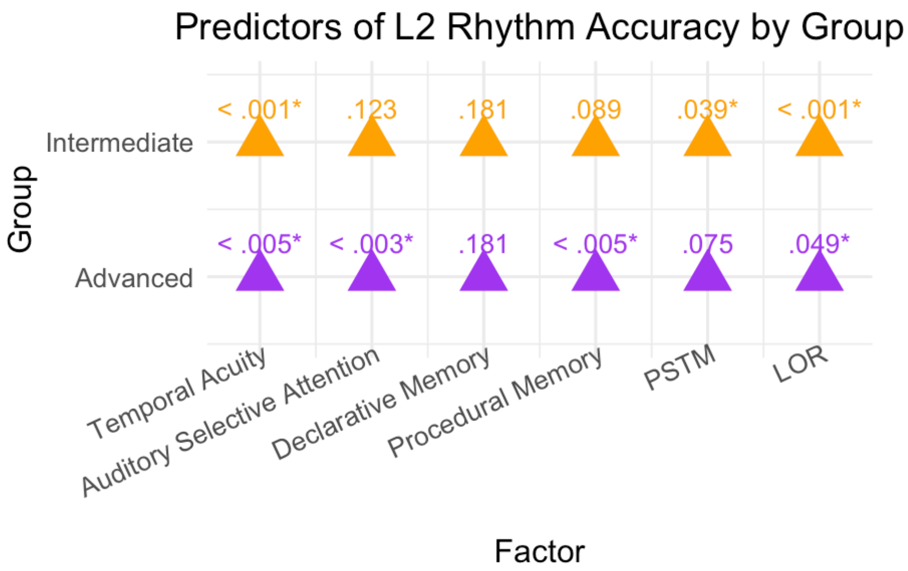


Figure 4



[1] Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition, 36*(2), 223-255. https://doi.org/10.1017/S0272263113000752
[2] Saito, K. (2018). Advanced second language segmental and suprasegmental acquisition. In *The Handbook of Advanced Proficiency in Second Language Acquisition* (pp. 282-303). Wiley Blackwell.
[3] Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*(3), 530–566. https://doi.org/10.1111/lang.12011
[4] Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition* (pp. 15-38). John Benjamins
[5] Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer (Version 6.1.16) [Computer software]. http://www.praat.org/
[6] Ullman, M.T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 953–68). CA: Elsevier.

Articulatory phonology and perception of gesture timing in aspiration

Chelsea Sanker
Stanford University

Introduction: Is aspiration part of the preceding consonant, part of the following vowel, or both? It behaves in many ways like part of the consonant which it is marked as occurring on. This association with the consonant is reflected in perceptual tests; Voice Onset Time (VOT) is a major factor in perception of laryngeal contrasts (e.g. Schertz et al 2015). Following vowel duration has an effect which is much smaller and in the opposite direction, i.e. longer vowels decrease perception that a consonant is aspirated (Toscano & McMurray 2012).

However, aspiration also contributes to listeners' perception of vowel duration; vowels are perceived as longer following an aspirated stop than following an unaspirated stop (van Wijk 1999; Gussenhoven & Zhou 2013). One potential explanation is that listeners compensate for the vowel duration expected in the consonantal environment, as vowels are shorter after aspirated stops (Crystal & House 1988). Alternatively, listeners might be sensitive to the overlapping consonant and vowel gestures within aspiration, while the beginning of the vowel gesture is not perceptible in unaspirated stops because it occurs during the full stop closure.

The current study tests whether within-category VOT manipulations impact listeners' perception of vowel duration. Vowels were more often identified as long when preceded by longer aspiration. The results are consistent with listeners perceiving an earlier beginning of the vowel gesture when the aspiration is longer.

Methods: 96 English speakers completed the task online. They heard monosyllabic English words beginning with aspirated stops and categorized the vowel in each one as being long or short in duration. Each participant heard 4 words with lengthened VOT (mean 137 ms) and 4 words with shortened VOT (mean 51 ms; see Figure 1 for an illustration), all with three different vowel durations (natural, 20% lengthened, 20% shortened), for a total of 24 items, each of which was evaluated 3 times. To ensure that listeners were not impacted by comparing lengthened-VOT and shortened-VOT versions of the same word, each listener only ever heard the same VOT manipulation for a particular word. Across listeners, there were 6 combinations of word set and which half of the words had shortened vs lengthened VOT.

Results: Model summaries for the results are given in Tables 1-2. Figure 2 visualizes the results. The duration of the vowel was a significant predictor of vowel duration identifications, confirming that listeners were able to do the task. The VOT manipulation was significant; vowels were more often identified as long when preceded by stops with lengthened VOT. The actual VOT of the specific stimulus was also significant (though highly collinear with VOT manipulation); vowels were more often identified as long with longer preceding VOT.

Conclusions: The within-category effects observed here exclude an account of compensation based on consonant category. Effects of within-category VOT on perceived vowel duration, in combination with across-category effects, support an analysis based in Articulatory Phonology (Browman & Goldstein 1989) and perception of gestures (Fowler 1996). Vowel gestures are often clearly reflected in formants during aspiration (e.g., Figure 1); articulators are already moving into their positions for the vowel.

Articulatory Phonology (AP) is usually used to model speech production, but perceptual theories like Motor Theory and Direct Realism propose that listeners also perceive articulatory gestures. The results here provide additional evidence specifically about the timing of those perceived gestures: As in AP, gestures can overlap. Perception which allows for overlapping gestures provides a straightforward explanation for why VOT contributes both to perception of consonants and also following vowels.

The results also have implications for stimulus design. Changing the VOT is also changing the acoustic cues to the vowel gesture, which could have a range of unintended side effects.

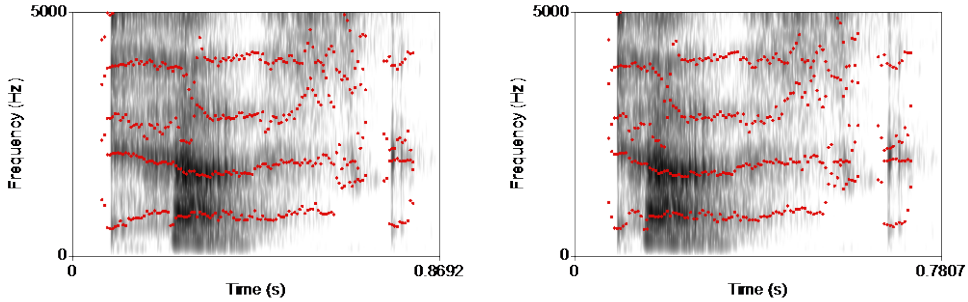


Figure 1. *Example stimuli, 'tusk' with lengthened VOT (left) and shortened VOT (right).*

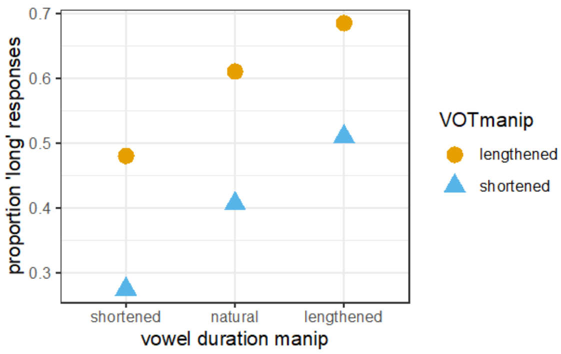


Figure 2. *'long vowel duration' responses by vowel duration and VOT manipulation*

	Estimate	Std. Error	Z value	P value
(Intercept)	0.47	0.198	2.37	< 0.0001***
VowelDuration	0.334	0.0201	16.6	< 0.0001***
VOTmanip Shortened	-1.02	0.0567	-17.9	< 0.0001***

Table 1. *Logistic model for 'long' responses. Intercept : VOTmanip = lengthened.*

	Estimate	Std. Error	Z value	P value
(Intercept)	-0.0376	0.184	-0.204	< 0.0001***
VowelDuration	0.335	0.0201	16.6	< 0.0001***
AbsoluteVOT	0.549	0.0294	18.7	< 0.0001***

Table 2. *Logistic model for 'long' responses.*

[1] Browman, C. & Goldstein, L. 1989. Articulatory gestures as phonological units. *Phonology* 6, 201-51.
[2] Crystal, T. & House, A.S. 1988. Segmental durations in connected-speech signals: Current results. *JASA* 83, 1553-1573.
[3] Fowler, C. 1996. Listeners do hear sounds, not tongues. *JASA* 99(3), 1730-1741.
[4] Gussenhoven, C., & Zhou, W. 2013. Revisiting pitch slope and height effects on perceived duration. *Proceedings of Interspeech 2013* (Lyon, France), 1365-1369.
[5] Schertz, J., Carbonell, K., & Lotto, A. 2020. Language specificity in phonetic cue weighting: Monolingual and bilingual perception of the stop voicing contrast in English and Spanish. *Phonetica* 77, 186-208.
[6] Toscano, J. & McMurray, B. 2012. Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics* 74, 1283-1301.
[7] van Wijk, M. 1999. The perception of aspiration: A cross-linguistic study. *Proceedings of ICPhS 1999* (San Francisco, USA), 853-856.

Effects of study abroad on L2 Perception of an innovative Spanish dialect variant

Angela Swain[1], Melinda Fricke[2]
[1]The Pennsylvania State University, [2]University of Pittsburgh

A study abroad context is conducive to observing shifts in a second language (L2) learner's phonetic knowledge due to increased input of social and linguistic information. Prior research has established that factors such as the L2 proficiency level, varieties of exposure, and learners' language attitudes towards local varieties can affect overall improvement [1, 2]. However, in the realm of phonetics and phonology, most research has focused on shifts in L2 production during study abroad programs, while L2 perception remains understudied [3].

Participants in the current study ($N = 15$; L1 English, undergraduate students at Penn State) spent four weeks studying Spanish in Cádiz, Spain, a coastal city in Andalusia associated with innovative phonetic traits such as the use of [s̪$^{\theta}$] (i.e., *ceceo*) [4]. The [s̪$^{\theta}$] variant is largely unfamiliar to learners and presents a discrepancy with students' prior expectations. As many varieties of Peninsular Spanish employ *distinción*, learners assume that an interdental [s̪$^{\theta}$] should be mapped to /θ/ rather than /s/.

We employed the Gating Paradigm [5], in which spoken words are progressively revealed over a series of short "gates." Target lexical items were frequent, disyllabic nouns; words were either control words or /s/ words. Half of the /s/ words contained the dental variant [s̪], which is associated with Seville capital and Latin American Spanish [4, 6], and half were produced with the local variant [s̪$^{\theta}$]. Talkers ($N = 4$; one man and one woman per variety) were native to Seville capital (i.e., producing [s̪]) or the outskirts of Seville (i.e., producing [s̪$^{\theta}$]).

Participants were presented with an auditory stimulus divided into five consecutive gates. For example, the word *sangre*, blood, would be heard as /s/, /sa/, /san/, /san.gɾ/, and /san.gɾe/. Participants were instructed to type the Spanish word after each gate presentation, regardless of how much acoustic information they had heard. At the end of the final gate presentation, participants rated the confidence of their last guess on a scale of 0 – 9 (i.e., least – most confident). Participants completed two experimental sessions: one shortly after their arrival to Cádiz and another at the end of the program, resulting in three weeks between sessions.

Results indicate differences in both accuracy and confidence ratings of /s/ words, with significantly lower accuracy for words produced with the innovative variant [s̪$^{\theta}$] when compared to control words and [s̪] words ($ps < 0.01$). Lower confidence ratings were observed for /s/ words when compared to control words ($p < 0.001$). However, answers for words produced with [s̪$^{\theta}$] were rated as significantly more confident ($p < 0.01$), despite being associated with lower accuracy. Differences also appear across experimental sessions; words containing [s̪$^{\theta}$] were recognized more accurately in Session 2 when compared to Session 1 [s̪$^{\theta}$] words ($p < 0.05$). However, confidence levels for /s/ words were significantly lower in Session 2 than in Session 1 ($p < 0.05$).

These findings suggest that participants' perception adjusted to accommodate the innovative variant by the end of the study abroad program, leading to a more successful mapping of [s̪$^{\theta}$] to /s/ instead of another fricative such as /θ/ or /f/. However, trends in confidence ratings indicate that participants were less certain with /s/ words overall in Session 2 after being exposed to variation in Cádiz, despite being more accurate, and that they were perhaps unaware of their initial difficulty in Session 1 with correctly mapping the [s̪$^{\theta}$] variant.

In conclusion, the current study demonstrates that short-term study abroad programs can lead to significant perceptual accommodation and changes in learners' perceived confidence levels, even when the variants of study are innovative and not explicitly taught in the classroom. Learners' positive attitudes towards the region of Andalusia, frequent interactions with locals, and entering Cádiz with at least an upper-intermediate level of Spanish likely also support the perceptual gains.

[1] Regan, B. (2022). Individual differences in the acquisition of language-specific and dialect-specific allophones of intervocalic /d/ by L2 and heritage Spanish speakers studying abroad in Sevilla. *Studies in Second Language Acquisition*, 1 – 28.
[2] Schmidt, L. B. (2018). L2 development of perceptual categorization of dialectal sounds. *Studies in Second Language Acquisition 40*, 857 – 882.
[3] Escalante, C. (2018). ¡Ya pué[h]! Perception of coda -/s/ weakening among L2 and heritage speakers in Ecuador. *EuroAmerican Journal of Applied Linguistics and Languages 5*, 1 – 26.
[4] Penny, R., & Sánchez-Méndez, J. (2004). *Variation and change in Spanish*. Madrid: Gredos.
[5] Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics 28*(4), 267 – 283.
[6] Villena Ponsoda, J. A. (2008). Sociolinguistic patterns of Andalusian Spanish. *International Journal of the Sociology of Language 193*, 139 – 160.

Subphonemic detail, resyllabification, and Spanish speech segmentation

Helena Escalera, Miquel Simonet
University of Arizona

In speech, words flow from one to another without pauses that delimit where they begin or end. In Spanish, words can both start and end with /s/, and they can also start and end with a vowel. This creates contexts that would seem to lead to ambiguity: *dices eso* and *dice seso*, for instance, could be homophonous, /diθeseso/. Furthermore, according to some phonological accounts, word-final consonants are resyllabified with the following word if the latter starts with a vowel (e.g., [1]). A strong interpretation of this claim predicts that the word-final /s/ in *dices eso* has identical phonetic characteristics as the word-initial /s/ in *dice seso*. Are these sequences indeed phonetically identical? How does word structure affect phonetic detail? If phonetic differences exist, are they exploited by listeners? Do fine acoustic differences, if found, impact speech comprehension, including word segmentation?

Speakers have been found to produce fine phonetic differences in their pronunciation of consonants preceding or following word junctures ([2]–[5]), and research suggests that listeners capitalize on these cues for disambiguation ([2], [6]). The present study revisits this theme and includes four experiments, two of them on speech segmentation.

A read-aloud production task with 10 L1 Spanish speakers from Spain focusing on ambiguous utterances examined whether different word affiliations led to phonetic differences in production. Speech recordings were acoustically analyzed for /s/ duration, /s/ spectral center of gravity, and vowel duration. Results revealed that /s/ duration was a reliable predictor of word affiliation, with word-initial /s/ being longer than word-final /s/. To investigate whether listeners exploit such duration differences, perception data were collected via a two-alternative forced-choice task (2AFC). In each trial in this task, participants (60 L1 Spanish listeners from Spain) were auditorily presented with one auditory stimulus ([diθes.eso]) and two written options (e.g., *dices eso* or *dice seso*). The findings confirmed that listeners' choices, while displaying low accuracy, were significantly above chance.

The remaining two experiments focus on speech segmentation. Sixty listeners participated in a word-monitoring task with the same materials as in the perception experiment above. In each trial, before playing the auditory stimuli, listeners were shown a word to monitor for (<seso>, <eso>, etc.) and they were instructed to press the space bar as soon as they recognized the word in the auditory stimulus, if at all present. Listeners were randomly played either a matching stimulus (<eso> [diθes.eso]), a mismatching stimulus (<eso> [diθe.seso]), or a distractor (<eso> [diθes.oso]). This was a go-nogo task for which we measured response times and accuracy. As expected, participants were found to react to both matching and mismatching trials (but not distractors). Most importantly, however, they were significantly faster when responding to matching than mismatching trials. This finding suggests that listeners exploit phonetic differences in the signal not only in perceptual identification but also speech segmentation. The last experiment was a fragment-monitoring task in which auditory stimuli had been cut to render them short and meaningless (e.g., rather than [diθes.eso], we had [θes.eso]), and participants were instructed to monitor for syllable-size fragments rather than word-sized fragments (e.g., <se>). Listeners accurately responded to matching and mismatching trials, but not distractors. However, there were no response-time differences between matching and mismatching trials in this experiment.

Evidently, listeners' phonological knowledge is lexically based, and their exploitation of subphonemic detail during speech comprehension is optimized for word-based speech segmentation. We conclude with a discussion of the theoretical relevance of our findings.

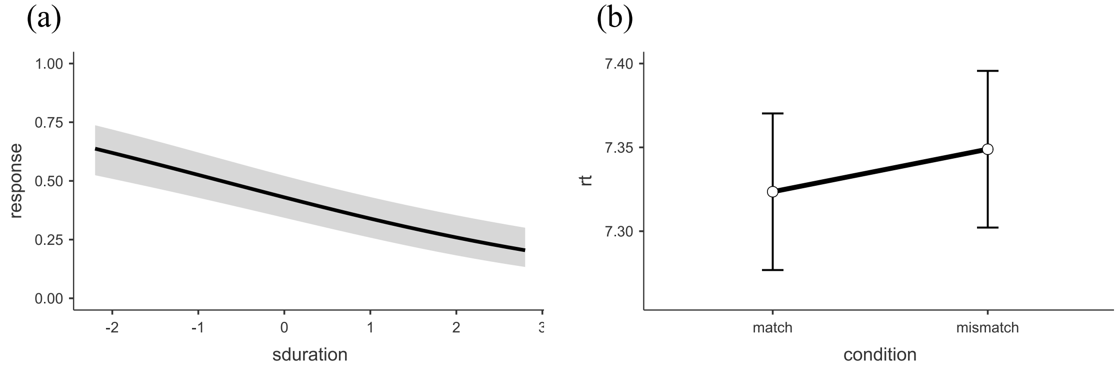


Figure 1. *(a) Regression between duration of /s/ (z-scored) and probability of responding "word-initial vowel" (that is, word-final /s/), from experiment 2 (two-alternative forced-choice perceptual identification); and (b) estimated marginal means (means and 95% CI) of multilevel regression model with reaction times as response and condition (matching, mismatching) as fixed predictors, from experiment 3 (word-monitoring task).*

[1] S. Colina, *Spanish Phonology: A Syllabic Perspective*. Washington, DC: Georgetown University Press, 2009.
[2] K. Ito and W. Strange, "Perception of allophonic cues to English word boundaries by Japanese second language learners of English," *J. Acoust. Soc. Am.*, vol. 125, pp. 2348–2360, 2009.
[3] P. Strycharczuk and M. Kohlberger, "Resyllabification reconsidered: On the durational properties of word-final /s/ in Spanish," *Lab. Phonol.*, vol. 7, p. 19, 2016.
[4] M. Jiménez-Bravo and J. M. Lahoz-Bengoechea, "Durational cues to resyllabification in Spanish," *Loquens*, vol. 10, no. 1–2, p. e099, 2023.
[5] J. M. Lahoz-Bengoechea and M. Jiménez-Bravo, "Ambisyllabic characteristics of Spanish resyllabification: Beyond durational cues," *EFE*, vol. 33, pp. 153–168, 2024.
[6] J. M. Lahoz-Bengoechea and M. Jiménez-Bravo, "Spoken word boundary detection in ambiguous resyllabification contexts in Spanish," vol. under review, 2021.

Preaspiration and the acoustics of Florentine Italian stop voicing and gemination

Chiara Repetti-Ludlow
Carnegie Mellon University

Introduction. Italian has both a voicing and a gemination contrast with stop consonants. While durational acoustic cues to these contrasts have been well documented (cf. Esposito and Di Benedetto, 1999; Kirby and Ladd, 2016), a growing body of literature suggests that additional cues may also play a role due to the presence of breathiness (Stevens and Reubold, 2014; Marotta, 2023). Specifically, several studies have noted preaspiration (also referred to as spread glottis noise) preceding voiceless geminate stops (Chasaide and Gobl, 1993; Stevens and Hajek, 2004). However, it is unclear whether this preaspiration is a correlate of voicing or gemination (Dian et al., 2023). The present study therefore employs both traditional durational cues *and* voice quality cues in order to investigate whether preaspiration is correlated with one of these contrasts. By considering these additional parameters, this research provides a more comprehensive description of Italian stops and explores why preaspiration may be arising.

Methods. A production study was carried out with 25 native speakers of Florentine Italian. Speakers read sentences with target words containing post-tonic intervocalic stops: [p, p:, b, b:, t, t:, d, d:, k, k:, g, g:]. After data was annotated in Praat, durational cues (constriction duration, preceding vowel duration, and C/V ratio) and voice quality cues (H1*-H2*, f0, and CPP) were obtained using VoiceSauce. Next, a linear discriminant analysis (LDA) was used to assess how these cues are employed by a computational model to distinguish stop voicing and gemination. Finally, linear mixed-effects models were used to test significance of these results. Further analyses accounted for potential effects of word frequency and lenition.

Results. The LDA demonstrated that both durational and voice quality cues are crucial in distinguishing stop categories, with a classification accuracy of 79.8%, although constriction duration had a higher importance in distinguishing gemination, while CPP had a higher importance in distinguishing voicing. This can be seen in Figure 1. Linear mixed-effects models expanded on these results, showing that constriction duration and C/V ratio are important in distinguishing gemination, while H1*-H2*, and CPP are important in distinguishing voicing. This suggests that preaspiration is correlated with voicing rather than gemination, as time course data shows there is increased breathiness (higher H1*-H2*) and noise (lower CPP) preceding voiceless stops but not preceding geminate stops.

Discussion. The findings of this study contribute to our understanding of how Florentine Italian stop contrasts are realized by considering voice quality cues in addition to durational cues. While only durational cues (constriction duration and C/V ratio) are relevant to gemination, voice quality cues (CPP and H1*-H2*) are only relevant for voicing. These results support an analysis in which voiceless stops are produced with greater glottal opening, leading to preaspiration before the stop closure, which could represent an anticipatory strategy for cutting off carryover voicing. That is to say, given the relevance of constriction duration to the voicing contrast, speakers might employ a spread glottis gesture during the preceding vowel to ensure that voicing not carry over into the closure.

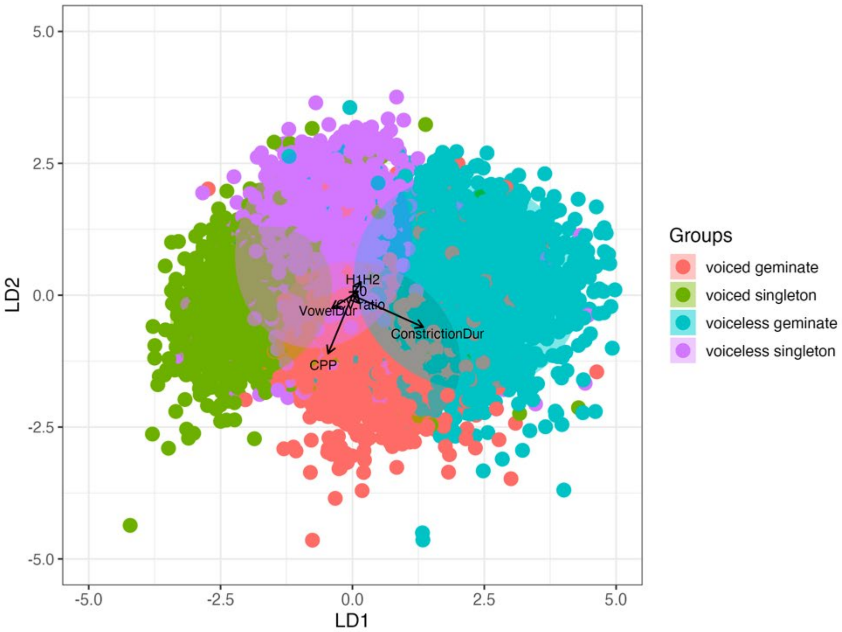


Figure 1. *Bi-plot showing distribution of tokens along the LD1 and LD2, with arrows reflecting acoustic correlates.*

Chasaide, Ailbhe N. & Christer Gobl. 1993. Contextual Variation of the Vowel Voice Source as a Function of Adjacent Consonants. Language and Speech 36(2-3). 303–330.

Dian, Angelo, John Hajek & Janet Fletcher. 2023. Preaspiration in Italian voiceless geminate and singleton stops. In International Congress of Phonetic Sciences (ICPhS), 888–892.

Esposito, Anna, & Maria Grazia Di Benedetto. 1999. Acoustical and perceptual study of gemination in Italian stops. Journal of the Acoustical Society of America, 106(4), 2051–2062.

Kirby, James & Robert D. Ladd. 2016. Stop voicing and closure duration in Italian: A cross-dialectal study. *Journal of Phonetics*, 55, 152-169. doi:10.1016/j.wocn.2015.11.003.

Marotta, Giovanna. 2023. Sulla preaspirazione in Toscano. Studi sul parlato. In onore di Renata Savy 189–203.

Stevens, Mary & John Hajek. 2004a. Comparing voiced and voiceless geminates in Sienese Italian: what role does preaspiration play? In Proceedings of the 10th Australian international conference on speech science and technology. Sydney, vol. 340, 340–345.

Stevens, Mary & Ulrich Reubold. 2014. Pre-aspiration, quantity, and sound change. Laboratory Phonology 5(4), 455-488.

Consonant voicing and quantity effects on vowel f0: A corpus study of Hungarian stops

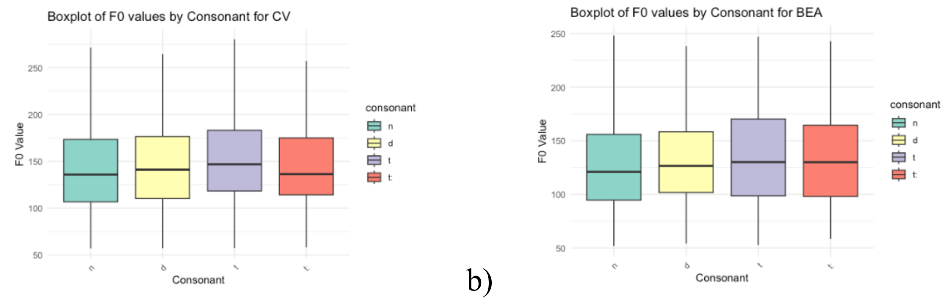
Jiachen Ma, Bogdan Pricop, Eleanor Chodroff
University of Zurich

The effects of consonant voicing on the f0 of following vowels have been widely documented across many languages and with varying effect sizes [1, 2, 3] (henceforth, CF0). Though the source of the effect has been widely debated, the effect size can vary across languages [1]. In addition, consonant gemination has also been shown to influence onset f0, in which higher mean f0 values are observed following geminates compared to singletons, regardless of voicing or manner of articulation [2, 4], though a few nuanced exceptions have been observed [5, 6]. This raises the question of how consonant voicing and quantity respectively influence following f0. Maspong et al. [2] investigated this question for Italian, finding main effects, but no interaction of voicing and quantity on f0 variability. Given the complex and often inconsistent effects of gemination on f0, it is worth examining whether similar patterns hold in other languages. Hungarian, a non-Indo-European language, has been less extensively studied in this regard and has both a voicing and geminate contrast in the alveolar stop series. Using two large-scale spoken corpora of Hungarian, we investigated the effects of consonant voicing and quantity on onset f0 within the alveolar stop series. Based on previous studies, we expected a higher onset f0 following voiceless than voiced stops and following geminate than singleton stops. We additionally tested the relative raising or lowering of f0 following oral stops relative to nasal stops, where f0 was not expected to be perturbed [7].

Methods. The study employed the Mozilla Common Voice Corpus [8] (CV) and the BEA – Hungarian Spontaneous Speech Database [9], each with over 70 hours of speech. Phonetic transcriptions were generated from the orthographic transcript using the XPF Corpus [10] and aligned to the audio using the Montreal Forced Aligner [11]. We extracted f0 using a two-pass approach in V_1CV_2 sequences from the onset of the V_2 force-aligned boundary. V_1 was any vowel, C was one of /n d t tː/ and V_2 was one of /i a u y/ or their long counterpart. Tokens were retained if the utterance contained more than two words, V_2 was between 50 and 1000 ms, the onset f0 was within 2 SD of the speaker-specific mean, and the speaker had 25 tokens of each stop (20 for BEA). The final analysis contained 47 speakers from CV and 22 from BEA.

Analysis and Results. A linear mixed-effect model was used to estimate onset f0 from the consonant type, vowel height, and maximal random-effect structure for speaker. Significant effects were assessed with 0.05. For each corpus, significantly higher f0s were observed following voiceless than voiced stops ($\beta_{CV} = -1.17$, $\beta_{BEA} = -1.15$). The effect of gemination was in the unexpected direction, but did not reach significance in either corpus (/t/ vs /tː/: $\beta_{CV} = 0.50$, $\beta_{BEA} = 0.52$). In addition, /n/ was not significantly different from /d/ alone ($\beta_{CV} = 2.576$, $\beta_{BEA} = 3.667$), but was significantly lower than both /t/ and /tː/ (/n/ vs /t/: $\beta_{CV} = 6.89$, $\beta_{BEA} = 7.65$; /n/ vs /tː/: $\beta_{CV} = 5.89$, $\beta_{BEA} = 6.60$). The main effect of vowel height was significant for CV but not BEA ($\beta_{CV} = 1.57$, $\beta_{BEA} = 0.26$), for CV, this effect was significantly enhanced for oral compared to nasal stops, voiceless compared to voiced stops, and geminates compared to singletons (/n/ vs /d t tː/: $\beta_{CV} = -0.34$; /d/ vs /t tː/: $\beta_{CV} = -0.73$; /t/ vs /tː/: $\beta_{CV} = -1.51$).

Discussion and Conclusion. Our results confirm CF0 effects in Hungarian across both read and spontaneous speech. However, gemination alone did not influence onset F0 but it did interact with vowel height in CV, lowering F0 with high vowels. In BEA, no effect or interaction was observed with geminate status, though figures suggest a general F0 reduction after geminates. This contrasts with previous findings such as [2] for Italian. This also suggests a hierarchy of f0 influences consistent with [5] for Italian, in which the voicing effect is larger than the quantity effect, if present. These findings suggest that differences in CF0 and geminate effects may stem from varying degrees of automaticity and speaker control. Further research is needed to explore these dynamics across languages and contexts.



a) b)

Figure 1. *Mean onset f0s (Hz) following each target consonant for a) the CV and b) BEA corpora.*

[1] Ting, C., Clayards, M. Sonderegger, M., and McAuliffe, M. 2025. The cross-linguistic distribution of vowel and consonant intrinsic F0 effects. *Language*.

[2] Maspong, S., Burroni, F., and Kirby, J. 2024. CF0 effect and articulatory strength of geminate consonants. In *Proceedings of Speech Prosody 2024*, pages 334–338.

[3] Pricop, B. and Chodroff, E. 2024. Consonant f0 effects: A case study on Catalan. In *Proceedings of Speech Prosody 2024*, pages 329–333, Leiden, The Netherlands, 2–5 July.

[4] Hamzah, M. H., Hajek, J., and Fletcher, J. 2020. Non-durational acoustic correlates of word-initial consonant gemination in Kelantan Malay: The potential roles of amplitude and f0. *Journal of the International Phonetic Association*, 50(1), 23–60.

[5] Di Benedetto, M. G., and De Nardis, L. 2021. Consonant gemination in Italian: The nasal and liquid case. *Speech Communication*, 133:62–80.

[6] Ridouane, R. 2007. Gemination in Tashlhiyt Berber: an acoustic and articulatory study. *Journal of the International Phonetic Association*, 37(2), 119–142.

[7] Hanson, H. M. 2009. "Effects of obstruent consonants on fundamental frequency at vowel onset in English." *Journal of the Acoustical Society of America,* 125: 425–441.

[8] Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., and Weber, G. (2020). Common Voice: A Massively-Multilingual Speech Corpus. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)* (pp. 4218–4222). European Language Resources Association (ELRA), Marseille, France.

[9] BEA Hungarian spontaneous speech database. (2013, January 1). Version 1.0.0 (automatically assigned). Dataset (Text and Audio Corpus). European Language Grid. Retrieved from https://live.european-language-grid.eu/catalogue/corpus/13040.

[10] Cohen Priva, U., Strand, E., Yang, S., Mizgerd, W., Creighton, A., Bai, J., Mathew, R., Shao, A., Schuster, J., and Wiepert, D. *The Cross-linguistic Phonological Frequencies (XPF) Corpus Manual*.

[11] McAuliffe, M., Socolof, M., Stengel-Eskin, E., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). Montreal Forced Aligner [Computer program]. Version 1.0.0. Retrieved May 5, 2017, from http://montrealcorpustools.github.io/Montreal-Forced-Aligner/.

Does VtoV coarticulation always require "look-ahead"? Evidence from an EMA and acoustic study of Campanian Italian

Pia Greca, Francesco Burroni, Jonathan Harrington
Institute for Phonetics and Speech Processing, LMU Munich, Germany

One of the most pervasive phenomena in speech is coarticulation: the fact that neighboring sounds influence one another. Two main models for coarticulation have been proposed: coproduction and look-ahead ([1], [2], [3]). Coproduction models hold that coarticulation is strictly rooted in local overlap and blending of articulatory gestures ([1]). Look-ahead models hold that coarticulation can be long-distance, with non-neighboring segments influencing each other ([3], [4]). A coarticulatory phenomenon that has attracted much attention to test the predictions of these two models is anticipatory trans-consonantal Vowel-to-Vowel (VtoV) coarticulation ([1],[4],[5],[6]). VtoV coarticulation can also be phonologized into metaphonic effects, which are, for instance, largely attested in Italo-Romance varieties ([8], [9]). These are particularly of interest because they suggest that, at least at diachronic timescales, VtoV coarticulation can lead to long-distance sound changes. The question we examine is whether synchronic coarticulation already betrays the hallmarks of these long-distance phenomena in an acoustic and electromagnetic articulography (EMA) study of VtoV coarticulation in Campanian Italian. **Methods.** ***Participants***. We collected data from 10 Campanian Italian speakers (6M, 4F, with a target of 15 for the conference). Beyond Italian, they all have knowledge of their local dialect. We focused on Italian produced by Campanian speakers because their native dialect displays phonologized metaphony ([10]). Given the lack of experimental work on VtoV effects in Italian, we considered Campanian speakers a good testing ground. ***Items.*** We elicited disyllabic words of the structure ˈ$CV_1C(:)V_2$, in which V was either /a, e, i, o/, while C was either bilabial, labiodental, alveolar, or nasal, and either singleton or geminate. We excluded velars because they can be allophonically palatalized before front vowels ([11]). The total of tokens collected was: 374 words x 10 speakers = 3740. ***Procedure***. Target items were produced in a carrier sentence. We collected simultaneous EMA and audio data using a three-dimensional Carstens AG501 system sampling at 1250 Hz. 10 sensors were used to track tongue, lips, and jaw movement. Since we focused on vowel production, we analyzed the vertical and horizontal dimensions of the tongue dorsum sensor (TD_y and TD_x) which reflect vowel height and frontness respectively. For the acoustic signals, we conducted formant analyses as coarse correlates of vowel height (F1) and frontness (F2). ***Analyses.*** Separately by stem vowel, we conducted mixed effect regression analyses at different time points of the stem vowel (V_1) to check when a significant effect as a function of the suffix vowel (V_2) is observed. Additionally, we also tested whether different consonant types affect the degree of VtoV coarticulation. **Results.** Our statistical analyses show that, at around 75-80% of V_1, V_2 starts exerting an influence on both TD_y and TD_x, indicating an anticipatory coarticulatory effect that is present in the last quarter of V_1. This pattern is robust for /a,e,o/ and weaker for /i/. We illustrate our findings with V_1= /a/ in

Figure 3A and B. In line with the EMA data, effects on acoustic formants, especially F2, are also limited to the final portion of the vowel. Finally, for most vowels, certain transvocalic consonants, specifically geminates, attenuate VtoV effects on the TD_x and TD_y position,

Figure 3C. **Conclusion.** To sum up, our findings indicate that, in terms of temporal extent, VtoV effects are limited to the last quarter of V_1 and get stronger the closer to V_2. Thus, VtoV coarticulation is local more than long-distance. Second, our findings also suggest that VtoV coarticulation is modulated by intervening consonants. Combined together, these two findings support the notion that in Campanian Italian VtoV effects are in agreement with the predictions of a coproduction model supplemented with a weak anticipatory gestural field ([1]) more than with those of look-ahead models. If this is correct, sound changes like metaphony, albeit originating in VtoV effects, need not to be the direct result of look-ahead coarticulation. Rather, local coarticulation, in the course of phonologization, needs to be magnified by additional mechanisms, such as e.g., perceptual reparsing, temporal realignment of gestural activation windows ([12]).

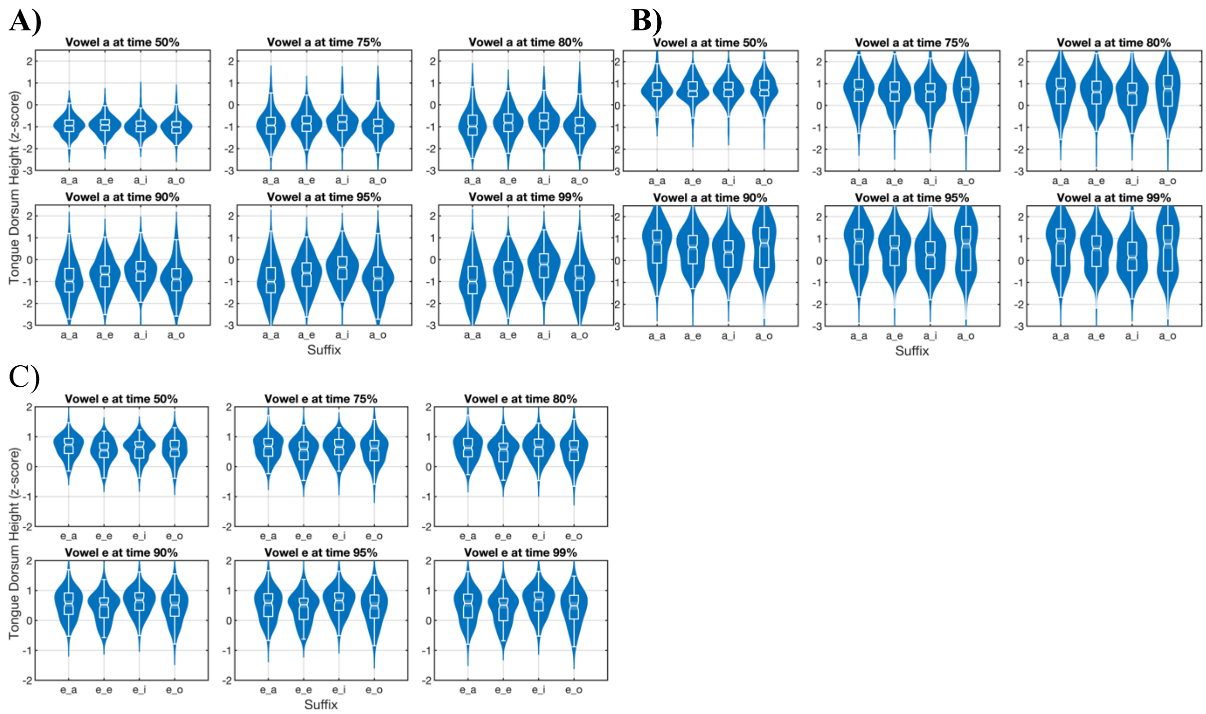


Figure 3. *A): Tongue Dorsum Height, and B): Tongue Dorsum Frontness at 50%, 75% … 99% for V_1 = /a/ , as a function of different suffix vowels (V_2). The y-axis represents in A) TD_y and in B) TD_x (z-scored) for each vowel, and the x-axis different V_2 contexts, e.g., a_a, a_e, a_i etc. Notice how, progressing across panels from 50% to 100% of V_1 duration, TD_y becomes higher (i.e. more raising) and TD_x lower (i.e. more fronted) in the a_i context, while TD_y becomes lower and TD_x higher in the a_a context, thus anticipating the position for V_2. C): Same plot as A) but with V_1 = /e/ and intervening consonants being only geminates. Notice how VtoV effects are absent for the /e/ vowel even at 99% of V_1 when the intervening consonants are geminates.*

[1] C. A. Fowler and E. Saltzman, "Coordination and coarticulation in speech production," *Language and speech*, vol. 36, no. 2–3, pp. 171–195, 1993.

[2] B. Kühnert and F. Nolan, "The origin of coarticulation," *Coarticulation: Theory, data and techniques*, pp. 7–30, 1999.

[3] B. H. Repp, "Some observations on the development of anticipatory coarticulation," *The Journal of the Acoustical Society of America*, vol. 79, no. 5, pp. 1616–1619, 1986.

[4] W. L. Henke, "Dynamic articulatory model of speech production using computer simulation.," 1966.

[5] D. H. Whalen, "Coarticulation is largely planned," *Journal of Phonetics*, vol. 18, no. 1, pp. 3–35, 1990.

[6] D. Recasens, "Long range coarticulation effects for tongue dorsum contact in VCVCV sequences," *Speech Communication*, vol. 8, no. 4, pp. 293–307, 1989.

[7] S. E. Öhman, “Coarticulation in VCV utterances: Spectrographic measurements,” *The Journal of the Acoustical Society of America*, vol. 39, no. 1, pp. 151–168, 1966.
[8] F. Torres-Tamarit, K. Linke, and M. van Oostendorp, *Approaches to Metaphony in the Languages of Italy*. De Gruyter Mouton Berlin, 2016.
[9] M. Maiden, Interactive morphonology: metaphony in Italy. Routledge, 2018.
[10] M. Loporcaro, “Metaphony and diphthongization in Southern Italy: reconstructive implications for sound change in early Romance,” *Approaches to Metaphony in the Languages of Italy*. Berlin: De Gruyter, pp. 55–87, 2016.
[11] D. Recasens, “Patterns of VCV coarticulatory direction according to the DAC model,” *Segmental and prosodic issues in Romance phonology*, vol. 282, p. 25, 2007.
[12] P. S. Beddor, “A coarticulatory path to sound change,” *Language*, pp. 785–821, 2009.

Oral Session 7: Information structure and prominence effects

Positional and stress effects on vowel glottalization in Tahitian

Janet Fletcher, Adele Gregory
University of Melbourne, Australia

It is well accepted that in many languages, phonemic /ʔ/ has multiple variants ranging from a canonical voiceless stop with full glottal occlusion to more vowel-like non-modal voiced variants ([1],[2]). A preliminary acoustic phonetic study of glottal stop variation in Tahitian, an Eastern Polynesian language spoken in French Polynesia ([3]), showed that the language has glottal variants ranging from a fully or partially-occluded "canonical" glottal stop to creaky voiced, breathy voiced or modal voiced allophones ([4]). However, compared to the closely-related Eastern Polynesian language Hawaiian, the proportion of canonical glottal stops is somewhat higher at 68% compared with 7% reported in [5] [6]. In general, glottal stops in Tahitian tend to be more "consonantal" rather than "vocalic".

The current study extends [4] and examines vowel glottalization in the vicinity of phonemic glottal stops in Tahitian. In Hawaiian, key variants identified by [5] include V1 and V2 glottalized variants where vowels either side of a phonemic glottal stop show extensive glottalization. This was not investigated explicitly in [4]. Additionally, we focus here on two important factors, lexical stress and word-context. Our interest was to see if stressed vowels in syllables with glottal-stop onsets show less coarticulation than unstressed vowels suggesting prosodic strengthening. Further we examined whether vowels following glottal stops in word initial position similarly show reduced levels of glottalization given many glottal stops in this context are produced with full glottal closure [4]. Finally we examined vowels following and preceding glottal stops in medial contexts to monitor the extent of carryover or anticipatory coarticulation and any potential interaction with stress.

Five female speakers in Papeete, French Polynesia were recorded performing two controlled reading tasks with tokens placed in narrow focus position and a third task which included multiple readings of La Bise et Le Soleil (the north wind and the sun) [8]. All files were coded for stress (accented and unaccented) following [9]. Various acoustic parameters were extracted for 9,984 vowels that occurred either side of glottal and oral stops including voicing fraction, Harmonic-to-Noise-Ratio (HNR15) and spectral tilt (H1*-H2*) which are typically used in analyses of creaky voice[1,2,7]. The first parameter was analysed separately for preceding and following vowels using *lmerTest* with fixed factors stress and position. The remaining parameters were submitted to separate generalised additive mixed models (GAMM) which included parametric difference terms and smooths for stop-type (oral versus glottal), word position, stress, vowel position, and random smooths by participant and word and experimental task.

For voiced fraction (Figure 1), vowels that precede and follow /ʔ/ show effects of stress with lower and more variable percentages of voicing in unaccented versus accented vowels ($p<0.0001$) suggesting more glottalization. A subsequent analysis showed that word initial position promotes higher voicing percentage values in following vowels regardless of stress, with differences more evident in medial contexts. Figure 2 compares the time course of spectral noise (HNR15) and spectral tilt for vowels flanking glottal stops and oral stops. Vowels in glottal environments have lower overall spectral noise values than in oral stop environments suggesting lower levels of periodicity and glottalization. Unaccented vowels have significantly lower spectral noise values compared to accented vowels in both contexts ($p<0.0001$), with lower periodicity in preceding vowels. For spectral tilt, vowels have significantly lower values in glottal compared to oral contexts regardless of stress ($p<0.001$), with lower values in following vowels. In general the data clearly show evidence of glottal coarticulation in either direction, with two of the three parameters also suggesting more vowel glottalization in unstressed versus stressed syllables.

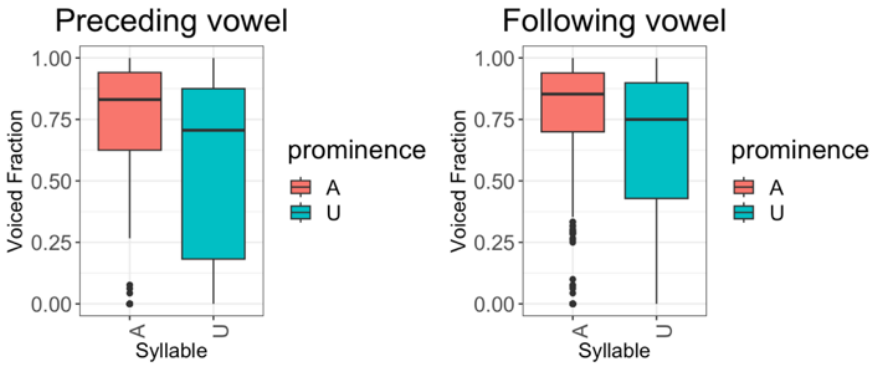


Figure 1. *Voiced Fraction for vowels preceding or following /ʔ/ plotted by stress (A for accented, U for unaccented)*

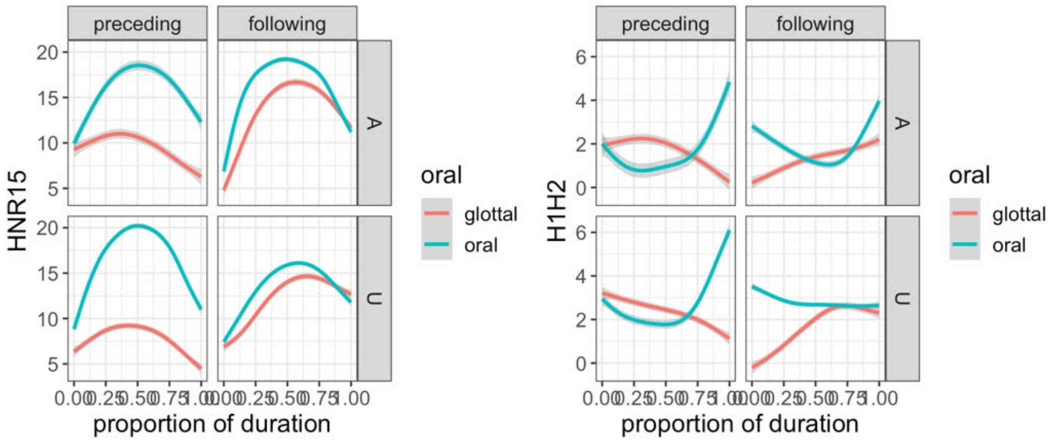


Figure 2. *Time course of Harmonic to Noise ratio (left) and Spectral Tilt (H1*-H2*) (right) for vowels either side of /ʔ/ and oral stops, plotted according to stress (A for accented, U for unaccented)*

[1] Garellek , M, Chai, Y., Huang, Y., and van Doren, M. (2021). Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association* 53, 305-332.
[2] Keating, P., Kuang, J., Garellek, M., Esposito, C.M., & Khan, S.D. (2023). A cross-language acoustic space for vocalic phonation distinctions: *Language 99*(2), 351-389.
[3] Lemaître, Y. (1973) *Lexique due Tahitien contemporain.* Paris Bureau de la Recherche Scientifique et Technique Outre-Mer.
[4] Fletcher, J. & Gregory, A. (2024). Speech task and prosodic context influence glottal stop variation in Tahitian. In Olga Maxwell & Rikke Bundgaard-Nielsen (Eds.), Proceedings of the Nineteenth Australasian International Conference on Speech Science and Technology (pp. 162-166). Australasian Speech Science and Technology Association.
[5] Davidson, L. (2021). Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian. *Journal of Phonetics* 88, 1-14.
[6] Davidson, L. and Parker Jones, O. (2023). Word-level prosodic and metrical influences on Hawaiian glottal stop realization. *Phonetica* 80 (3-4), 225-258.

[7] Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics*, 45, 106-113.
[8] Boula de Mareüil, P. Vernier F., Adda G., Rilliard, A, Vernaudon, J. (2019). A speaking atlas of indigenous languages of France and its Overseas Territories. *International Conference Language Technologies for All* (LT4All), Paris, 155–159.
[9] Bickmore, L. (1995). Refining and formalizing the Tahitian stress placement algorithm. *Oceanic Linguistics*, 34, 410-442.

The information status of locally decelerated content words in Czech newsreading

Michaela Svatošová, Jan Volín
[1]*Institute of Phonetics, Faculty of Arts, Charles University, Prague*

Introduction. Information structure of utterances is thought to be most prominently marked by melodic patterns, however, other prosodic domains might serve this purpose as well. The present study therefore explored the relation of the articulation rate variations to information structure. Speech tempo is often referred to on a global scale, e.g. stating the speaker's mean articulation rate in a given recording or phrase. Unfortunately, this averaging conceals the local differences in articulation rate that occur inside utterances [1], while these might convey affective and discourse meanings [2] (cf. a mean F0 value vs. a nuclear pattern).

Method. In order to investigate temporal changes in utterances, we designed a metric for measuring the local articulation rate (LAR), which normalizes for inherent durational properties of phones. This approach relates the observed duration of each phone to its inherent duration and thus quantifies the extent to which each phone is accelerated or decelerated. (The inherent phone durations were obtained from a corpus of continuous speech containing over 170 000 manually annotated phones from 32 speakers.) Figure 1 illustrates the difference between simple rate values (inverted phone durations) and the LAR values. Since durations of phones are also affected by phones adjacent to them, the metric was developed by accounting for the inherent durations of 173 most frequent phone bigrams in Czech. This results in a smoother and possibly more appropriate temporal contour (thick line in panel B of Figure 1).

Analysis. In this study, the LAR metric was used as a tool for identifying locally decelerated content words in recordings of 16 radio news presenters. The material represents a real-life communicative situation, in which professional speakers intend to be intelligible while informing the public about current events. The recordings amounted to almost 120 minutes of speech, but the analysis was restricted to 1 304 major prosodic phrases which were not divided into minor prosodic phrases and contained three to eight stress-groups. We calculated the local articulation rate (LAR) contours of these phrases and smoothed them with a 3-point moving average to reduce the effect of outliers. Mean LAR values were obtained for individual words. Subsequently, we searched for substantially decelerated content words, here defined as the slowest non-final word of the phrase with the mean local articulation rate of 80% or less in relation to the average LAR of the phrase (excluding the final stress-group). The information structure of utterances containing the target words was annotated and discussed by the authors.

Results and discussion. The analysis provided 43 markedly decelerated words. They were most frequently located in the initial or penultimate stress-group. Nearly 80% of these words occurred in prominence-lending roles, forming a part of the rheme (72%) or contrastive topic (7%). Figure 2 shows two such examples. The target words represented various word classes, nouns being the most dominant category (37% of cases, many of which were proper nouns).

The majority of the target words were monosyllabic (30%) or disyllabic (58%). Since content words in Czech tend to be polysyllabic, it might be useful to further normalise the LAR metric for the number of syllables in each word, which has been described as one of the factors influencing the duration of phones [3, p. 96]. Moreover, changes of articulation rate also reflect aspects of speech production other than signalling the functional prominence described above. For instance, some of the most decelerated words were likely slowed down for rhythmical reasons – in order to resolve a stress-clash (two accented syllables standing next to each other).

Conclusion. The results support the connection between variations of articulation rate and information structure of utterances (tested by [4], [5]). In future studies, LAR contours could be analysed jointly with F0 contours to evaluate if these two domains function together or independently. Furthermore, perception experiments using the described phrases could provide insight into the perceptual significance of these spontaneously produced local decelerations.

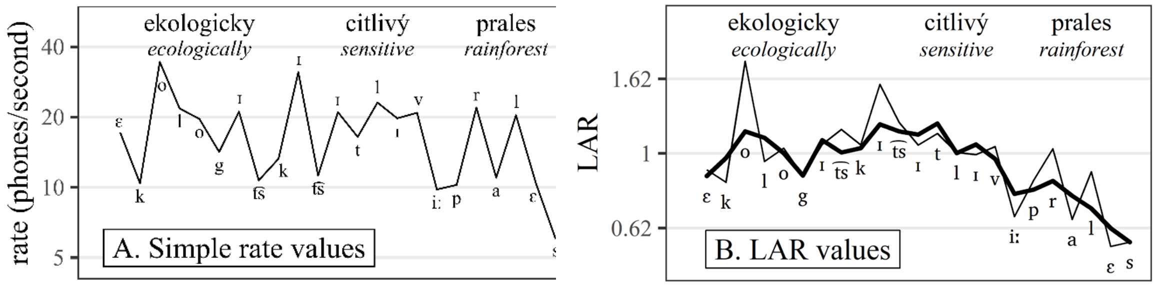


Figure 1. *Comparison of the simple rate values (inverted phone durations, panel A) and the local articulation rate (LAR) metric (panel B). In panel B, the thin line represents the values normalised only for inherent phone durations and the thick line represents the values normalised also for the inherent durations of the most frequent phone bigrams.*

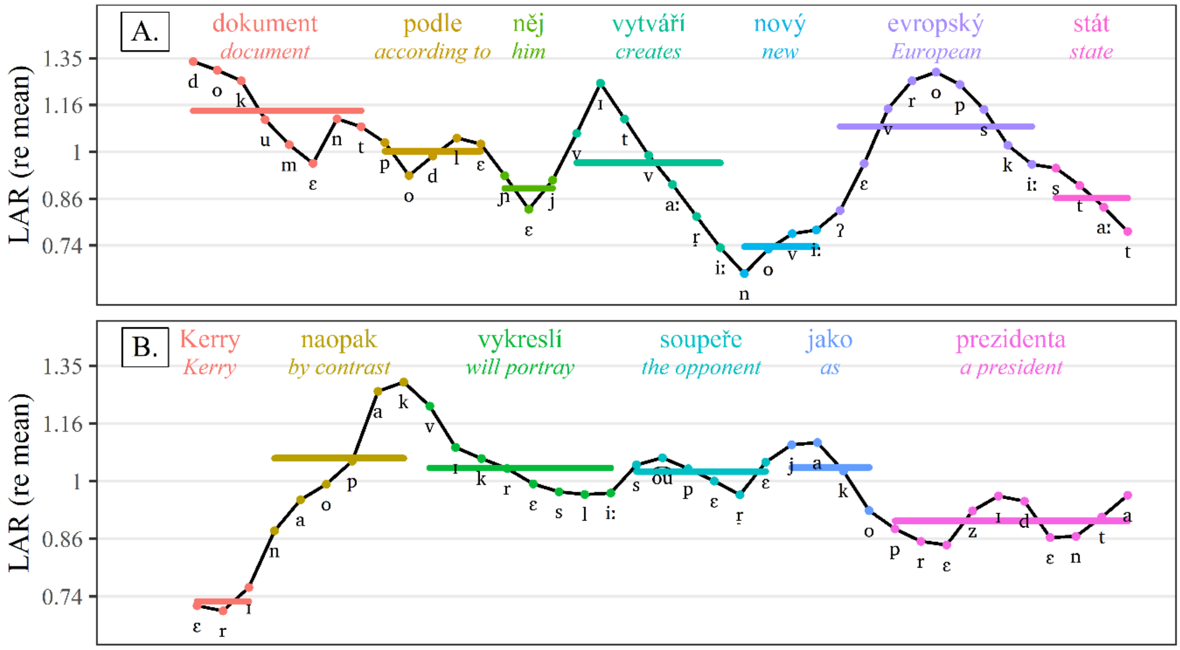


Figure 2. *The temporal contours of phrases containing substantially decelerated words. The black lines represent the LAR values smoothed with a 3-point moving average and the coloured lines correspond to mean LAR values in individual words. – A. "According to him, the document creates a new European state." The decelerated word "nový (new)" is part of the rheme. – B. "By contrast, Kerry will portray the opponent as a president [who makes bad decisions…]." The decelerated word "Kerry" forms a contrastive topic.*

This work was financially supported by Charles University Grant Agency, project no. 32424.

[1] Miller, J. L., Grosjean, F., & Lomanto, C. 1984. Articulation Rate and Its Variability in Spontaneous Speech: A Reanalysis and Some Implications, *Phonetica* 41 (4), pp. 215–225.

[2] Uhmann, S. 1992. Contextualizing Relevance: On Some Forms and Functions of Speech Rate Changes in Everyday Conversation, in *Pragmatics & Beyond New Series* 22, P. Auer and A. Di Luzio (Eds.) Amsterdam: John Benjamins Publishing Company, 297–336.

[3] Dankovičová, J. 2001. *The Linguistic Basis of Articulation Rate Variation in Czech*. in Forum Phoneticum. Frankfurt am Main: Hector.

[4] Baumann, S., Becker, J., Grice, M., & Mücke, D. 2007. Tonal and articulatory marking of focus in German, in *Proceedings of the XVIth ICPhS*, 1029–1032.
[5] Kügler, F. 2008. The role of duration as a phonetic correlate of focus, in *Proceedings of the 4th Conference on Speech Prosody*, 591–594.

Variation in French *continuation* rises: Is there a tone on extrametrical schwa?

Cristel Portes[1], Emilie Marty[2], Caterina Petrone[1], Martine Grice[3]
[1]Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France, [2]Université de Tours, INSERM, iBraiN U1253, 37032, Tours, France, [3]University of Cologne, Cologne, Germany

In French, continuation (e.g. indicating an unfinished turn) is mostly conveyed by a simple rise occurring at the end of a prosodic phrase. In AM approaches, it is generally modelled as three consecutive tones on the final tonic syllable: a bitonal pitch accent, LH*, and an edge tone, H- or H% (Figure 1). However, when the final syllable ends in a coda consonant, an extrametrical schwa can occur in phrase-final position. In this case, many realisations of the continuation contour show a rapid fall in f0 on this final schwa (Figure 2), which is clearly audible and thus often annotated with a final L%. In this paper, we provide distributional evidence for treating this final fall as phonetic, and thus analyse these continuation contours with a final fall on schwa nonetheless as LH*H%. This analysis is only possible if the H% is associated with the tonic syllable rather than simply with the edge.

As early as in 1685, French grammarians have analysed the optional realisation of a schwa after a final coda consonant in French like in the word "animal" pronounced [animalə] as the reflection of the release of a final coda consonant and a way to render it more sonorous [1]. More recently, a reinforcement and generalisation of the production of this final post-tonic schwa has been observed, even in Parisian French where it was considered to have almost disappeared [2, 3, 4]. We follow [4] in considering this final post tonic schwa as the "grammaticalisation of articulatory relaxation at the end of a major prosodic unit". [4] also shows that French listeners have no difficulty identifying the continuation contour despite the f0 drop on the final schwa syllable. This is surprising, because the French intonational inventory includes a rise-fall on the last tonic syllable, tonally analysed as LH*L%, that could be confused with the schwa realisation of the continuation rise. However, this rise-fall has been shown to be aligned earlier than the continuation rise and to convey a quite different meaning, i.e. of conviction, or obviousness [5].

In this pilot study, we compare IP final realisation of the continuation contour with and without schwa. We used an existing prosodically annotated corpus [6] consisting of 36 scenarios about everyday news, drawn from professional and institutional contexts, each read aloud by 30 French female native speakers (mean age = 38 y.o., SD = 6.5). In each scenario, we examined one IP-final intonational contour at the end of a preface where 2/3 of the final words ended in a coda consonant (henceforth coda condition), that were likely to elicit a schwa, and 1/3 ended in a vowel (henceforth, no coda condition). We focused on the items that were annotated as LH*L% and LH*H% and compared the proportion of each annotation across the coda and no coda conditions. Interestingly, we found that LH*L% was hardly ever annotated in the no coda condition (5%) while it was almost exclusively annotated on the coda items (95%). We conclude that the presence of schwa after the coda is linked to the annotation of L%. For LH*H% the distribution was more balanced, 42% of which were annotated in the no coda condition and 58% in the coda condition. We argue that the presence of both LH*H% and LH*L% in the coda condition is linked to the optionality of schwa. A preliminary perceptual analysis suggests that most of the analysed LH*L% tunes were interpretable as continuation rises, and do not appear to be functionally different from the LH*H% tunes.

From the distributional properties and preliminary perceptual analysis, we argue that the f0 fall on the schwa syllable cannot correspond to a tonal target. Consequently, at the phonological level the continuation contours with a schwa syllable can be analysed as LH*H%, regardless of their phonetic properties. To account for this contour, the H% boundary tone can be primarily associated to the penultimate accented syllable rather than to the edge of the constituent. Such a conclusion is possible in a flexible conception of the AM model whereby edge tones may acquire a secondary association to a metrically strong syllable [7].

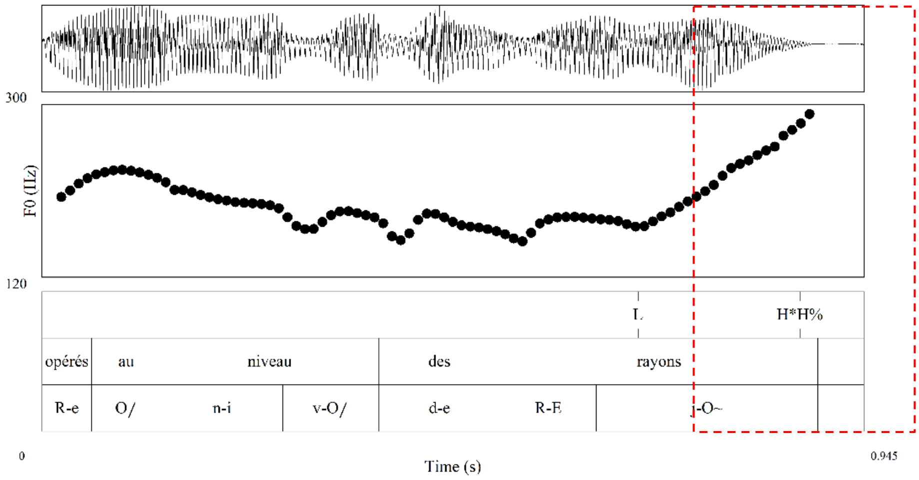


Figure 1. *Continuation rise without final schwa syllable. The dashed red line delimits the accented syllable*

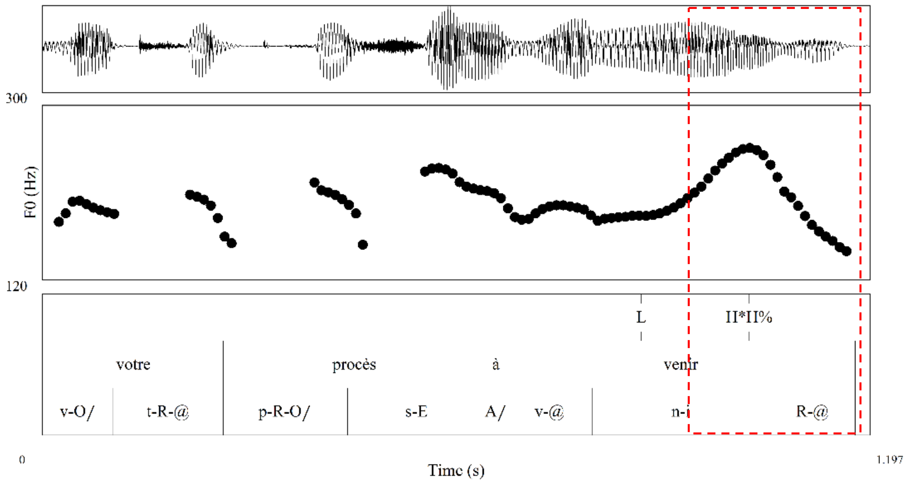


Figure 2. *Continuation rise with final schwa syllable [R-@]. The dashed red line delimits the accented syllable*

[1] Carton, F. (1999). L'épithèse vocalique en français contemporain: étude phonétique. Faits de langues, 7(13), 35-45.
[2] Fónagy, I. (1989). Le français change de visage?. *Revue Rornane*, *24*(2), J989.
[3] Hansen, A. B. (1997). Le nouveau [ə] prepausal dans le français parlé à Paris. Polyphonie pour Iván Fónagy, 173-198.
[4] Fagyal, Z. (2000). Le retour du e final en français parisien: changement phonétique conditionné par la prosodie. In Actes du XXIIe Congrès International de Linguistique et de Philologie Romanes: Vivacité et diversité de la variation linguistique (pp. 151-160). Max Neimeyer.
[5] Portes, C., & Lancia, L. (2017). Earlier or Higher? Comparing French rising-falling contour with rising contour in a corpus of conversation. *Journal of phonetics*, *63*, 35-52.
[6] Marty, E., Bertrand, R., Petrone, C., German, J. (2022) Prosodic Correlates of Discourse Structure and Emotion in Discourse Markers that Preface Announcements of News. Proc. Speech Prosody 2022, 605-609.

[7] Grice, M. (2022). Commentary on Chapter 1: Introducing Flexibility into Autosegmental-Metrical Phonology. In *Prosodic Theory and Practice*, MIT Press.

A dynamic system of informativity grades between prosody and syllable distribution

Jinhee Kwon, Johan Frid, Mikael Roll
[1]*Lund University, Sweden*

A system of optimal information transmission adheres to the principle of maximum entropy, in which all events are assigned equal probabilities if there is no reason to think otherwise, and where entropy denotes uncertainty [1]. Message-oriented phonology (MOP) proposes that a language's sound patterns are shaped by the goal of effectively transmitting meaning-bearing units, such as words and syllables, even synchronically [2]. This study begins with the idea that a phonological representation, based on a message, must operate in a non-committal default state. In this state, the members of the inventory have equal probabilities, and the system exhibits high uncertainty. This uncertainty allows the message to influence the dynamic distribution of phonological categories. If the inventory system of segments and suprasegments follows this principle, the probabilities of each category in a phonological event would be uniformly distributed.

Swedish has a quasi-distinctive pitch accent realised on the stressed syllable of each word. In South Swedish, accent 1 has an H*L contour whereas accent 2 has L*HL. Across dialectal variations, tone accent 1 is known to reduce uncertainty in words due to its relatively sparse distribution in the lexicon facilitating rapid speech processing [3–6]. This imbalance is partly due to the fact that most compound words carry accent 2. A corpus-based study on the distribution of the two tone accents found that accent 1 had a disproportionately low type frequency compared to a more balanced token frequency [7]. This resulted in a significant reduction in cohort entropy for accent 1 compared to accent 2, observed after the first vowel in words. In terms of the probability distribution in an optimal system, while accent 1 words benefit from the prosodic domain in word recognition, it can be assumed that accent 2 words exhibit a distribution more focused on segmental scope.

In this study, we explore whether accent 1 and accent 2 have different impact on lexical identification. We approach this by testing whether the principle of maximum entropy applies to the second syllable following the word-initial syllable, considering both syllables and tone accents within a shared inventory. To measure syllable likelihoods, we introduce a *syllabic transition entropy* based on the conditional probability of the second syllable given the first syllable. The effects of tone accents are quantified by calculating the entropy difference for each tone accent across all possible syllable combinations at the first syllable transition point. These probabilities are derived from a phonologically tagged lexical frequency database of Swedish, developed using a genre-balanced corpus of 50 million tokens, along with a pronunciation lexica by Nordisk språkteknologi and the Swedish Language Council [8].

Our preliminary results reveal a strong trade-off between syllable informativeness and accent 1 informativeness. The entropy reduction rate associated with accent 1 was higher for syllables with greater probability, where higher probability indicates lower informativeness in the generated model (Figure 1). This pattern does not hold for accent 2, where the prosodic contributions showed no correlation with syllabic probability (Figure 2).

Additionally, we observed a strong link between word length and syllabic transition entropy (Figure 3), suggesting that longer words convey less information at word onset. These findings support previous research that tone relevant for lexical identification [9], with a varying degree of contribution across different word-initial combinations. Future investigations into the informativeness in the shared syllabic and prosodic landscape in simplex words, as well as transitions between phonemes, could further elucidate the impact of the tone accent distinction within the segmental and suprasegmental inventory.

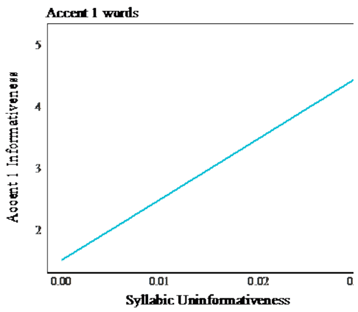


Figure 1.
Shared informativeness between syllabic and prosodic domains for accent 1 words

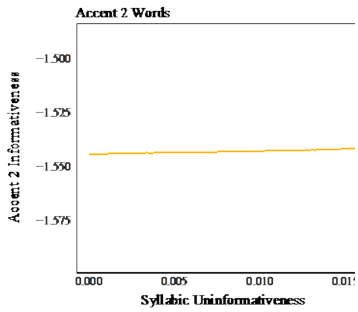


Figure 2.
Absence of prosodic contribution among accent 2 words

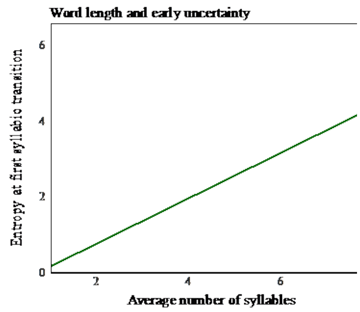


Figure 3.
Correlation between word length and the uncertainty level at the first syllabic transition

[1] Jaynes ET. 1957. Information Theory and Statistical Mechanics. *Phys. Rev.* **106** (**4**): 620–630.
[2] Hall K, Hume E, Jaeger TF, Wedel AB. 2016. The Message Shapes Phonology. https://doi.org/10.31234/osf.io/sbyqk
[3] Hjortdal A, Frid J, Novén M, Roll M. 2024. Swift Prosodic Modulation of Lexical Access: Brain Potentials From Three North Germanic Language Varieties. *J. Speech Lang. Hear. Res.* **67** (**2**): 400–414.
[4] Roll M. 2015. A neurolinguistic study of South Swedish word accents: Electrical brain potentials in nouns and verbs. *Nord. J. Linguist.* **38** (**2**): 149–162.
[5] Roll M. 2022. The predictive function of Swedish word accents. *Front. Psychol.* **13**: 910787.
[6] Söderström P, Horne M, Roll M. 2017. Stem Tones Pre-activate Suffixes in the Brain. *J. Psycholinguist. Res.* **46** (**2**): 271–280.
[7] Kwon J, Frid J, Söderström P, Roll M. Computational analysis of Swedish lexical prosody. In: *Fonologi i Norden 2024*. Stockholm: 2024. pp. 19–20.
[8] Kwon J, Frid J, Söderström P, Roll M. Lexical phonological statistics of Swedish. (In Preparation)
[9] Kwon J, Roll M. 2024. Neural semantic effects of tone accents. *NeuroReport*

The early acquisition of lexical and post-lexical prominence in Italian. A case study.

C. Avesani[1], S. Bonifacio[2], G. Calignano[3], V. D'Aloia[4] F. Olivucci[2], M. Vayra[5], C. Zmarich[1]
[1]Institute of Cognitive Sciences and Technologies, Italian National Research Council; [2]Independent Researcher; [3]University of Padova; [4]University of Chieti-Pescara; [5]University of Bologna

Many experimental studies have shown that children from a very early age are sensitive to prosody [1]. Such sensitivity is rooted in prenatal experience with speech, which consists mainly of prosodic information, and it has been shown to affect how newborns perceive speech and produce communicative sounds [2]. At only 2 days of age, Italian newborns are sensitive to the alternation of stressed and unstressed syllables [3]. Between 6 and 10 months, infants are sensitive to lexical stress contrasts and show a preference for the predominant stress pattern of the ambient language, which, in turn, partly determines the capacity to extract word forms from fluent speech [3]. This early sensitivity to prosody contrasts sharply with the infants' limited production capabilities. Several studies have shown that English children produce lexical stress contrasts only in the second year of age [e.g., 4]. However, in none of them was a distinction made as to whether the prominent syllables considered were stressed at the word or phrase/sentence level. As for Italian, a couple of studies have investigated how children produce stressed and unstressed syllables, but one of them considered children in the late preschooler years of life, i.e. starting well past the stage of the first vocabulary, and did not try to delineate developmental trajectories for every single child [6], the other traced children's individual development [7] but did not separate lexical from post-lexical stress in the same way as [6].

In the present study, we examine the spontaneous productions of one male child from North-East Italy recorded every 3 months from 18 to 36 months of age while interacting with a caregiver. Each session contained a "free" combination of naming, repetition, and conversational speech tasks. The sample of language collected in each recording session was considered valid and representative of the child's linguistic abilities only if the number of lexical forms produced represented at least 50% of the words attested in the lexical list compiled by the parents, based on the Italian version of the MacArthur-Bates Communicative Development Inventory [8]. At 18 months of age, the child's expressive vocabulary reached the 50th percentile; at 36 months, it reached the 90th percentile.

Our longitudinal corpus allows us to investigate when the child begins to produce prosodically phrased utterances and when the internal structure of the prosodic constituents in terms of relatively strong and weak elements will emerge. The question we address is whether, when, and how a child learns to differentiate lexical from post-lexical prominence (stress vs. accent). Utterances of one, two, and multi-words have been (extended) IPA transcribed and prosodically coded with *Praat*. Each syllable was categorized within the framework of *Prosodic Phonology* [9] and the *Autosegmental Metrical Theory of Intonation* [10] as unstressed (U), lexically stressed (with no pitch accent) (S), prenuclearly accented (lexically stressed with a pitch accent in prenuclear position of an Intermediate/Phonological Phrase) (P), nuclearly accented in a Phonological/Intermediate phase (Nphp), nuclearly accented in an Intonational Phrase (N). Of the total number of syllabic occurrences (N=3454), the analysis focuses only on vowels occurring in open syllables (N=1157) of bi-, tri-, and quadrisyllabic words, with stress on the penultimate and ante-penultimate syllable and uttered spontaneously (repetitions were not considered). For each vowel of the selected syllables, we measured duration, spectral balance, and formant trajectories. Here we report results for vowel duration, as it was shown to be the most reliable correlate of lexical stress in Italian also for child speech [7].

Generalized additive mixed-effects models (GAMMs) were used to model the data for random and fixed effects as a function of time (duration.ms. ~ Age * Prominence + s(Age, by = Prominence), with the the *mgcv package* of R software. Fig. 1 shows the nonlinear regression lines for the estimated vowel duration of each syllable type across months. The estimated smoothed effect indicates that unstressed vowels predicted shorter duration compared to all the others since18 months of age. At the highest level of prominence, nuclear vowels in Intonational Phrases (N) predicted a significant increase in duration from 18 to 36 months of age compared to all the other vowels. At 21 months, with the emergence of argument structure and prosodic phrasing, Prenuclear accents appear, but it is only at 24 months that all vowel types show a significant difference. Still, nuclear vowels in Intonational Phrases (N) are shorter than in Intermediate Phrases (Nphp). At 27 months, the prominence hierarchy appears to be in place, as in adult speech, with vowel duration significantly increasing from U to N. However, the system is unstable: at 30 months, unstressed and stressed vowels become durationally undifferentiated and remain so also at 33 months. At 36 months, the child's production is again adult-like: the distribution of prominences within the prosodic constituents is what is expected in adult speech, with a linear increase in vowel duration from the lowest (U) to the highest (N) prominence level. Overall, results show that during the child's prosodic development the duration of Intonational Phrase-nuclear vowels increases linearly, the duration of unstressed vowels decreases linearly, and the duration of stressed, prenuclear, and Intermediate Phrase-nuclear vowels is progressively but non-linearly adjusted, consistent with the adult prosodic hierarchy.

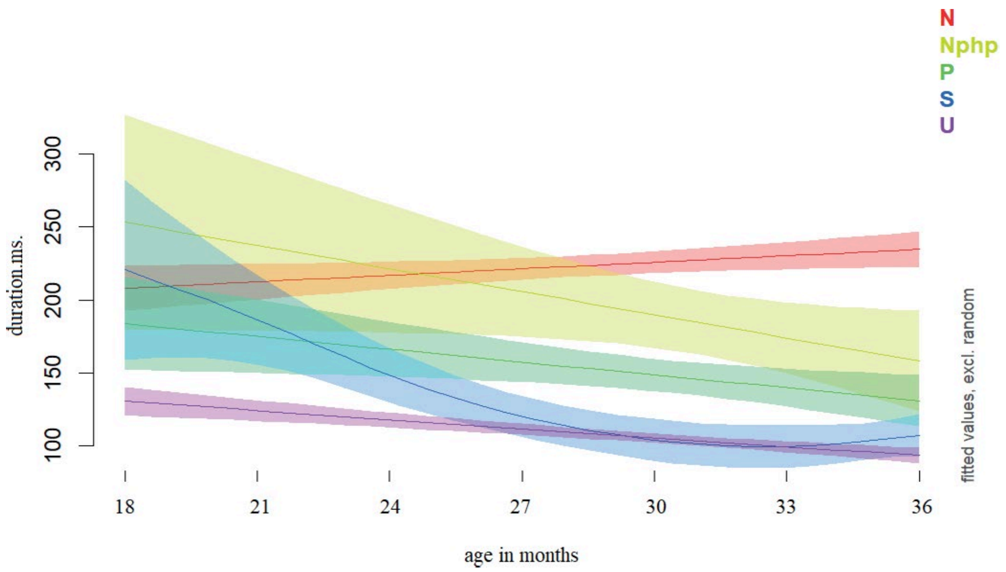


Figure 1. *Partial effects (fixed effects only) of the GAMM model showing the nonlinear regression lines for each of the five syllables (U=Unstressed, S=Stressed, P=Prenuclear, Nphp=Nunclear in Phonological Phrase, N=Nuclear in Intonational Phrase) with pointwise 95% confidence intervals*

[1] Prieto, P. & Esteve-Gibert, N. (Eds) 2018. *The Development of Prosody in First Language Acquisition*, John Benjamin Publishing, Amsterdam.

[2] Gervain, J. (2015). Plasticity in early language acquisition: the effects of prenatal and early childhood experience. *Current Opinion in Neurobiology*, 35, 13-20.

[3] Bahatara, A. Boll-Avetysian N., Höhle B., & Nazzi T. (2018). Early sensitivity and acquisition of prosodic patterns at the lexical level. In Prieto & Esteve-Gibert (Eds), 2018, 38-57.
[4] Pollock, K. E., Brammer, D. M., & Hageman, C. F. 1993. An acoustic analysis of young children's productions of word stress. *Journal of Phonetics* 21,183- 203.
[6] Arciuli, J., & Colombo, L. 2016. An acoustic investigation of the developmental trajectory of lexical stress contrastivity in Italian. *Speech Communication*, 80, 22-33.
[7] Olivucci, F., Pasqualetto, F., Vayra, M., & Zmarich, C. 2016. Lo sviluppo dell'accento lessicale nel bambino in età prescolare: una prospettiva fonetico-acustica. *Studi AISV*, vol.II, AISV, Milano.
[8] Caselli, M.C., Pasqualetti, P. & Stefanini, S. (2007). *Parole e frasi nel Primo Vocabolario del Bambino*. Milano, FrancoAngeli.
[9] Nespor, M., Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris. Berlin: Mouton de Gruyter.
[10] Beckman, M., Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook*, 3, 255–310.

Exploring first pretonic prominence in varieties of Russian with a multivariable reading task

Margje Post
University of Bergen

The East-Slavic languages Ukrainian, Belarusian and Russian have a typologically uncommon word prosody. The syllable preceding the stress is relatively long compared to other unstressed syllables, especially in Central Standard Russian (CSR). In CSR, non-high vowels /a, o/ in this first pretonic syllable (a_{-1}) also stand out in quality (after non-palatalized consonants), where they are produced as a low vowel, whereas second pretonic a_{-2} is heavily reduced to a short *schwa*. The prosodic word in CSR thus has vowel reduction in two degrees and a strong, disyllabic nucleus, opposed to a weak periphery.[1] Dubina (2012) explains this as the word prominence being realized over two syllables[2] and connects it to an abstract phonological high tone in Belarusian and Russian on the first pretonic syllable.[2,3] More empirical data are needed to assess how commonly this abstract tone surfaces in the phonetic realization of the word under various prosodic conditions and in various regions,[3] which differ in the alignment of high turning points.[4]

A recent empirical study of vowel durations and formant values in read speech of trisyllabic $Ca_{-2}Ca_{-1}$'CaC words in Moscow and Perm (Urals) (26 participants) indicates that historic geographical differences between central and non-central Russian dialects are retained in today's urban speech.[5,6] Here, a_{-1} was on average twice as long as a_{-2} in Moscow, and had a substantially higher F1, but, unexpectedly, in Perm only a small, not statistically significant difference in duration and F1 between the vowel positions was found,[5,6] suggesting that the prosodic word may not have a disyllabic nucleus there. This reading task was a subset of an exploratory, multivariable reading task in two cities, covering multiple prosodic conditions and patterns, with both falling and rising nuclear and prenuclear pitch accents.

In the current study we used the full dataset – with additional prosodic conditions and more speakers – to explore the role of F0 in first pretonic prominence, and its possible covariation with duration. Fluctuations in F0 were in part caused by differing pitch accents.

The F0 values for each item, measured at the midpoint of each vowel, were converted into semitones with a reference of 1 Hz to facilitate cross-speaker comparisons. We used multiple linear mixed effects regression models (LMM) to assess between-group and between-location changes in duration, and Spearman's rho to assess the association between the change in duration and change in F0 when moving from position a_{-2} to a_{-1}.

LMM analysis with Tukey posthoc tests in our extended dataset confirmed that the change in duration from position a_{-2} to a_{-1} was statistically highly significant in Moscow ($p = <0.001$ for both groups), but non-significant in Perm ($p = >0.33$ for both genders) (Fig. 1), as in the earlier dataset,[5,6] with comparable mean durations. A Bartlett Test indicated greater variance in change in F0 from a_{-2} to a_{-1} in Moscow ($p = <0.001$), but this greater F0 fluctuation does not appear to be associated with duration. There was a non-significant association between change in vowel duration and change in F0 in both genders in both Moscow and Perm (Fig. 2). This result suggests that F0 does not change consistently as a function of change in duration across word positions, in either city.

The high number of variables and some gaps in this larger, multivariable dataset restrict potential statistical analyses. Preliminary qualitative analyses suggest: 1) more frequent first pretonic peak alignment in Moscow, but it can occur in Perm as well, 2) a stable difference between Moscow and Perm, since all speakers followed the pattern of their city, 3) an unexpected difference in durations between disyllabic and trisyllabic words among some

Moscow speakers. These occurrences counter previous claims of Russian as stress-timed. We highlight the need for further data and exploration of this multifaceted phenomeno[1].

(1)

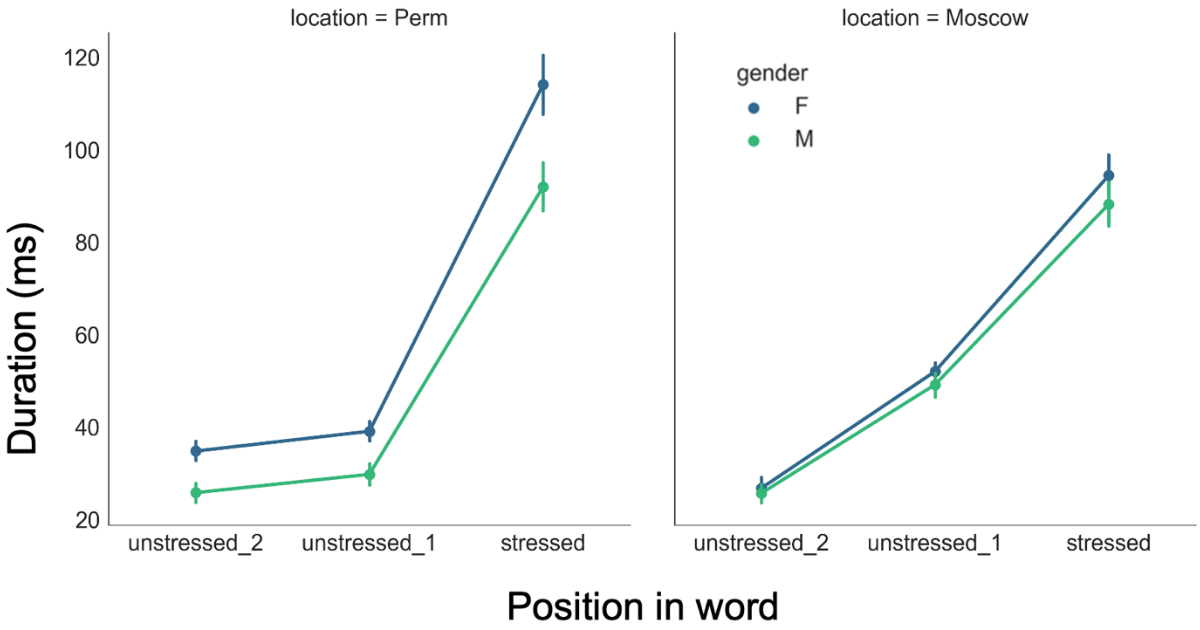


Figure 1. *Duration (ms) in positions* a_{-2}, a_{-1} *and 0 (stressed) by male and female speakers in Moscow and Perm.*

(2)

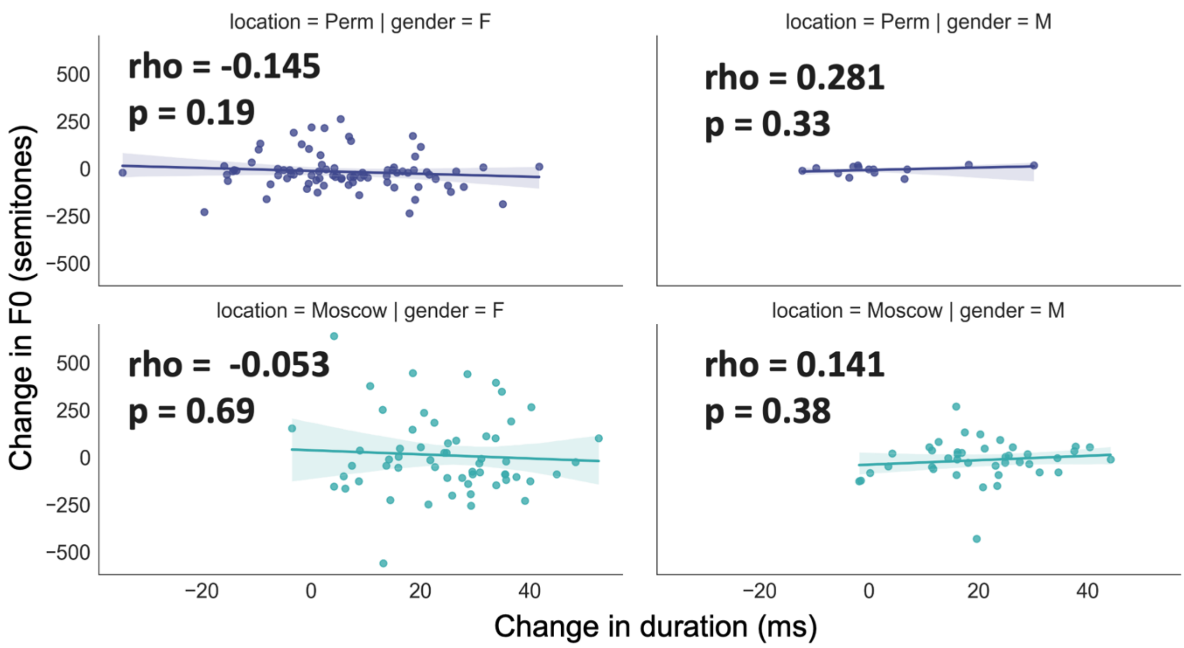


Figure 2. *Spearman's rank correlation of change in F0 from position* a_{-2} *to position* a_{-1} *by change in duration.*

[1] *Acknowledgements*: I thank Benedikte Fjellanger Vardøy for the recordings in Perm and Madeleine Rees for advice and statistics.

[1] Kodzasov, S. 1999. Russian. In Hulst, H. van der (Ed.), *Word Prosodic Systems in the Languages of Europe*, 852-870. Berlin: De Gruyter.
[2] Dubina, A. 2012. *Towards a tonal analysis of free stress*. PhD Dissertation. Utrecht: LOT: Radboud Universiteit Nijmegen.
[3] Mołczanow, J. 2015. The interaction of tone and vowel quality in Optimality Theory: A study of Moscow Russian vowel reduction. *Lingua* 163, 108-137.
[4] Duryagin, P., & Knyazev, S. 2022. Prosodic diversity in Standard Russian: pitch alignment in Central and Northern varieties. *Russian Linguistics*, *46*(2), 55-75.
[5] Post, M. 2025 (a). Regional prosodic variation in the speech of young urban Russians: Quantitative vowel reduction in Moscow and Perm. In Wagner, S., & Stange-Hundsdörfer, U. (Eds.), *(Dia)lects in the 21st century: Selected papers from Methods in Dialectology XVII*. (Language Variation). Berlin: Language Science Press.
[6] Post, M. 2025 (b). Word prosodic structure and vowel reduction in Moscow and Perm Russian. In Gehrke, B., Lenertová, D., Meyer, R., Seres, D., Szucsich, L. & Zaleska, J. (eds.), *Advances in Formal Slavic Linguistics 2022*. (Open Slavic Linguistics). Berlin: Language Science Press.

Oral session 8: Pragmatic and social effects

Greek polar question tunes in informal podcast interviews: A pilot study

Maria Lialiou[1], Mary Baltazani[2], Elinor Payne[2], Martine Grice[1]
[1]University of Cologne, [2]University of Oxford

Questions have garnered significant attention in linguistic research, particularly in the study of intonation [1]. The intonation of questions has mainly been investigated in laboratory settings, using controlled elicitation and read speech. Nonetheless, such settings do not create a natural communicative need, and thus speakers might not fully engage with the task, producing, in turn, a restricted and possibly non-typical set of question tunes, unlike in real-life interactions [1]. The present study investigates polar question intonation in Standard Modern Greek (hereafter, GR) using data from an informal podcast interview, exploring the extent to which the typical polar question tune observed in lab speech is representative of real-life conversations, the typical tune reported being a low (L*) nuclear pitch accent (NPA), followed by a high phrase accent and a low boundary tone combination (H-L%) [e.g. 2-5].

For this study, we extracted the polar questions (n = 13) from multiple speakers in one episode[2] of the GR podcast "Krima" [6], hosted by the GR radio producer Zoe Pre. Analysis of the question intonation followed the GR_ToBI system [7]. We found that polar questions are produced not only with the reported typical tune, but also with another, new, tune that has so far been unattested. Figure 1a) and b) illustrate examples of two polar questions, one with the typical tune (L* H-L%), the other with a new tune tentatively analysed as L+H* H-L% (where we leave open the possibility that the phrase accent may be a complex phrase accent LH- [e.g. 8]). Only four questions are produced with the typical L* H-L%, while the rest are produced with L+H* H-L%. Interestingly, both tunes have H-L% in common. This might be because the H-L% is an important cue to interrogativity, while the pitch accents potentially have a different pragmatic meaning. To further explore the differences between the two tunes, we extracted periodicity-modulated intonation contours (see Figures 1a) and b)), using the ProPer toolbox [e.g. 9-10], and implemented the Mass metric calculated as the sum of duration and power (i.e. the area under periodic energy curve). Specifically, we measured relative Mass, i.e. the area under the periodic energy curve of each syllable relative to the other syllables in the same utterance, allowing for comparisons across speakers and utterances. As shown in Figure 2, the syllables bearing the L+H* NPA tend to show less dispersed and higher mean Mass than the L* NPA, indicating that it is prosodically stronger. Focusing on the H-, it appears that the Mass of syllables bearing the H- is lower when the H- is preceded by a L+H* than when it is preceded by a L* NPA. Within tunes, H- tends to be higher in Mass compared to its preceding L* NPA, while the opposite trend is observed for the tune with the L+H* NPA.

To explore the pragmatic interpretation of the attested tunes, we employed the next-turn-proof practice [e.g. 11-12]. Analysing the recipient response as a source of evidence about how a question is interpreted, it appears that the L+H* H-L% tune conveys confirmation-seeking or positive bias, whereas the typical tune appears to be used for neutral information-seeking questions. In one example with the typical tune there appears to be negative bias, but further data is needed to confirm this interpretation as representative.

Summing up, in real-life informal conversations, we find not only the typical GR polar question tune but also a new tune. The differences between the tunes, both in categorical (NPA) and continuous (Mass) dimensions seem to encode different pragmatic meanings. Importantly, the new contour is more frequent, potentially because, in communicative settings like informal interviews, the majority of questions asked involve some kind of positive bias as the

[2] Annotation is ongoing and by the time of the conference, data from more podcast episodes will be available.

interviewer attempts to engage the interviewees on topics that are potentially relevant to them, initiating an elaborate and lively discussion.

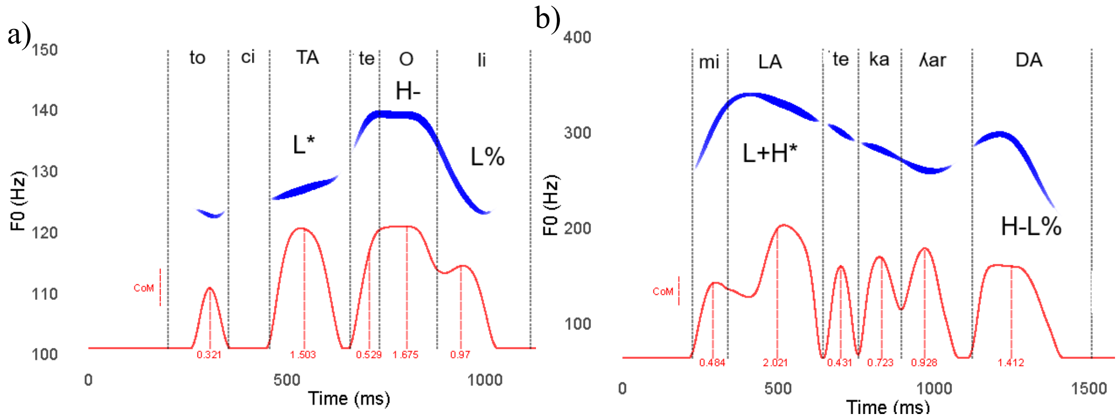


Figure 1. *Periogram & periodic energy curves of the questions 'Do you all see it?' (left panel) and 'Do you speak the (idiolect) Kaliarda?' (left panel; capitals indicate lexical stress). The blue line illustrates the f0 curve, with thickness reflecting its strength. The red solid line depicts the periodic energy curve. The Hz values on the y-axis correspond only to the f0 curve. The dotted vertical lines denote syllable intervals. The red vertical dashed lines indicate the position of the centre of Mass (CoM) within syllabic intervals. Mass values on a relative scale are given in numbers below each interval.*

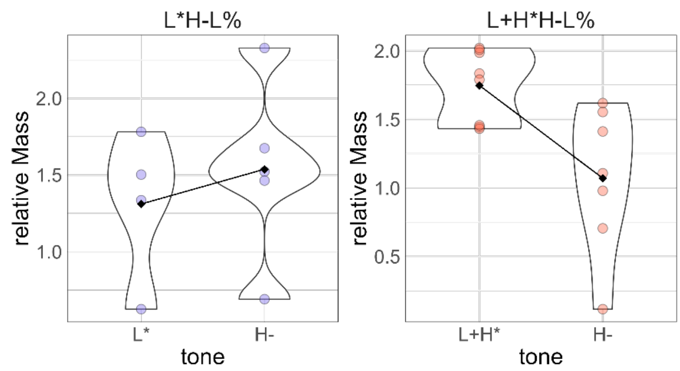


Figure 2. *Violin & individual dot plots illustrate the distribution of relative Mass associated with the syllables bearing the NPA (L* or L+H*) and the H- in the L*H-L% and L+H* H-L% tunes. Mean relative Mass is depicted by the black diamonds.*

[1] Cangemi, F., Grice, M., Jeon, H.-S., & Setter, J. 2023. Contrast or context, that is the question. *Speech Prosody*.

[2] Baltazani, M. 2007. Intonation of polar questions and the location of nuclear stress in Greek [Series Title: Phonology and Phonetics]. In A. Lahiri, C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes: Experimental Studies in Word and Sentence Prosody, 2*. Mouton de Gruyter: 387–406.

[3] Baltazani, M. 2003. Broad focus across sentence types in greek. *Proc. Eurospeech 2003*, 89–92.

[4] Arvaniti, A. 2002. The intonation of yes-no questions in Greek. *Selected papers on theoretical and applied linguistics 14*, 71–83.

[5] Grice, M., Ladd, D. R., & Arvaniti, A. 2000. On the place of phrase accents in intonational phonology. *Phonology, 17* (2), 143–185.
[6] Pre, Z. 2024. The first drag queen marriage in Greece (translation). Youtube.
[7] Arvaniti, A., & Baltazani, M. 2005. Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press: 84–117.
[8] Arvaniti, A., Ladd, D. R., & Mennen, I.. 2006. Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication 48*, 667-696.
[9] Albert, A., Cangemi, F., & Grice, M. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Proceedings of Speech Prosody 2018*. [10] Albert, A., Cangemi, F., Grice, M., & Ellison, T. M. 2020. ProPer: PROsodic analysis with PERiodic energy [Publisher: OSF].
[11] Heritage, J. 1984. *Garfinkel and ethnomethodology*. Polity Press.
[12] Sacks, H., Schegloff, E. A., & Jefferson, G. 1978. A Simplest Systematics for the Organization of Turn Taking for Conversation*. In J. Schenkein (Ed.), *Studies in the Organization of Conversational Interaction*. Academic Press: 7–55.

Speech timing and pragmatic influences on the production and perception of conversational turn transitions

Damar Hoogland, Laurence White

School of Education, Communication and Language Sciences, Newcastle University

The temporal coordination between successive speakers' utterances is typically characterised in terms of the interval between the end of one turn and the start of the next ("turn transition time" or TTT). Since TTT is frequently reported to be narrowly distributed around 200ms – often shorter than typical vocal reaction time – it is hypothesised that listeners must anticipate turn endings, timing the onset of their own turns according to consistent cues in the preceding utterance [1]. Some approaches propose that precise turn timing arises from perception-action entrainment [2,3], while others view it as a combination of prediction and reaction [1]. Here, based on three convergent studies of production and perception of turn timing in spontaneous conversation, we argue for more precise phonetic operationalisations of turn timing.

Study 1, via a systematic review of 25 cross-linguistic studies, examined variation in TTT distributions. All studies operationalised turn timing in line with the above TTT definition, but few provided phonetic definitions of turn offsets or onsets. TTT distributions had narrow peaks, thick tails and were generally right-skewed, with measures of central tendency that ranged between -830ms (i.e., overlap) and 3300ms, but most clustered between 0ms and 500ms. We also assessed the influence of dialogue factors on measures of central tendency, using forest plots and subgroup analysis. As Figure 1 shows, mean TTTs in task-oriented conversations ranged more widely than in free conversations, likely due to specific task demands. In addition, TTT variation was not systematically related to the particular language used (Figure 2).

Studies 2 and 3 tested the prediction, arising from entrainment theories [2,3], that faster speech rates should be associated with shorter TTTs. In previous work, the expected negative rate *vs* TTT associations are reported for experimental question-response paradigms [4,5], but corpus studies of spontaneous conversation found this negative association was only present in interruptions [6] or even found numerical evidence for a positive rate *vs* TTT relationship [7]. We analysed speech rate and turn timing in question-response sequences from 17 spontaneous Dutch and English conversations using linear regression in a model selection procedure. Study 2 found no language differences, but – across Dutch and English – speech rate improved the fit of a model predicting turn timing, $\chi^2(1) = 3.91$, $p = .048$; contrary to the entrainment hypothesis, however, higher question rate was associated with *longer* TTTs. Moreover, polar ("yes"/"no") questions were answered faster than open questions, $\chi^2(1) = 9.47$, $p = .002$, and relevant answers were had shorter TTTs than non-relevant answers, $\chi^2(1) = 10.93$, $p = .001$. When these dialogue factors were included, rate was no longer a robust predictor, $\chi^2(1) = .00$, $p = .998$, suggesting that it is less important than pragmatic features of conversation.

Of course, not all turns are produced "on time". Study 3 elicited listener judgements of the response "timeliness" of the Study 2 question-answer exchanges. Speech rate was a predictor of TTT only for responses that were judged timely (i.e., scoring in the middle third of the distribution of a 7-point Likert scale from "early" to "late"), $F(1) = 6.68$, $p = .011$, again with longer TTTs being associated with faster question speech rate. Timeliness ratings themselves were predicted by an interaction between TTT and turn-final syllable duration for English, $\chi^2(1) = 5.93$, p = .015, with a trend for the influence of turn-final syllable duration for Dutch, $\chi^2(1) = 3.04$, p = .081. For both, longer final syllables were overall associated with later ratings.

Allied to other findings, this suggests that the usual operationalisation of TTT may not be the most appropriate for describing or predicting turn-transition behaviour. More generally, these studies argue against entrainment theories of turn timing and indicate the relevance of dialogue factors, such as question type, answer type and task demands, to predict turn-timing.

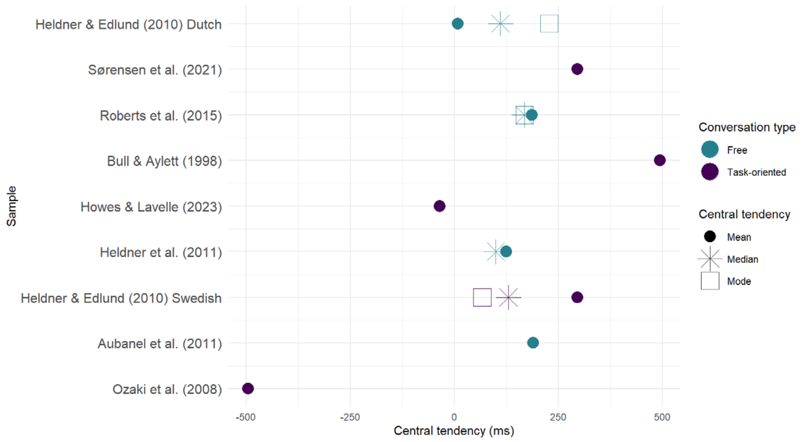


Figure 1. *Means, medians and modes of the nine studies that: (1) reported turn timing from full conversations; (2) did not exclude TTT values above/below a threshold; (3) did not reanalyse a previously used corpus.*

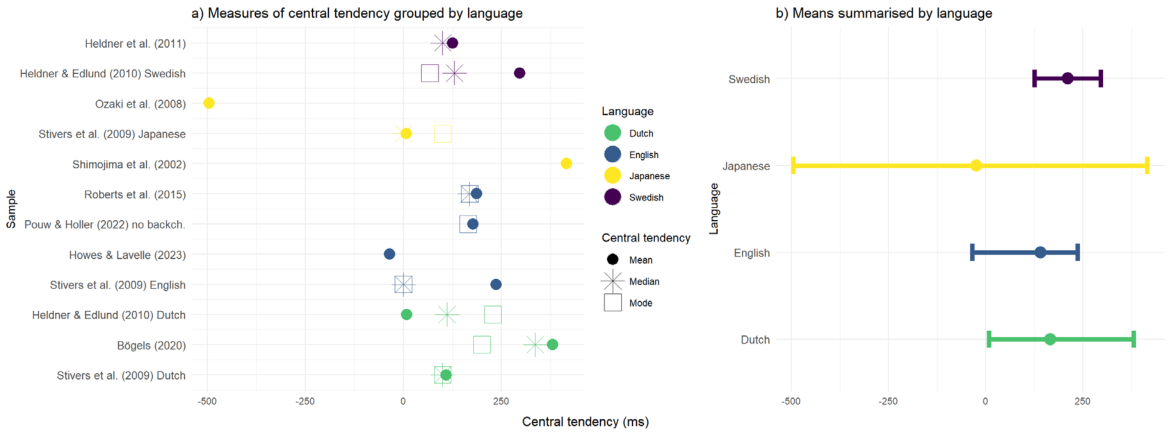


Figure 2. *Means, medians and modes of the twelve studies that: (1) did not exclude TTT values above/below a threshold; (2) concerned a language that was studied at least twice.*

[1] Levinson SC, Torreira F. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology* 6:731.
[2] Wilson M, Wilson TP. 2005. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*12:957–68.
[3] Garrod S, Pickering MJ. 2015. The use of content and timing to predict turn transitions. *Frontiers in psychology* 6, 751:1-2.
[4] Corps RE, Gambi C, Pickering MJ. 2020. How do listeners time response articulation when answering questions? The role of speech rate. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46, 4:781.
[5] Franich K. 2013 Prosodic predictors of temporal structure in conversational turn taking. *Proceedings of the International Congress of Phonetic Sciences 2023.*
[6] Beňuš Š. 2009. Are we in sync': turn-taking in collaborative dialogues. *Proceedings of Interspeech 2009*, 2167-2170.
[7] Roberts, Torreira F, Levinson SC. The effects of processing and sequence organization on the timing of turn taking: A corpus study. *Frontiers in Psychology* 6:509.

Gender and individual variation in spontaneous speech intonation

Stella Gryllia, Amalia Arvaniti
Radboud University, The Netherlands

According to recent research, individual characteristics, such as empathy, affect intonation processing (e.g., [1]), but it is unclear whether such characteristics also influence production. Here we examined the intonation of the Greek negative polar particle [ˈoxi] "no" used as a response to positive questions to understand whether tune choice is related to individual speaker characteristics. We additionally considered speaker gender, since gender has been shown to affect intonation production, especially the choice of preferred tune (e.g., [2]).

Our analysis is based on data from 18 (9F) functionally monolingual speakers of Standard Greek, 18-33 years old (mean = 23; S.D. = 3.8), recorded at the studio of the Musicology Department, University of Athens, in five groups of four (data from two speakers are not included here). Each group played the board game "Guess What + Who" in which the main player has to guess the entity depicted on a card they wear on a headband by asking the other players questions, which they can only answer with *yes* or *no*. Our corpus consisted of the F0 curves of 1614 [ˈoxi] responses (see Table 1). The F0 curves were first examined using hierarchical cluster analysis, which groups time-normalized F0 curves modelled as time series of 20 equidistant points [3]; this analysis included 1165 datapoints (out of 1614) due to filtering. The entire corpus of 1614 [ˈoxi] responses was also analysed using Functional Principal Component Analysis (FPCA), a data-driven approach that returns the major modes in which the F0 curves vary [4]. FPCA was used to a) ensure the output of clustering reflected genuine differences in the corpus, b) understand where changes took place in the tunes, and c) how tunes connected to individual traits, namely the speakers' Autism Quotient (AQ; [5]), Empathy Quotient (EQ; [6]), and musicality (measured using mini-PROMS [7]).

Fig. 1 shows the output of clustering, separately for male and female participants. Clustering suggests that speakers largely used falls, plateaux, and rising tunes, with falls being predominant among both groups, followed by plateaux for women and rises for men (which were rarely used by the female speakers). Fig. 2, which shows the first two principal components (PCs), supports the clustering findings. The first two PCs, which accounted for 84.6% of curve variance, showed differences related to the pitch accent and edge tones used. The pitch accent on the first syllable of [ˈoxi] was high-falling or low (red and blue lines of PC2, respectively), while the edge tones were rising or falling (red and blue lines of PC1, respectively). To assess whether individual characteristics affected tune choice, we ran linear mixed effects models using as predictors the speakers' AQ, EQ, and mini-PROMS score with the coefficients of the first two PCs as dependent variables. There were not significant effects for any of the three predictors.

Our results suggest that, in this communicative situation at least, Greek speakers used falls (H* L-L%), rises (L* H-H%), and plateaux (L* L-L% and H* H-!H%), with women showing a substantial preference for plateaux, which were not attested in the data of the male speakers. This difference, which may indicate more hesitancy on the part of the female speakers in our sample, suggests that gender is important in intonation production as it affects tune preference. This finding contrasts with the individual cognitive traits we tested: our results suggest that while these traits affect perception (cf. [1]), they may not be critical for production. This finding is in line with previous research, which shows sporadic or null effects of individual traits on speech production features unrelated to prosody, such as vowel quality [8], [9]. However, this should not be taken to mean that production is not affected by individual traits: production studies typically involve much smaller samples than perception studies, while detection of individual traits effects requires large samples [1]. Thus, it is possible that significantly larger production studies will be able document the effect of individual traits in production.

Table 1 *Number of [ˈoxi] tokens (#) per female (F) and male (M) speaker (SP)*

SP	#	SP	#	SP	#	SP	#	SP	#	SP	#
SP.F03	107	SP.F07	79	SP.F10	55	SP.M01	85	SP.M13	85	SP.M16	56
SP.F05	146	SP.F08	121	SP.F11	120	SP.M02	57	SP.M14	77	SP.M17	93
SP.F06	108	SP.F09	83	SP.F20	67	SP.M04	98	SP.M15	97	SP.M22	79

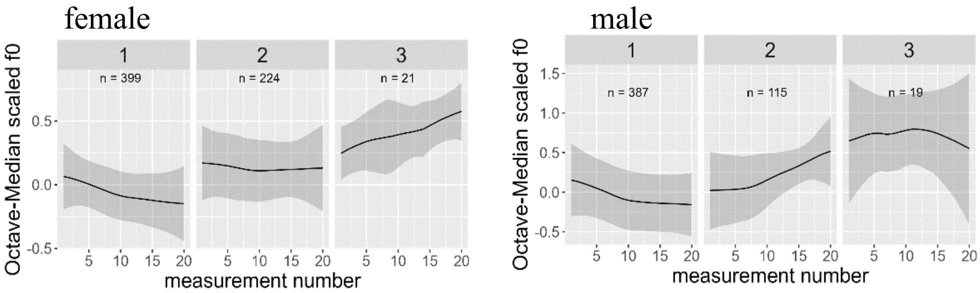


Figure 1. *Clustering, showing mean curves (black lines) and std error (gray bands), for female (left) and male speakers (right).*

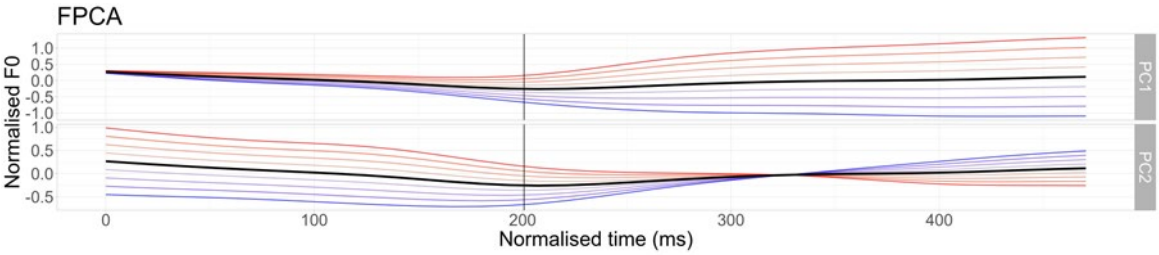


Figure 2. *FPCA output as applied to the entire data set of F0 curves; colour-coded curves illustrate the effect of each PC on the mean curve (solid black line); the vertical line indicates the offset of the accented vowel of [ˈoxi].*

[1] Orrico, R., S. Gryllia, J. Kim, A. Arvaniti. 2025. Individual variability and the H* ~ L+H* contrast in English. *Language and Cognition* 2025;17:e9. doi:10.1017/langcog.2024.62

[2] Arvaniti, A., & M. Atkins. 2016. Uptalk in Southern British English. *Speech Prosody 2016*. https://www.isca-archive.org/speechprosody_2016/arvaniti16_speechprosody.html

[3] Kaland, C. 2023. Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *JIPA* 53(1), 159-188.

[4] Arvaniti, A., Katsika, A., Hu, N. 2024. Variability, overlap, and cue trading in intonation. *Language* 100 (2): 265-307.

[5] Baron-Cohen, S., S. Wheelwright, R. Skinner, J. Martin, E. Clubley. 2001. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5-17.

[6] Baron-Cohen, S. & S. Wheelwright. 2004. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163-175.

[7] Zentner, M., & H. Strauss. 2017. Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS. *Ann. N.Y. Acad. Sci.* 1400, 33-45.

[8] Bishop, J., et al. 2022. Autistic traits predict spectral correlates of vowel intelligibility for female speakers. *Journal of Autism and Developmental Disorders* 52(5): 2344-2349.

[9] Turnbull, R. 2019. Listener-oriented phonetic reduction and theory of mind. *Language, Cognition and Neuroscience* 34(6): 747–768.

What do individual differences in pragmatic skill predict about prominence perception?

Jason Bishop[1,2], Mei-Ying Ki[2], Chen Zhou[2]
[1]*City University of New York-CSI ;* [2]*City University of New York-Graduate Center*

Perceiving prosodic prominence, like perceiving other aspects of speech, emerges from some combination of (a) bottom-up cues related to the signal and (b) top-down cues related to: linguistic knowledge (e.g., phonology); experience using language (e.g., lexical stats); and properties of the discourse context and the message itself (meaning) [1],[2],[3]. However, we know much less about the role played by properties of listeners themselves—i.e., individual differences. This study explores such differences.

The variety of individual differences we focus on here is related to a particular set of cognitive mechanisms underlying perception and processing of language, what we will refer to as (following work in sentence processing; e.g. [4]) *pragmatic skill*. Individuals differ in their level of this construct, i.e., in their propensity to integrate context, to engage in perspective taking (relevant to both Theory of Mind and empathy), and to interpret speakers' intentions. Work in recent years has shown that measures of pragmatic skill, such as those targeting neurotypical autistic traits (e.g. [3],[5]) and empathy (e.g. [6]), are to some extent predictive of listeners' sensitivity to prosodic cues in various perception and processing tasks.

The present study explored how such variation influences listeners' performance in Rapid Prosody Transcription ([7]), a speeded identification task widely used to test the perceptual salience of prosodic events, both phonetic and phonological. In particular, we asked whether individual differences in pragmatic skill related to autistic traits (based on subscales of the AQ and BAPQ; [8][9]) and empathy (the Reading the Mind in the Eyes task, henceforth "EYES"; [10]) mediate listeners' sensitivity to phonological structure, phonetic realization, and various top-down factors during prominence perception. Our hypothesis, based on recent work (e.g. [6]), was that higher scores on measures of empathy/ToM in particular may predict a more gradient strategy, making listeners' identification of prominence more sensitive to the following distinctions: *accent status* (a word's status as prenuclear accented or nuclear accented in the ToBI framework); *accent type* (whether a word bears a !H*, H*, or L+H*); and phonetic variation (f0 peak height) for words bearing H* and L+H*. Data were based on the judgments of 160 American English-speaking listeners who heard approximately 5 minutes of connected speech that was subsequently annotated for intonational structure (using ToBI conventions; [11]) and for discourse/information structure (using the RefLex scheme; [12]).

Mixed-effects regression analyses showed the following. First, higher scores on all measures of pragmatic skill were associated with the identification of an overall higher proportion of accented words, but no measure of pragmatic skill predicted differences in sensitivity to *accent status* contrasts (see Fig 1). Second, higher scores on EYES (our measure of empathy/ToM), but no other measure, was associated with significantly greater sensitivity to *accent type* (Fig 2). We also found EYES to be predictive of listeners' sensitivity to within-category phonetic cues; for words bearing L+H*, higher levels of empathy were associated with greater sensitivity to the height of f0 peaks (not shown). Moreover, higher EYES scores predicted stronger effects of discourse newness (Fig 3) and weaker effects of lexical frequency (not shown), the latter of which we note has been reported for clinical groups as well [e.g. [13]). Taken together, these findings suggest pragmatic skill—especially tied to measures ostensibly related to empathy/ToM in particular—can provide insight into the divergent strategies and/or sensitivities listeners exhibit in prominence perception tasks like Rapid Prosody Transcription.

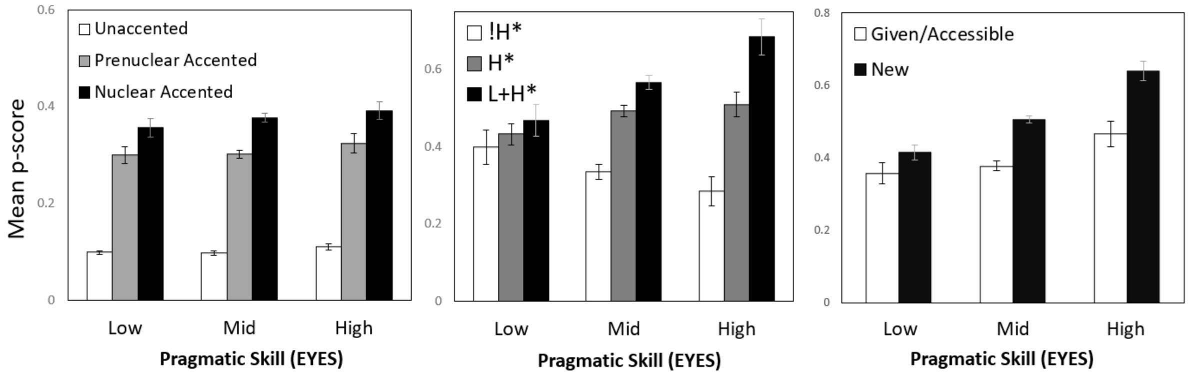


Fig 1 *(Left): Mean P-scores (proportion of words marked prominent) as a function of accent status and EYES scores (higher EYES scores indicate greater pragmatic skill related to empathy). Fig 2 (Middle): Mean P-scores as a function of accent type and EYES scores. Fig 3 (Right): Mean P-scores as a function of discourse new status (lexically new in RefLex) and EYES scores for words marked by L+H*. [Note: EYES was a continuous predictor in all models; binning is for visualization only]*

[1] Cole, J., Hualde, I., Smith, C., Eager, C., Mahrt, T., de Souza, R. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics* 75, 113-147.

[2] Cole, J., Mahrt, T., Hualde, J.I. (2014) Listening for sound, listening for meaning: Task effects on prosodic transcription. Proc. Speech Prosody 2014, 859-863.

[3] Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics,* 82, 100977.

[4] Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology*, 4, 708.

[5] Jun, S-A. & Bishop, J. (2015). Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier & E. Gibson (Eds.): *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor*. Springer.

[6] Arvaniti, A., Gryllia, S., Zhang, C., Marcoux, K. (2022). Disentangling emphasis from pragmatic contrastivity in the English H* ~ L+H*. *Proc. Speech Prosody 2022*, 837-841.

[7] Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1(2), 425–452.

[8] Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001a). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *J. of Autism and Developmental Disorders*, 31(1), 5-17.

[9] Hurley, R. S., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. *J. of Developmental Disorders*, 37(9), 1679–1690.

[10] Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I. (2001b). The "Reading the Mind in the Eyes" Test revised version. *J. of Child Psych. & Psychiatry*, 42, 241-251.

[11] Beckman, M., & Ayers Elam, G. (1997). Guidelines for ToBI labeling (Version 3). Unpublished manuscript, The Ohio State University.

[12] Riester, A., & Baumann, S. (2017). The RefLex scheme – annotation guidelines. In SinSpeC: Working Papers of the SFB 732 "Incremental Specification in Context." Stuttgart, Germany: University of Stuttgart. DOI: http://doi.org/10.18419/opus-9011
[13] Grice, M., Krüger, M. & Vogeley, K. (2016). Adults w/Asperger syndrome are less sensitive to intonation than control persons listening to speech. *Cult.&Brain* 4, 38-50.

Oral session 9: Language variation

Articulatory properties of the (pre)palatal affricate across two varieties of Spanish – an ultrasound and acoustic investigation

Karolina Broś, Małgorzata Ćavar, Simón González
[1]*University of Warsaw,* [2]*University of Indiana,* [3]*Australian National University*

Most Spanish dialects have only one affricate, described as a prepalatal (or postalveolar) voiceless /ʧ/, e.g. *mucho* ['muʧo] 'a lot' [1,2]. In the Canary Islands, however, the affricate is reported to be **shorter**, more **retracted** and tense, and is often referred to as an **'adherent'** *ch* with a greater tongue-palate contact surface [2,3,4]. Its closure part tends to be either partially or fully voiced due to post-vocalic voicing lenition. As the frication phase of this affricate is shorter and may go down to zero, it was proposed that the sound in question is a palatal stop rather than an affricate [4,5]. Recent phonetic work on the Canarian dialect [6] suggests that the shorter duration, frication portion and tenseness of *ch* can be the **result** of the voicing lenition process affecting stops. It is unclear, however, whether the place of articulation of *ch* is inherently different than in Peninsular Spanish without taking lenition into account. To investigate this, we conducted an ultrasound study using the Micro system from Articulate Instruments Ltd, collecting synchronized ultrasound and audio data.

13 speakers (7 from the Canary Islands, 6 from the Iberian Peninsula) took part in the study. The stimuli list consisted of 39 target words embedded in carrier phrases, with 16 words having *ch* in word-initial position, preceded by a pause or the vowel [o], e.g. *(cuatro) cheques de viaje* '(four) traveler's checks', and 20 words with *ch* produced word-medially preceded by a vowel, e.g. *diez cucharadas de azúcar* 'ten spoonfuls of sugar', *una pechuga de pollo* 'a chicken breast'. The participants produced three repetitions of each target phrase based on a prompt presented visually on the screen, using one of the leading sentences (*He comprado* 'I have bought' or *En la tienda no pueden faltar* 'The shop must not lack'). Duration, center of gravity and voicing measurements were extracted from annotated acoustic data using a Praat script. Ultrasound data were preprocessed using a dedicated R-based programme which calculated 20 angles marking the distance between an automated point of origin for each speaker and individual tongue contours. This allowed for the extraction of normalized within-participant data, which were then scaled to allow for between-group comparisons. Mean scaled distance and tongue displacement from the articulatory onset to the maximum constriction of the sound were used as dependent variables. We used cluster analysis to determine angles corresponding to the key sections of the tongue (the blade/predorsum, the dorsum and the root) for analysis.

The results of the study show that Canarian speakers produce a more palatal variant of *ch*, with a longer constriction in the blade and pre-dorsum area (group effect: χ^2=5.37, p<.05). At the same time, the root of the tongue is more advanced in these speakers (χ^2=11.08, p<.001). The major constriction in Peninsular speakers involves the blade and is shorter as compared to the Canarians (Fig.1). Also, the affricates are more palatalised when followed by high vowels, when unstressed and word-medial, compared to word-initial in both speaker groups. The results concerning the preceding environment are not consistent across tongue parts, hence we cannot confirm lenition-driven articulation differences in Canarian productions. Interestingly, the data show a significant difference between groups in the acceleration of tongue movement from sound onset to maximum constriction (χ^2=6.15, p<.05), while there are no group differences in the amount of displacement, confirming that the Canarian variant is briefer and more dynamic, and possibly more stop-like. Acoustic data show that the Canarian *ch* is shorter (χ^2=15.64, p<.001), and has a more retracted pronunciation (lower COG, χ^2=8.44, p<.01), confirming what we found in the articulatory analysis. This retraction may be due to the preceding context as there is bimodality in the Canarian productions after vowels but not after pauses (Fig. 2), suggesting that a different variant is produced in lenition contexts, perhaps due to simultaneous voicing and intensity changes. Further acoustic analyses of the data are ongoing.

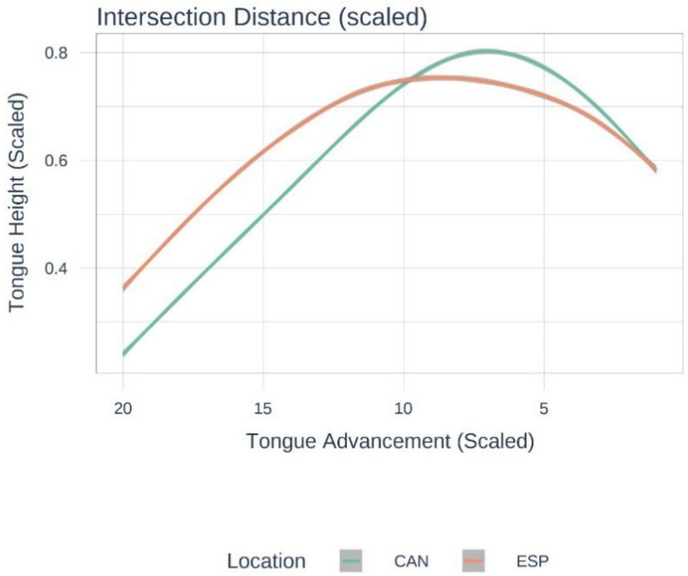


Figure 1. *Tongue trajectories showing averaged tongue contours by group. CAN – Canarian speakers, ESP – Peninsular speakers; tongue tip is on the right side of the graph.*

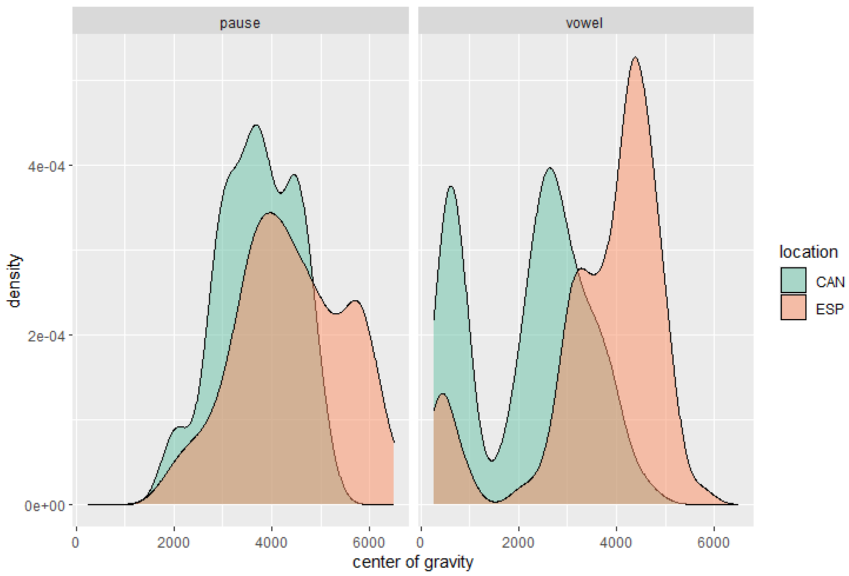


Figure 2. *Center of gravity differences in Hz by preceding context.*

[1] Martínez Celdrán, E. & Fernández Planas, A.M. (2007). *Manual de fonética española*. Ariel, Barcelona.
[2] Quilis, A. (1966). Datos para el estudio de las africadas españolas. *Travaux de Linguistique et de Literature 6:*403–412.
[3] Alvar, M. (1968). *Datos acústicos y geográficos sobre la ch adherente*. Estudios Canarios, Las Palmas de Gran Canaria.
[4] Dorta, J. (1997). Datos acústicos y percepción de la /c/ adherente de Canarias y de la prepalatal castellana. In Manuel Almeida & Josefa Dorta (eds.), *Contribuciones al*

estudio de la lingüística hispánica (Homenaje al profesor Ramón Trujillo). 57-72. Montesinos.
[5] Almeida, M. & Díaz Alayón, C. (1989). *El español de Canarias.* Santa Cruz de Tenerife: edición de los autores.
[6] Authors, to appear.

Contact-induced adaptation in a multidialectal context: Perception and production of front lax vowels in Indian Australian English

Elinor Payne[1], Olga Maxwell[2], Debbie Loakes[2], Mitko Sabev[3]
[1]Oxford University, [2]Melbourne University, [3]Saarland University

The Indian diaspora, some 35 million globally [1], is the second fastest-growing ethnolinguistic community in Australia [2], resulting in large-scale inter-dialect contact between Australian English (AusEng) and Indian English (IndE) [3]. Our study constitutes a first analysis of Indian Australian English (IndAusEng), and the effects of this multidialectal context. We consider evidence from perception and production, thus also contributing to the wider question of how these are integrated in processes of sound change [4].

Our focus is on English front lax vowels /ɪ, e, æ/, motivated by rapid changes in their production and perception in AusE, i.e. a change in perceptual boundaries between KIT-DRESS-TRAP from 1998-2004 [5], a pull chain lowering shift resulting in very open TRAP vowels in younger speakers [6], and highly significant perceptual DRESS-TRAP differences, including pre-lateral merger for some, including listener age [7]. The pre-lateral DRESS-TRAP merger is complex, being sometimes absent in younger AusE listeners, and, importantly in multilingual Australia, with its presence influenced by listeners' language experiences [8].

120 participants, comprising 34 AusE (age Mdn=27.3; Melbourne), 42 IndAusE (Mdn=33.8; Melbourne; varied length of residency (LoR)) and 44 IndE speakers (Mdn=27.9; India) took part in a perception and production study. Part I consisted of an online forced-choice auditory identification task. Target words in */CVC(V)/* and */(C)Vl(V)/*, of differing lexical frequency and shape (e.g., 1 vs 2 syllables), were synthesised to create a 7-step continuum. Part II consisted in the self-recording of target vowels in carrier words/sentences (a viable approach in phonetics [9]). Participants also completed a background questionnaire. A series of GLM analyses with STEP, COHORT, and GENDER as predictors, were performed using RStudio [10] in combination with a non-parametric method of conditional inference trees [11] to examine effects of LoR, gender and age.

Our findings reveal no merger effect with respect to the perception of DRESS-TRAP for AusE listeners, in line with [8] for a similarly young cohort. However, while IndE and IndAusE cohorts patterned with AusE responses for pre-lateral contexts (Fig.1a), both IndE cohorts showed a significantly later crossover to TRAP in non-lateral contexts, suggesting divergent perceptual expectations shaped by exposure to IndE vowels. That this difference is detectable in non-lateral contexts could be an artefact of poorer carryover coarticulatory cues in the voiceless obstruent coda (compared with the lateral context). The richer carryover cues in the pre-lateral context could override divergent expectations: there is simply more acoustic information on which to judge the difference in this context. This highlights the importance of factoring the dynamic, contextual nature of speech perception into research on sound change. Critically, however, IndAusE listeners showed a significantly earlier crossover than other IndE listeners (Fig.1b), indicating a perceptual adaptation towards the AusE system. Examination of conditional inference trees suggests adaptation may be further mediated by gender and LoR.

We interpret these findings as evidence for a gradient perceptual adaptation resulting from dialect-contact. We are now investigating production data for the same participants, with the hypothesis that adaptation exists also in production, but variably and possibly more weakly. We also consider the role of wider systemic adaptations, e.g. examining potential changes in the lateral (more clear articulation in IndE > more dark articulation in AusE) and how these could interact with perceptual expectations for the vowel. We discuss the implications of our findings for both theories of sound change as a result of dialect-contact, and theories of lexical representation, with regard to the integration of production and perception.

Figures 1a and 1b. *Observed probabilities/proportions by continuum step for /set-sat/ (1a - left panel) and /shell-shall/ (1b - right) in red (AusE), green (AusIndE) and blue (IndE).*

[1] Indian Diaspora. n.d. *Global Footprint of Indian Diaspora*. Retrieved 17 November 2024. https://www.indiandiaspora.org.

[2] Australian Bureau of Statistics. 2021. Census 2021. https://www.abs.gov.au/census

[3] Wiltshire, C. (2020). *Uniformity and Variability in the Indian English Accent* (Elements in World Englishes). Cambridge: Cambridge University Press.

[4] Pinget AF, Kager R, Van de Velde H. 2020. Linking Variation in Perception and Production in Sound Change: Evidence from Dutch Obstruent Devoicing. *Language and Speech*, *63*(3), 660-685.

[5] Mannell, R. 2004. Perceptual vowel space for Australian English lax vowels: 1998 and *2004.* In S. Cassidy, F. Cox, R. Mannell, & S. Palethorpe (Eds.), *Proceedings of the 10th Australian Speech Science and Technology Association Conference 2004* (Sydney*,* Australia), 221–226.

[6] Cox, F. & Palethorpe, S. 2008. Reversal of short front vowel raising in Australian English. In J. Fletcher, D. Loakes, R. Göcke, D. Burnham & M. Wagner (Eds.), *Proceedings of Interspeech 2008 incorporating ASSTA 2008* (Bonn, ISCA), pp. 342-345.

[7] Loakes, D., Clothier, J., Hajek, J., & Fletcher, J. 2024. Sociophonetic Variation in Vowel Categorization of Australian English. *Language and Speech, 67*(3), 870-906. https://doi.org/10.1177/00238309231198520

[8] Diskin-Holdaway C, Loakes D, Clothier J. 2024. Variability in cross-language and cross-dialect perception. How Irish and Chinese migrants process Australian English vowels. *Phonetica, 2024 Jan 15, 81*(1):1-41. doi: 10.1515/phon-2022-0007

[9] Zhang, C., Jepson, K., Lohfink, G. & Arvaniti A. 2021. Comparing acoustic analyses of speech data collected remotely. *J Acoust Soc Am. Jun, 149*(6), 3910. doi: 10.1121/10.000513

[10] R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[11] Levshina, N. 2015. *How to do Linguistics with R.* John Benjamins.

A Crosslinguistic Analysis of Vowel Intrinsic F0

Miao Zhang, Eleanor Chodroff
University of Zurich

Introduction. This study investigates vowel intrinsic F0 (VF0), the systematic fundamental frequency difference observed between high vowels (e.g., /i, u/) and low vowels (e.g., /a/) across languages [1], [2]. VF0 might arise as an automatic, biomechanical consequence of speech production, where the tongue pulling on the larynx during high vowel production could raise F0 or the jaw lowering during low vowel production could lower F0 [3]. Alternatively, speakers could be "in control" of such differences and might even enhance the biomechanical effect to serve as a secondary cue to the height contrast [4], [5].

This research seeks to answer whether VF0 arises as an automatic byproduct of speech articulation or is under speaker control to enhance the vowel height contrast. We investigate the following hypotheses: If the effect is biomechanical, high vowels should be produced with a slightly higher F0 than low vowels at a consistent magnitude across speakers and languages. If speaker control is apparent, two broad possibilities emerge: first, speakers could theoretically specify the high and low vowel duration targets in any way possible, resulting in no observable structure in the relationship. Alternatively, speakers could conform to their language population, revealing structure in the relationship on a language-specific level.

Methods and Data. Speech data was sourced from Common Voice Dataset [6], with phone- and word-level alignments from VoxCommunis Corpus [7]. Languages were considered for analysis if at least 5 speakers produced at least 10 tokens of high front, high back, and low vowels, respectively. The final analysis included data from 68 languages (i-a) and 64 languages(u-a). The speakers of each analyzed language ranged from 5 to 6818 (median: 70) for the /i/-/a/ analysis and 5 to 5824 (median: 50) for the /u/-/a/ analysis. Mean F0 was measured in hertz over the middle 10% of the vowel for high vowels [i, ı, y, Y, u, U, W] and low vowels [a, A, 5, 6] using Parselmouth [8]. The raw F0 in hertz was then converted to semitones for each speaker.

Results and Discussion. Results from the by-language linear mixed-effect models confirm that VF0 is a prevalent feature across languages, with high vowels exhibiting higher F0 (Figure 1 only shows the i-a results). Our study highlights significant inter-speaker variability. While most speakers across many languages tend to produce a higher F0 for high vowels, exceptions are evident in certain languages. Notably, tone languages such as Mandarin and Cantonese stand out as unusual cases where several speakers consistently produce a lower F0 for high vowels. This observation is particularly important as it suggests that while VF0 may be a general tendency, individual and language-specific factors can introduce variability in vowel production. The by-speaker linear regression further indicates that VF0 is a preferred tendency even at the individual level. While there are always speakers in all languages, except Taiwanese, who produce a significantly higher F0 for high vowels, there may be no speakers in some languages who produce a significantly lower F0 for high vowels. Regardless of statistical significance, speakers with a higher F0 for high vowels almost always make up the majority in all languages except Cantonese, Mandarin, and Japanese.

Conclusion. In conclusion, this study enriches the understanding of VF0 as a critical phonetic feature shaped by both universal physiological processes and language-specific characteristics. The results underscore a consistent tendency for higher F0 in high vowels while also revealing the complexities inherent in vowel production due to various factors. Future research could further investigate the implications of these findings, especially in the context of tonal languages and geographic distribution, and explore how VF0 variations affect speech technology and phonetic theory. Overall, this study contributes important data and insights to the ongoing discourse regarding vowel production and the nature of phonetic variation across languages.

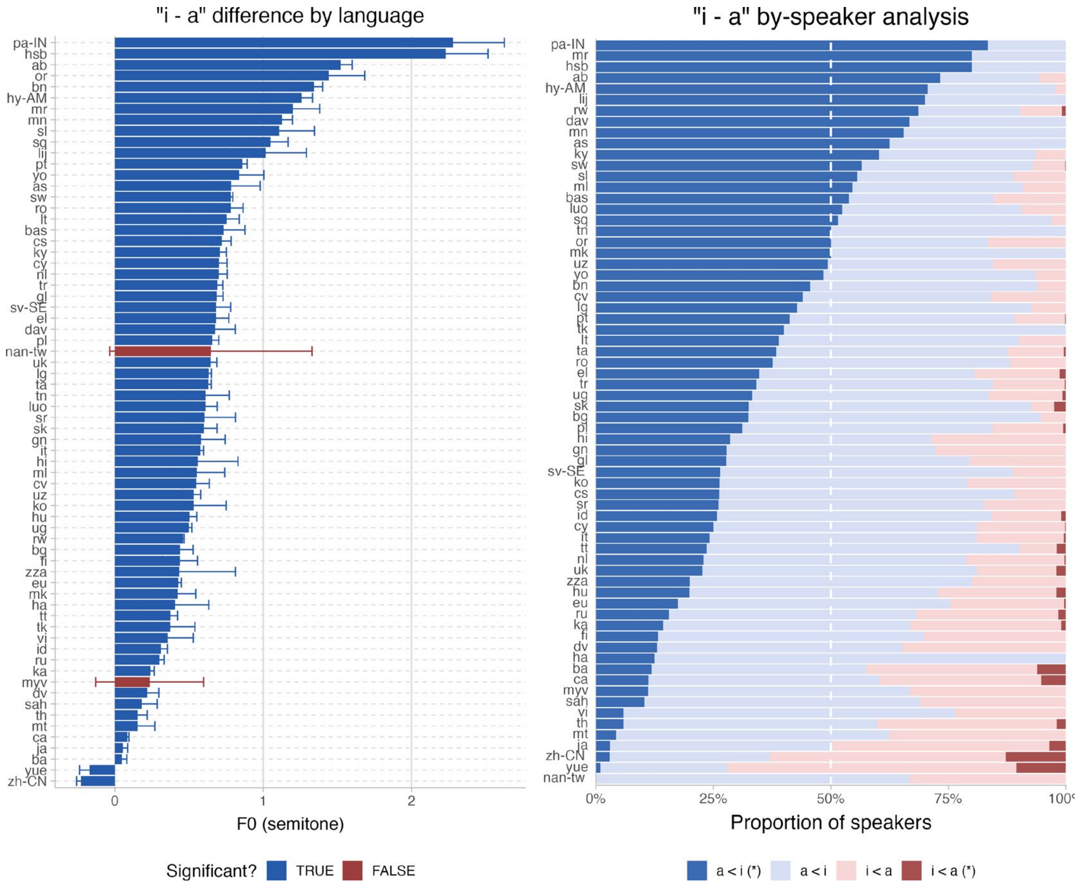


Figure 1: *"i-a" Comparison.*

[1] D. H. Whalen and A. G. Levitt, "The universality of intrinsic F0 of vowels," *Journal of Phonetics*, vol. 23, no. 3, pp. 349–366, 1995.

[2] C. Ting, M. Clayards, M. Sonderegger, and M. McAuliffe, "The cross-linguistic distribution of vowel and consonant intrinsic F0 effects," *Language*,

[3] W.-R. Chen, D. Whalen, and M. K. Tiede, "A dual mechanism for intrinsic f0," *Journal of Pho- netics*, vol. 87, p. 101 063, 2021.

[4] R. L. Diehl, "The Role of Phonetics within the Study of Language," *Phonetica*, vol. 48, no. 2-4, pp. 120–134, Mar. 1991, Publisher: De Gruyter Mouton.

[5] R. L. Diehl and K. R. Kluender, "On the Objects of Speech Perception," *Ecological Psychology*, vol. 1, no. 2, pp. 121–144, 1989.

[6] R. Ardila, M. Branson, K. Davis, *et al.*, "Common voice: A massively-multilingual speech cor- pus," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, European Language Resources Association, 2019, pp. 4218–4222.

[7] E. Ahn and E. Chodroff, "VoxCommunis: A Corpus for Cross-linguistic Phonetic Analysis," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, Jun. 2022, pp. 5286–5294.

[8] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, vol. 71, no. 2018, pp. 1–15, 2018, Publisher: Elsevier Ltd.

Vowel dispersion and focalization in the lexicons of 30 languages

Benjamin Storme[1], Kowsar Amiri[1], Alba Hermida Rodriguez[1,2]
[1]Leiden University, [2]Ghent University

Following Liljencrants and Lindblom's seminal paper on the Dispersion Theory [1], a number of works have shown that vowel inventories across the world's languages tend to favor acoustically dispersed vowels [2-4]. However there has been little research on whether acoustic dispersion also shapes the frequency of sounds *within* languages. Preliminary anecdotal evidence suggests that it does: peripheral vowels are on average more frequent than central vowels in the lexicons of languages [5].

This paper further investigates this question by fitting dispersion-based models to the frequency distribution of short oral vowels in the lexicons of 30 genetically and geographically diverse languages, sampled from the DoReCo multilingual speech corpus [6]. DoReCo was used to get estimates of both (i) the frequency of use of short oral vowels in the lexicons of languages and (ii) the acoustic realization of these vowels, corresponding to the first four formants measured at vowel midpoint, using the FastTrack plugin in Praat.

The lexical frequency of vowels was then modeled in a MaxEnt model [7] including a constraint favoring more distinct vowel contrasts. The specific implementation of this constraint closely follows [2], with vowel contrasts being penalized according to the square of their (weighted) Euclidean distance in the Bark-transformed F1 × F2' space, where F2' is a linear function of F2, F3 and F4 (see Figure 1). Following [2], we allowed languages to vary in the way F1 and F2' are weighted when calculating the Euclidean distance between vowels. Finally, we used the marginalization method proposed in [8] to infer the weight of the dispersion constraint from the lexical frequency of individual vowels in MaxEnt.

The dispersion-based model was then compared to three alternative models: (i) an unbiased model without any pressure towards dispersion, (ii) a model with a bias favoring focal vowels (i.e. with formants in close proximity), as in Quantal Theory [9], and (iii) a model combining vowel dispersion and focalization [2]. The unbiased model just predicts that vowels should be equally probable in the lexicon. The focalization constraint was operationalized as in [2], allowing the distance between F1 and F2 to be weighted higher than the distance between the other formants (F2-F3, F3-F4) in the overall formant-proximity measure.

All models were implemented as Bayesian multinomial regressions, using the JAGS package in R. Models were compared using the Deviation Information Criterion (DIC). The results show that models with a bias towards acoustic dispersion generally provide a better fit to the data than models that do not include this bias, in line with the hypothesis that the pressure for acoustic distinctiveness shapes not only the sound inventories of languages but also their lexicons. But the best model also includes the focalization constraint in addition, in line with the dispersion-focalization model of vowel inventories [2]. Dispersion-focalization models were generally found to be characterized by a smaller DIC than the other models, indicating a better fit to the frequency distribution of vowels in the lexicon. This was the case in 29 out of 30 languages (in the remaining language, the dispersion theory without focalization provided a better fit). Figure 2 compares the attested lexical counts of vowels in the 30 languages and the corresponding lexical counts predicted by the dispersion-focalization model. This model provides a reasonable fit to the data (R^2 = 0.62). As expected, peripheral vowels, and in particular [a], tend to be more frequent than non-peripheral vowels in the lexicons of languages. All in all, this paper provides quantitative evidence that acoustic dispersion and focalization play a role beyond sound inventories and also shape the way lexicons are built across languages.

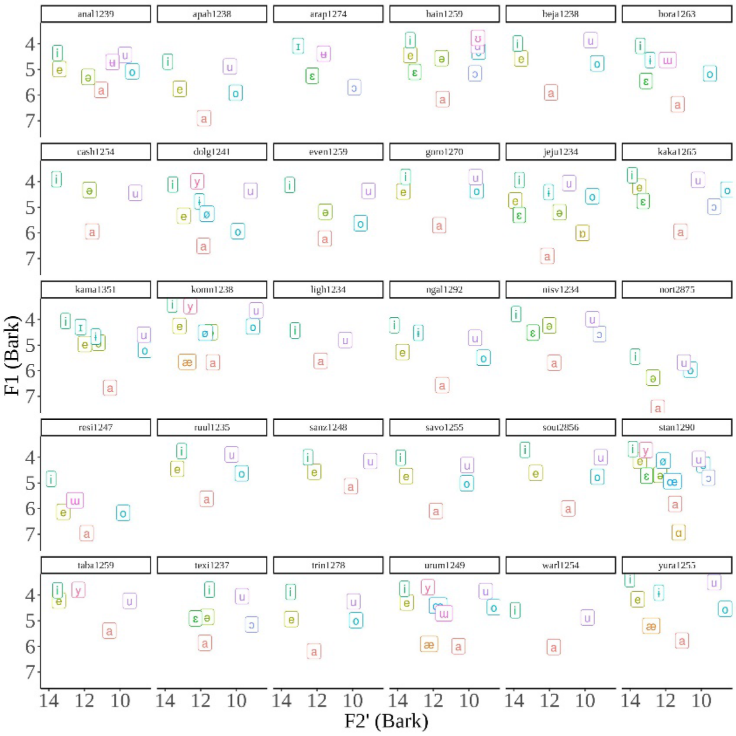


Figure 1. *Realization of short oral vowels in the F1xF2' space. The 30 languages are identified by their glottocodes, labels indicate mean values.*

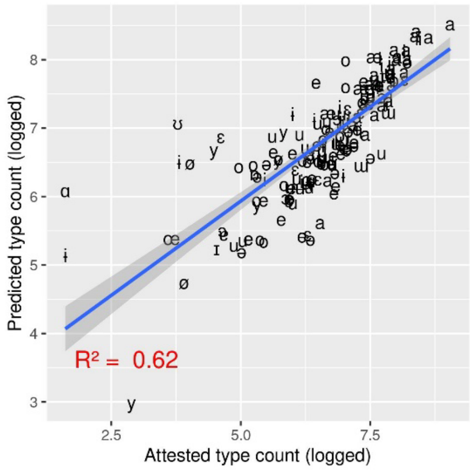


Figure 2. *Vowel count in the lexicon (logged): attested vs predicted by the dispersion-focalization model. Each label represents a vowel in one of the 30 languages.*

[1] Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839-862.

[2] Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of phonetics*, *25*(3), 255-286.

[3] Becker-Kristal, R. (2010). Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus. University of California, Los Angeles.

[4] Cotterell, R., & Eisner, J. (2018). A deep generative model of vowel formant typology. *arXiv preprint arXiv:1807.02745*.

[5] Gordon, M. K. (2016). *Phonological typology* (Vol. 1). Oxford University Press.

[6] Seifart, Frank, Ludger Paschen & Matthew Stave (eds.). 2022. Language Documentation Reference Corpus (DoReCo) 1.2. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

[7] Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, *39*(3), 379-440.

[8] Storme, B. (2023). A method to evaluate systemic constraints in probabilistic grammars. *Linguistic Inquiry* (advance online publication).

[9] Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics*, *17*(1-2), 3-45.

Oral Session 10: Nasality and spectral changes

What is left when nasalization is gone? Formant shift in formerly nasalized vowels in Larraine Basque

Inigo Urrestarazu-Porta[1,2,3], Ander Egurtzegi[1]
[1]CNRS – IKER UMR5478, [2]UPV/EHU, [3]UPPA

The loss of a nasal aspiration /h̃/ created phonologically nasal vowels (/Ṽ/) in all Basque varieties. Gradually, contrastive nasalization was lost in most of them, and has only been preserved in the endangered Zuberoan dialect until recent times (Hualde 2003: 31, Egurtzegi 2014: 151).

However, not all Zuberoan varieties seem to have maintained vowel nasalization to the same degree; Lafon (1958: 118) already observed that the Larraine variety of Zuberoan Basque had lost contrastive nasalization in the 1950's, with nasalized vowels only found in contact to nasal obstruents. Alas, the fact that both vowel nasalization in Basque and the Zuberoan dialect are understudied makes it difficult to have a clear picture of this contrast.

Nasal acoustic cues in vowels are not limited to audible nasalization. As is known, nasalized vowels show differences in their formant structure when compared to their oral counterparts (Whalen & Beddor 1989; Carignan 2018), and these differences can also be used for the perceptual discrimination between oral and nasal sounds.

With this in mind, we have analyzed the production of (formerly) nasalized vowels /ĩ/ and /ũ/ to check 1) if they have lost the nasality they previously had, as Lafon suspected, and 2) whether the opposition of between the formerly nasal and oral vowels has been lost or it has been transphonologized to a new, height-based opposition.

To answer to these questions, we recorded 11 speakers (4 women) of the Larraine variety of Basque. The experiment involved reading isolated words that were specifically extracted from Larrasquet's (1932) glossary (which includes phonetic transcriptions of all its entries). Thus, the elicited words were transcribed including one of the sounds of interest in a Zuberoan variety of the 1930's. For the recordings, we used a nasometer. This device is composed of two microphones, divided by a wooden board that serves to separate the airflow coming from the mouth from that coming out of the nose, and record them in separate channels. This device is increasingly popular in phonetic research, and it has already been used to analyze nasal sounds in many languages (Rodriquez et al. 2023).

The stereo nasalance data was processed using Praat. For the acoustic analysis, both the nasal and oral channels were band-pass filtered (80 Hz-10000 Hz) and the nasalance —ratio of the nasal amplitude to the sum of the oral and nasal amplitudes, i.e. $A_n/(A_o + A_n)$ (Carignan 2018)— was computed every 5 ms.

We have fit two Bayesian regression models to answer to our research questions. The first model analyzed the nasalance values as a function of the interaction between the place of articulation (I and U) and the etymological group (nasalized, oral, and contextually nasalized vowels), with by-speaker correlated varying slope and intercept adjustments and by-word varying intercept adjustments. We included weakly informative priors to stabilize the posteriors. In our data, /V/ and /Ṽ/ are identical in terms of nasality, and both are different from contextually nasalized vowels —i.e. vowels adjacent to a nasal obstruent (see Figure 1).

The second model was used to check whether the neutralization of /V/ and /Ṽ/ is complete, or a new acoustic cue currently distinguishes these two sets of segments. We fitted a multivariate model that included the logarithm of F1, F2, F3 and duration as a function of, again, the interaction between vowel and etymological group. In our results, /i/ and /ĩ/ are completely neutralized, with no differences in any of the first three formants. Nonetheless, our results are consistent with /ũ/ being uttered further back than /u/, which could be enough to maintain an opposition that has been described as lost (see Figure 2).

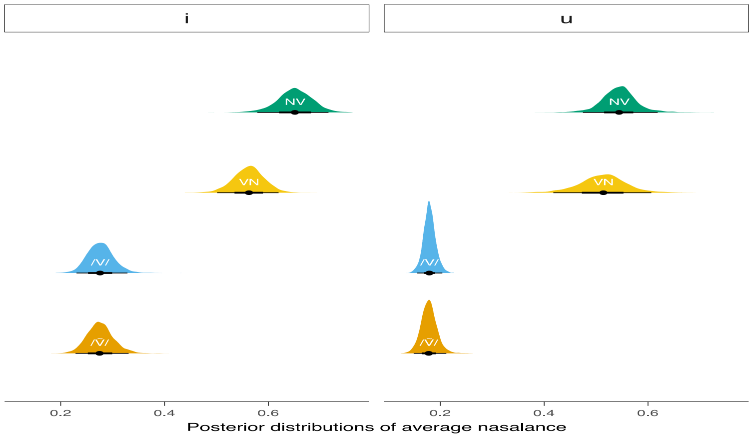


Figure 4: *Posterior nasalance*

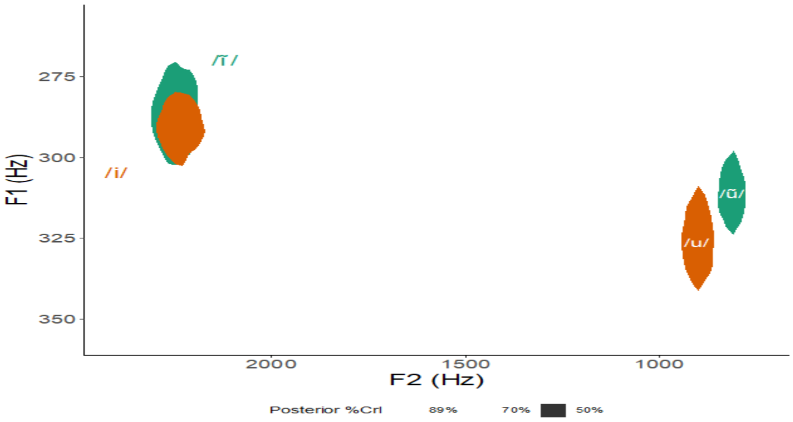


Figure 5. *Posterior distributions of F1 and F2*

Carignan, C. (2018). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *JASA* 143 (5), pp. 2588–2601.

Egurtzegi, Ander. 2014. Towards a phonetically grounded diachronic phonology of Basque [PhD dissertation of the UPV/EHU].

Hualde, José Ignacio. 2003. Segmental Phonology. In José Ignacio Hualde & Jon Ortiz de Urbina (Eds.), *A grammar of Basque* (pp. 16-65). Mouton de Gruyter.

Lafon, René. 1958. Contribution à l'étude phonologique du parler basque de Larrau (Haute-Soule). In Diego Catalán (Eds.), *Miscelánea Homenaje a André Martinet*, II (pp. 77-106). Biblioteca filológica.

Larrasquet, Jean. 1939. *Le Basque de la Basse-Soule Orientale*. Paris: C. Klincksieck.

Rodriquez, Francesco; Pouplier, Marianne; Alderton, Roy; Lo, Justin H.; Evans, Bronwen;
Reinisch, Eva; Carignan, Christopher. 2023. What French speakers' nasal vowels tell us about anticipatory nasal coarticulation. In *ICPhS20* (pp. 848-852). Guarant International.

Whalen, D. H., & Beddor, Patricia S. (1989). Connections between nasality and vowel duration and height: Elucidation of the Eastern Algonquian intrusive nasal. *Language*, 65, pp. 457–486. https://doi.org/10.2307/415219

Delayed nasality in Quebec French nasal vowels: Mechanical or controlled?

Josiane Riverin-Coutlée

Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München

The phonological inventory of the variety of French spoken in Quebec (henceforth QFrench) possesses four nasal vowels: /ẽ, œ̃, ã, ɔ̃/ [7]. This paper is concerned with the time course of oral and nasal intensity in these nasal vowels. The main aim is to determine whether a trend reported in earlier literature [e.g. 2,4,5], where the onset of nasality is delayed within QFrench nasal vowels, is a mechanical or a controlled phenomenon [9].

Nasality – contrastive or coarticulatory – is known to be realized in a language- and/or variety-specific manner [e.g. 2,3,4,5,8,10]. For contemporary QFrench nasal vowels, recent aerodynamic studies have suggested that nasality was "right-shifted", i.e. nasal airflow starts later than vowel onset, is very high at vowel offset, and only decreases after vowel offset [2,4,5]. This has been taken as evidence that QFrench nasal vowels are composed of an oral part followed by a nasal part, the latter being heavily nasalized. However, the temporal extent of the oral part is not necessarily clear or consistent across studies, which may partly be due to the coarse time resolution of nasal airflow measurements. Most importantly, earlier literature is not explicit about whether the origin of the oral part could be mechanical (coarticulatory), or a controlled (phonologized) property of QFrench nasal vowels. A constant *duration* of the oral part across vowel durations would support a mechanical account; whereas a constant *proportion* across vowel durations could point to a controlled phenomenon [9].

In this paper, we assess vowel nasality over the course of vowels using oral and nasal intensity data collected separately with a nasometer (see details below). We examine oral and nasal intensity over the course of short and long variants of QFrench nasal vowels to assess which scenario, mechanical or controlled, finds most support.

Twenty native speakers of Quebec French were recorded with a commercial nasometer, which allows two microphones separated by an acoustic baffle to pick up sound waves from the nose and mouth separately. The participants performed a reading task where target words embedded in carrier sentences appeared at the end of intonational phrases. The ±1250 nasal vowels produced by 8 speakers (6W, 2M, 25-63 y.o.) analyzed here were extracted from word-final syllables ending with two different types of codas: no coda (open syllable), e.g. [pɔ̃] *pont* 'bridge', where vowels are short; and voiceless codas, e.g. [pɔ̃ːp] *pompe* 'pump', where vowels are long [7]. The speech signals were forced-aligned with WebMAUS [6], then hand-corrected. Praat [1] was used to sample separate intensities for nasal and oral channels 101 times from vowel onset to offset. The data were modeled with a generalized additive mixed model (GAMM, [12]) in which the response variable was *intensity* over normalized time; a fixed-term parameter for the interaction between *channel* (nasal, oral) and *coda* (no coda, voiceless) was included; as well as random smooths per *speaker* and *lexical item* (the latter accounting for potential differences across /ẽ, œ̃, ã, ɔ̃/). Residuals were modeled as scaled-t distributed and an autocorrelation term was added [11].

The oral and nasal intensity curves predicted by the model are shown in Figure 1. In the first 30% or so of the normalized duration, oral intensity (dashed) is high, while nasal intensity (solid) starts low and increases. Oral and nasal intensities then cross-over, after which oral intensity steadily declines until vowel offset, while nasal intensity plateaus (no coda, solid black) or keeps increasing (voiceless, solid red), and remains higher than oral intensity despite a final drop. Taken together, these qualitative trends corroborate earlier descriptions of an oral part followed by a heavily nasalized part. Since these oral and nasal parts occupy similar *proportions* of the short and long vowels' duration, late nasalization seems to be a phonologized, controlled property of QFrench nasal vowels, which calls for further investigation of velum activity at an even finer temporal level, using e.g. real time MRI [3].

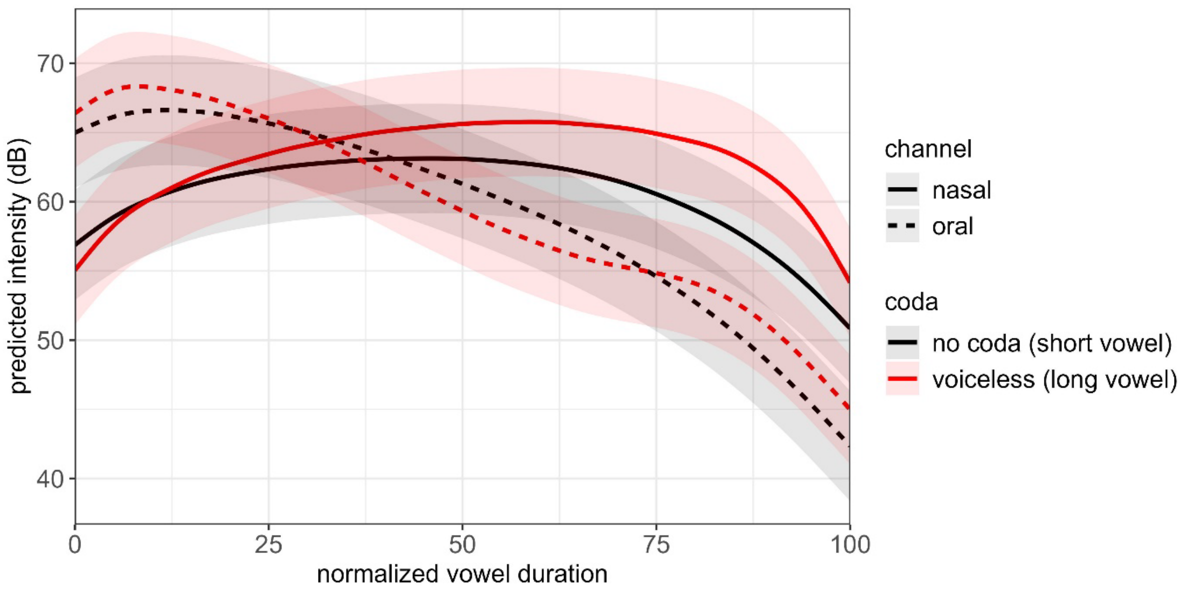


Figure 1. *GAMM predictions for nasal and oral intensity over normalized duration of QFrench nasal vowels followed by no coda and a voiceless coda*

[1] Boersma, P., & Weenink, D. 2019. *Praat: doing phonetics by computer*. Computer program, version 6.1.x.

[2] Carignan, C. 2013. When Nasal Is More Than Nasal: The Oral Articulation of Nasal Vowels in Two Dialects of French. University of Illinois at Urbana-Champaign PhD dissertation.

[3] Cunha, C., Hoole, P., Voit, D., Frahm, J., & Harrington, J. 2024. The physiological basis of the phonologization of vowel nasalization: A real-time MRI analysis of American and Southern British English. *Journal of Phonetics* 105, 101329.

[4] Delvaux, V. 2006. Production des voyelles nasales en français québécois. *Actes des XXVIes Journées d'études sur la parole* (Dinard, France), 383–386.

[5] Desmeules-Trudel, F., & Brunelle, M. 2018. Phonotactic restrictions condition the realization of vowel nasality and nasal coarticulation: Duration and airflow measurements in Québécois French and Brazilian Portuguese. *Journal of Phonetics* 69, 43–61.

[6] Kisler, T., Reichel, U. D., & Schiel, F. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45, 326–347.

[7] Martin, P., Beaudoin-Bégin, A.-M., Goulet, M.-J., & Roy, J.-P. 2001. Les voyelles nasales en français du Québec. *La Linguistique* 37(2), 49–70.

[8] Pouplier, M., Rodriquez, F., Lo, J. J. H., Alderton, R., Evans, B. G., Reinisch, E., & Carignan, C. 2024. Language-specific and individual variation in anticipatory nasal coarticulation: A comparative study of American English, French, and German. *Journal of Phonetics* 107, 101365.

[9] Solé, M.-J. 2007. Controlled and mechanical properties in speech: A review of the literature. In Solé, M.-P., Beddor, P. S., & Ohala, M. (Eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 302–321.

[10] Styler, W. 2017. On the acoustical features of vowel nasality in English and French. *Journal of the Acoustical Society of America* 142(4), 2469–2482.

[11] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86–116.

[12] Wood, S. N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd edn. Boca Raton: Chapman and Hall/CRC.

Relating spectral and articulatory changes in sequences of alternating sounds

Leonardo Lancia[1], Ekaterina Biteeva[2], Cécile Fougeron[2]
[1]*Laboratoire Parole et Langage (CNRS / Aix-Marseille Université),* [2]*Laboratoire de Phonétique et Phonologie (CNRS / Université Sorbonne Nouvelle)*

The amount of articulatory displacement is related to the saliency and discriminability of the produced speech sounds [1-2] and as such it is an important feature in the characterization of both healthy and disordered speech production behaviour. When lacking articulatory data, the tendency toward hyper- or hypo- articulated speech can be inferred from the acoustic signal by measuring the distance between different vocalic sounds in the formant space. This approach is however not well suited for the analysis of disordered speech or of other kinds of strongly variable speech that challenge formant tracking methods [3, 4].

Goldstein [5], shows that the amount of articulatory change is strongly related to the amount of acoustic change as quantified by computing at each point in time the distance between consecutive configurations of the first 13 MFCCs coefficients. In the present work, we further analyse the relation between the change of the MFCC coefficients and the amplitude of articulatory movements through the analysis of controlled speech material for which we recorded both acoustic and articulatory signals (through EMA sensors attached to the UL, LL, TT and JAW). From the acoustic signal, we compute cepstral change (henceforth ΔMFCC) as the square root of the squared and summed frame-by-frame differences of coefficients from 2 to 13 (as the first one reflects the amount of acoustic energy averaged over Mel bands). ΔMFCC constitutes itself an estimate of the amplitude of the underlying changes in the vocal tract. A further estimate is the maximum envelope of the ΔMFCC, which is computed via interpolation of the ΔMFCC peak values, expected to reflect peak articulatory velocities.

We recorded seven French speakers repeatedly producing three-syllabic utterances containing alternating sequences of sounds (/ajajaj/, /ujujuj/, /wiwiwi/) at normal and fast rates. So far, the complete analysis pipeline has been applied to two of them. Once the frontiers between segments were manually identified based on acoustic data, we used productions recorded at slow speech rate to train a Linear Discriminant Analysis classifier to predict the sound identity at each point in time from the corresponding configuration of articulatory positions and velocities. This procedure permits obtaining a "uni-dimensional articulatory trajectory" corresponding to the time-varying discrimination function that encodes the probability that at each point in time the speaker is producing one or the other sound (e.g. /u/ or /j/). This uni-dimensional articulatory trajectory is illustrated in figure 1a (dashed line) with its amplitude (i.e. amplitude of articulatory displacements, solid line). Our estimates of acoustic changes are aligned in Figure 1b with ΔMFCC (orange) and maximum envelope (blue).

The data in figure 1 correspond to eight tokens of the utterance /ujujuj/ produced at normal (first 4) and fast rate (last 4) by one speaker. In these productions, the amplitude of articulatory trajectory is clearly smaller at fast speech rate and word-internally (i.e. on the medial syllable of the tri-syllabic token). This hypoarticulation is also captured by a reduction of the ΔMFCC peaks and maximum envelope. The correlation between the acoustic estimates (ΔMFCC and maximum envelope of ΔMFCC) and the amplitude of articulatory trajectory, is displayed in Figure 2 for all tokens produced by the two speakers analyzed. When averaged over the duration of each sound (columns 1 & 2), articulatory amplitude seems to be better captured by ΔMFCC (confidence interval are smaller and more uniform) than by its maximum envelope. However, when averaged over the duration of each token, the amplitude of articulatory trajectories is better captured by the maximum envelope of ΔMFCC (columns 3 & 4).

Our preliminary results validate the use of ΔMFCC as a proxy of articulatory changes, which are related to the effect of rate or prosodic organization in this study. The behaviour of the remaining five recorded speakers will be discussed at the conference.

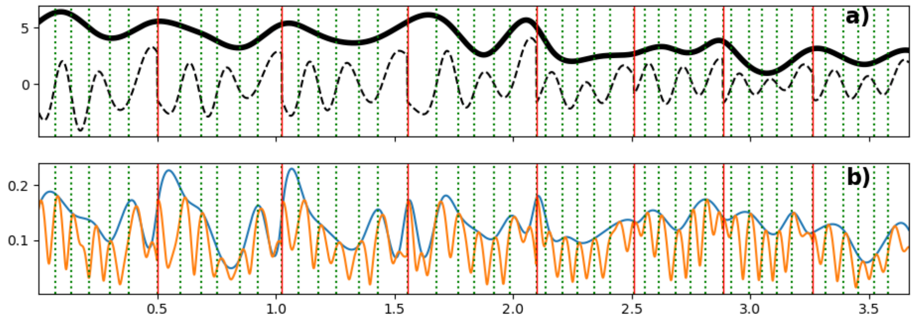


Figure 6: *Production of eight token of /ujujuj/ by on speaker at normal (first 4) and fast (last 4) rate. 1a: articulatory changes estimated by a uni-dimensional articulatory trajectory (dashed line) and its amplitude (solid line), 1b: acoustic changes estimated by ΔMFCC (orange) and maximum envelope of ΔMFCC (blue). Green dotted vertical lines separate segments, Red continuous lines separate /ujujuj/ tokens.*

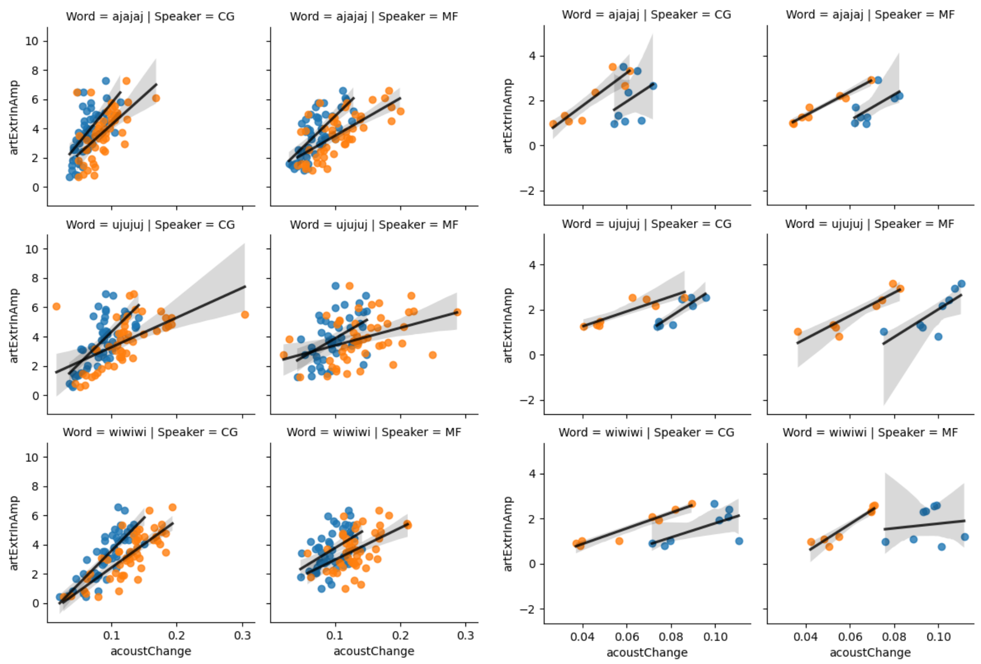


Figure 7 Amplitude of the uni-dimensional articulatory trajectory vs. ΔMFCC (orange dots) and vs. maximum envelope (blue dots). Regressions in column 1&2 comprise one data point per segment, and one data point per token in columns 3&4.

[1] Lindblom, B. (2014). The status of phonetic gestures. In *Modularity and the motor theory of speech perception* (pp. 7-24). Psychology Press.
[2] Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. *Linguistics*21.263 (1983): 181-203.
[3] Lévêque, N., A. Slis, L. Lancia, G. Bruneteau, C. Fougeron (2022). Acoustic change over time in spastic and/or flaccid dysarthria in motor neuron diseases. *Journal of Speech, Language, and Hearing Research,* 65(5):1767-1783.

[4] Slis A., N. Lévêque, C. Fougeron, M. Pernon, F. Assal, L. Lancia (2021). Analysing spectral changes over time to identify articulatory impairments in dysarthria. *Journal of the Acoustical Society of America*, 149 (2), pp.758 - 769.
[5] Goldstein, L. (2019). The role of temporal modulation in sensorimotor interaction. *Frontiers in Psychology*, *10*, 2608.

Poster Session 1a

1. Accommodation of pitch-related parameters to interlocutors' power: a gendered perspective

Paula Barriendo-Cebrián[1], Nuria Polo[1], Filipa M.B. Lã [2]
[1]Faculty of Philology, UNED, [2]Faculty of Education, UNED

This research investigates the role of fundamental frequency (f_o) in conveying non-verbal information during a conversation. Interlocutors rarely respond solely to words; thus, consider the relationship between discourse and social contexts is necessary [1]. Factors such as social distance, power relationships, and imposition levels between speakers are crucial to shaping these interactions [2]. Social expectations, including sex and gender-specific speaking styles, further influence conversational dynamics. In informal verbal interactions, speakers typically adjust their speech features to the gender of the pairing (male-male, female-female, male-female) and to the interlocutors' power and familiarity [3].

Previous studies have focused on conversational analysis such as, relative talkativeness, speech intrusions, interruptions, overlaps, assent production, and turn-taking. However, vocal characteristics, such as variations in mean f_o and related parameters, remain largely unexplored, particularly among Spanish-speaking population. This study aims at determining whether f_o of young native Spanish speakers depends on their partner's gender, power status, and/or familiarity level.

One professional Spanish actor and one actress in their twenties were recorded in a soundproof booth. They were instructed to enact six roles in daily conversational scenarios: (1) "unfamiliar-powerful," (2) "unfamiliar-powerless," (3) "unfamiliar-equal power," (4) "familiar-powerful," (5) "familiar-powerless," and (6) "familiar-equal power." These roles were designed to motivate specific linguistic choices in listeners' responses. The listeners were 42 Spanish native speaking young adults (20 males and 22 females). Each listener was prompted to use the words 'lavandería' (laundry) and 'domingo' (Sunday) to standardize responses across situations. The spoken interactions were captured using an omnidirectional headset microphone, and the mean f_o and related parameters (standard deviation, maximum, minimum, range) were extracted using a Praat script.

The preliminary findings indicate that f_o and its related measures vary more with conversational contexts than with the speaker's gender. When talking to a stranger, male speakers with a powerless position significantly decreased mean f_o ($p = 0.009$), its variation ($p = 0.004$), its maximum values ($p < 0.001$) and range ($p = 0.017$). On the other hand, females exhibited a decrease in both mean f_o ($p = 0.019$), its variation ($p = 0.08$) and its range ($p = 0.015$) but when in a powerful position. When interacting with an acquaintance, both male and female speakers show a similar trend: to increase the same parameters when in a situation with power.

Overall, it seems that the sex of the partner is not relevant to conversational dynamics for young Spanish native speakers. However, social power and degree of familiarity are crucial in mediating f_o and related parameters.

[1] Norris, S. (2011). Identity in (inter)action: Introducing multimodal (inter)action analysis. Berlin, Germany: De Gruyter.
[2] Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (No. 4). Cambridge university press.
[3] Holmes, J., & King, B. W. (2017). Gender and Sociopragmatics 1. In *The Routledge handbook of pragmatics* (pp. 121-138). Routledge.

2. Phonological phrase in Spanish: the role of fundamental frequency as an acoustic cue

Mario Casado Mancebo
Universidad Nacional de Educación a Distancia

Prosodic structure is encoded by speakers through a range of different articulatory and acoustic cues, i.e., gestural stretching, time lengthening (Cho, 2016) and pitch movements (Baek, 2019) among other possibilities. In Spanish, (Lahoz-Bengoechea, 2015) confirmed the presence of prosodic cues to phonological word boundaries from a production perspective. The influence of phonological phrase boundary in phonological operations that involve segmental changes has also been studied in Spanish (Polo Cano & Elordieta, 2016). This work focuses on fundamental frequency (f_o) as one of these cues and aims to study how phonological phrase boundaries model f_o contours. This task is completed using the Autosegmental-Metrical approach (Pierrehumbert & Beckman, 1988) and Sp-ToBI notation (Estebas-Vilaplana & Prieto, 2010).

30 Spanish speakers from Madrid were recorded in a booth reading aloud 8 text frames with 60 target sentences. Those sentences were grouped in pairs that had the same two syllables before and after a phonological word boundary (PW) and a phonological phrase boundary (PP). For instance, (a) El algo*dón de*cente no causa esos problemas, (b) Las débiles fibras de ese algo*dón de*jarán bolas al lavarlo (*Quality cotton doesn't cause those issues*, *That cotton's weak fibers will bobble after washing it*), where syllables *dón* and *de* appear across a PW boundary (a) and a PP boundary (b). After segmenting each sample with Praat (Boersma & Weenink, 2023), f_o data were extracted from the words before and after each boundary using Python's Parselmouth package (Jadoul et al., 2018). The whole set of f_o contours was processed in R using Functional Principal Components Analysis (Gubian et al., 2015) and Linear Mixed Models (LMM) (Bates et al., 2024) in order to understand the effects of PP boundaries. Some speakers produced a pause after the PP boundary. Samples containing a pause are included as a separate boundary category. Consequently, three boundary categories are modelled – *PW, PP* and *PP+pause*. Other categories that have been proven to have an influence on f_o contours such as speaker sex (Henton, 1989) and speech rate (Torres & Fletcher, 2020) are also controlled in LMM.

Results show that f_o countours are shaped differently across PW and PP boundaries as shown in fig. 1 below. On the one hand, PW boundaries display an L+<H* pattern with the high tone delayed with respect to its bearing unit. On the other hand, PP boundaries display an L+H* pattern with an early peak aligned with its bearing unit. However, results do not show differences between those PP contours with or without a pause. Both PP boundaries with and without a pause showed an L+H* contour with their peaks aligned with their bearing unit.

In each of the three boundary categories (PW, PP, PP+pause) a basic L+H pattern was found as is expected in Spanish declarative sentences (Estebas-Vilaplana & Prieto, 2008). Focusing on each category allowed us to observe that prosodic strengthening on f_o contours does not increase the magnitude of the acoustic cue (i. e., higher frequency) as happens with other cues such as duration. All H tones (PW, PP and PP+pause) are estimated around 125 Hz by the model (Fig. 1). Results showed that prosodic strengthening on f_o affects time axis with an early alignment of H tones in the context of a PP boundary. Early alignment of tones is described in previous studies (Pierrehumbert, 1980) in relation to a delimitative function (Hirst, 2022) that would help listeners to parse major boundaries. As for L tones, frequency does not result as a relevant cue. All L tones (PW, PP and PP+pause) are estimated around 114 Hz (Fig. 1). Likewise, LMM models showed that they keep the same alignment across conditions as is described in previous studies for different languages (Santen & Hirschberg, 1994) for American English, (Prieto et al., 1995) for Spanish, (Caspers & Heuven, 1993) for Ducth, (Xu, 1998) for Chinese, (Igarashi, 2004) for Russian among others). Finally, the fact that samples with a pause could not be set as a separate category is not unexpected as they do not always signal prosodic boundaries. The presence of pauses has been shown to be determined by very different reasons such as hesitations (Shriberg, 2001) and speech rate (Crystal & House, 1990). Therefore, pauses found in this analysis may be related to any of these factors and not to a higher prosodic boundary.

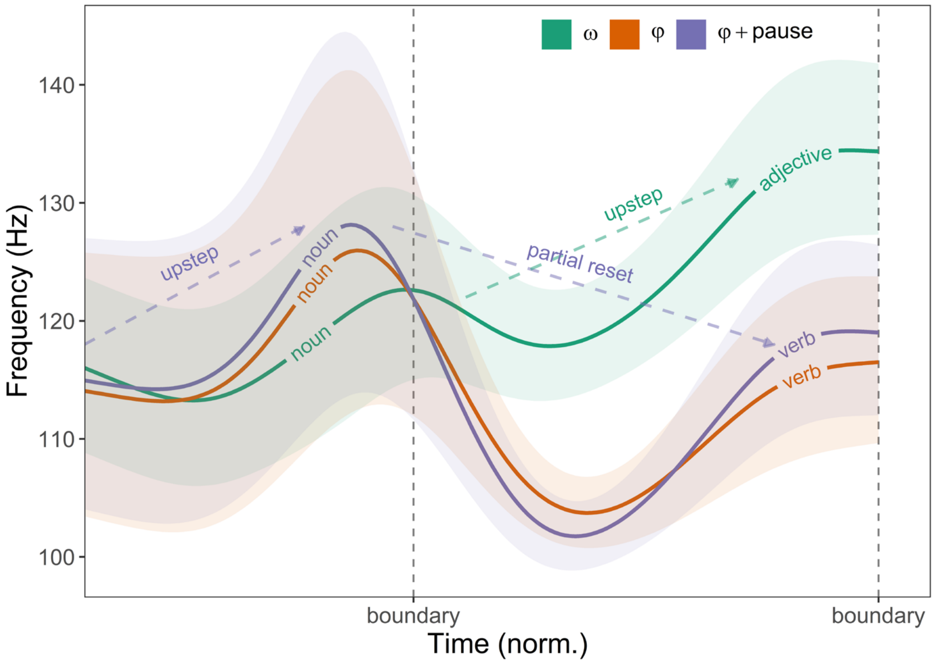


Fig. 1*: Line plot with LMM estimations for each category f0 contours.*

Baek, H. (2019). A cross-linguistic comparison on the use of prosodic cues for ambiguity resolution. 36, 060005. https://doi.org/10.1121/2.0001094

Bates, D., Maechler, M., Bolker [aut, B., & Walker, S. (2024). *lme4: Linear Mixed-Effects Models using «Eigen» and S4* (Versión 1.1-35.5) [Software]. https://cran.r-project.org/package=lme4

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Versión 6.3.08) [Software]. http://www.praat.org/

Caspers, J., & Heuven, V. J. van. (1993). Effects of Time Pressure on the Phonetic Realization of the Dutch Accent-Lending Pitch Rise and Fall. *Phonetica*, *50*(3), 161-171. https://doi.org/10.1159/000261936

Cho, T. (2016). Prosodic Boundary Strengthening in the Phonetics–Prosody Interface. *Language and Linguistics Compass*, *10*(3), 120-141. https://doi.org/10.1111/lnc3.12178

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, *88*(1), 101-112. https://doi.org/10.1121/1.399955

Estebas-Vilaplana, E., & Prieto, P. (2008). La notación prosódica del español: Una revisión del Sp-ToBI. *Estudios de fonética experimental*, 264-283.

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. En *Transcription of Intonation of the Spanish Language* (pp. 17-48). Lincom Europa.

Gubian, M., Torreira, F., & Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, *49*, 16-40. https://doi.org/10.1016/j.wocn.2014.10.001

Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. *Language & Communication*, *9*(4), 299-311. https://doi.org/10.1016/0271-5309(89)90026-8

Hirst, D. (2022). A Multilevel, Multilingual Approach to the Annotation and Representation of Speech Prosody. En J. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic Theory and Practice* (pp. 117-149). The MIT Press. https://doi.org/10.7551/mitpress/10413.003.0006

Igarashi, Y. (2004). «Segmental anchoring» of F0 under changes in speech rate: Evidence from Russian. *Speech Prosody 2004*, 25-28. https://doi.org/10.21437/SpeechProsody.2004-5

Jadoul, Y., Thompson, B., & Boer, B. de. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1-15. https://doi.org/10.1016/j.wocn.2018.07.001

Lahoz-Bengoechea, J. M. (2015). *Fonética y fonología de los fenómenos de refuerzo consonántico en el seno de las unidades léxicas en español* [Tesis doctoral, Universidad Complutense de Madrid]. https://hdl.handle.net/20.500.14352/38657

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [PhD Thesis, Massachusetts Institute of Technology]. http://www.phon.ox.ac.uk/jpierrehumbert/publications/Pierrehumbert_PhD.pdf

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese Tone Structure*. MIT Press. https://mitpress.mit.edu/9780262660631/japanese-tone-structure/

Polo Cano, N., & Elordieta, G. (2016). Evidencia segmental del sintagma fonológico en español. *LEA: Lingüística española actual*, *38*(1), 43-67.

Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, *23*(4), 429-451. https://doi.org/10.1006/jpho.1995.0032

Santen, J. P. H. van, & Hirschberg, J. (1994). Segmental effects on timing and height of pitch contours. *Proc. 3rd International Conference on Spoken Language Processing*, 719-722. https://doi.org/10.21437/ICSLP.1994-182

Shriberg, E. (2001). To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*(1), 153-169. https://doi.org/10.1017/S0025100301001128

Torres, C., & Fletcher, J. (2020). The alignment of F0 tonal targets under changes in speech rate in Drehu. *The Journal of the Acoustical Society of America*, *147*(4), 2947-2958. https://doi.org/10.1121/10.0001006

Xu, Y. (1998). Consistency of Tone-Syllable Alignment across Different Syllable Structures and Speaking Rates. *Phonetica*, *55*(4), 179-203. https://doi.org/10.1159/000028432

3. “Stress-deafness” variation in Singapore English listeners

Adam J. Chong
Queen Mary University of London

Singapore English (SgE) is a Contact English variety [1] spoken alongside at least three other languages – Mandarin Chinese, Malay and Tamil. These three languages are typically spoken by the three main ethnic groups, Chinese, Malays and Indians, respectively. Previous work has found ethnicity-related variation in lexis (e.g. choice of discourse particles; [2]) and segmental productions [3, 4]. In this study, we examine whether ethnicity-related variation also occurs in the SgE’s prosodic system [5]. Recent work has found evidence for f0 differences in production based on stress contrasts (based on British English stress placement) [6]. However, not all speakers in [6] showed a distinction in their f0 patterns between words with different stress placement. This variability in production raises the question of what the status of lexical prominence representations in SgE are and whether these are uniform across the population.

In this study, we use a perception task to examine the status of word-level prominence in SgE. 112 bilingual SgE listeners (54 Chinese, 29 Malay and 29 Indian; age range: 21-37, mean: 25.3) participated in a Sequence Recall Task (SRT), designed to tap into phonological encoding, to examine their ability to discriminate world-level prominence contrasts. To allow for direct comparability with previous results from other languages, we used the SRT design and stimuli from [7], implemented online using jsPsych [8]. Listeners were trained on two contrasts: (1) segmental [ˈmuku]~[ˈmunu], and (2) prosodic [ˈnumi]~[nuˈmi]. In the test phase, listeners had to recall sequences of these words (3, 4, or 5-word sequences) (e.g. [ˈmuku]-[ˈmunu]-[ˈmuku]-[ˈmunu]). In total, there were 60 test sequences, with 30 test sequences for each contrast (segmental and prosodic). Half of these were produced by 2 Dutch speakers and the other half by 2 Persian speakers.

Accuracy for SgE listeners by ethnicity (red box) is shown in Figure 1 alongside overall accuracy from listeners from the 5 languages examined by [7] (blue box) - French, Persian and Indonesian (‘stress-deaf’ (SD) languages) and Dutch, Japanese (‘non-stress-deaf’ (non-SD) languages). SgE listeners’ accuracy on the Segmental contrast did not differ significantly from listeners from the 5 languages from [7]. On the Prosodic contrast, however, SgE listeners were significantly less accurate than Dutch ($p = 0.0013$) and Japanese listeners ($p = 0.0004$; non-SD languages) but not significantly different from French, Indonesian and Persian ($p > 0.05$ on all comparisons; SD languages). Within SgE, Chinese SgE listeners were more accurate than Malay ($p = 0.003$) and Indian SgE ($p = 0.03$) listeners on the Prosodic contrast. Interestingly, while Malay and Indian SgE listeners’ performance patterned with listeners from the SD group, Chinese SgE listeners showed behaviour in between listeners from the non-SD and SD groups.

The results indicate that SgE on the whole lacks contrastive lexical prominence, given SgE listeners pattern with listeners from languages argued to be ‘stress-deaf’. Within SgE, Chinese listeners show less of a stress-deafness effect than listeners of other ethnicities, and are in between SD and non-SD languages. While similar in-between performance on the same task elsewhere has been attributed to English dominance [9], Chinese SgE listeners’ better performance here is unlikely due to English dominance alone. Almost all our listeners regardless of ethnicity reported being more English dominant and reported acquiring both English and their other language from a young age, as is typical. We suggest instead that (at least some) Chinese listeners have transferred their use of (lexical) pitch cues from Mandarin Chinese to do better on the task [10]. Taken together, the variability in both production [6] and perception suggests that the degree to which lexical prominence is contrastively encoded might vary across the SgE population. The implications of these results for modelling intonation and language acquisition in multilingual dialectal contexts are also discussed.

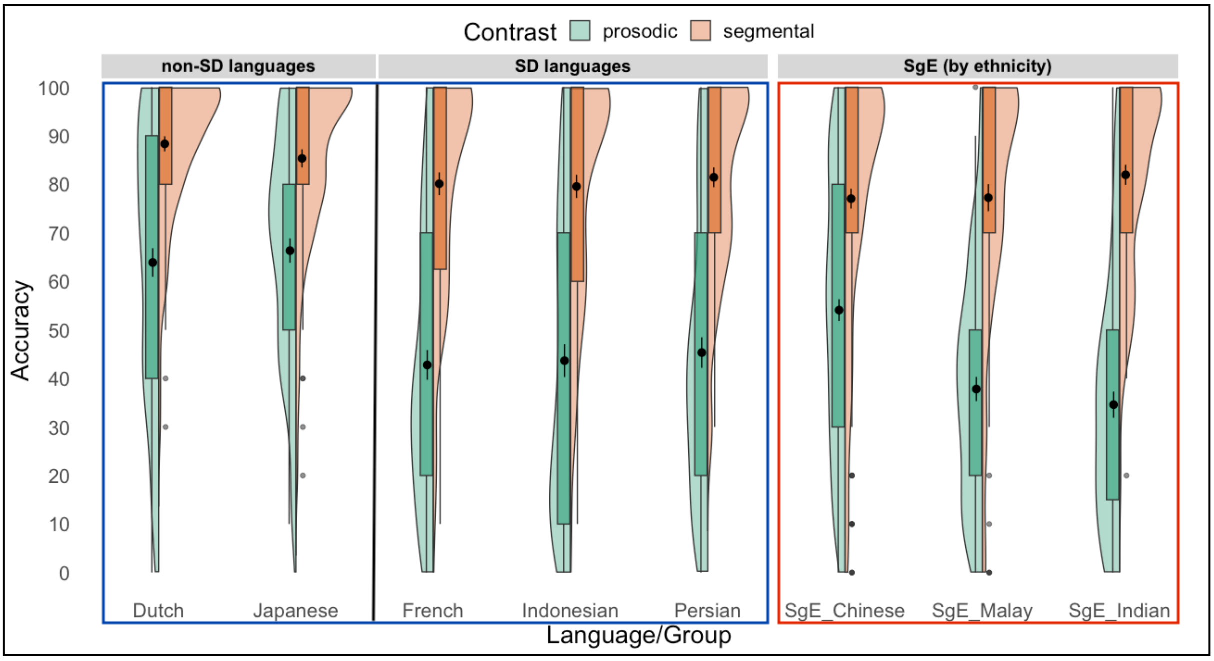


Figure 1. *Raw accuracy scores by language group (SgE split by ethnicity – red box; languages from Rahmani et al. (2015) are in the blue box) and contrast (prosodic vs. segmental). Accuracy is collapsed across sequence length. SD = "stress-deaf" languages; non-SD = "non-stress-deaf" languages.*

[1] Grice, M., German, J. S., & Warren, P. 2020. Intonation systems across varieties of English. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language prosody* (pp. 285–302). Oxford: Oxford University Press.

[2] Alsagoff, L. 2007. Singlish: Negotiating culture, capital and identity. In V. Vaish, S. Gopinathan, & Y. Liu (Eds.), *Language, Capital, Culture* (pp. 24–46). Rotterdam: Sense Publishers.

[3] Lim, L. (Ed.). (2004). *Singapore English: A grammatical description*. Amsterdam: John Benjamins.

[4] Sim, J. H. (2021). Sociophonetic variation in English /l/ in the child-directed speech of English-Malay bilinguals. *Journal of Phonetics*, 88.

[5] Tan, Y.-Y. (2010). Singing the same tune? Prosodic norming in bilingual Singaporeans. In M. C. Ferreira (Ed.), *Multilingual Norms* (pp. 173–194). Frankfurt: Lang.

[6] Chong, A.J. & German, J.S. 2023. Prominence and intonation in Singapore English. *Journal of Phonetics*, 98.

[7] Rahmani, H., Rietveld, T. & Gussenhoven, G. 2015. Stress 'deafness' reveals absence of lexical marking of stress or tone in the adult grammar. *PLoS One*, 10(12), e0143968.

[8] de Leeuw, J.R., Gilbert, R.A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, 8(85), 5351.

[9] Lialiou, M., Bruggeman, A., Vella, A., Grech, S., Schumacher, P. & Grice, M. (2023). World-level prominence and 'stress deafness' in Maltese-English bilinguals. *Proceedings of ICPHS 2023.*

[10] Qin, Z., Chien, Y.-F. & Tremblay, A. (2017). Processing word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics*, 38, 541-570.

4. Accessing the content of factive embedded clauses: interacting effects of intonational tunes and continuation

Timo Buchholz
Universität zu Köln

Factive constructions like (1) can introduce the content of their embedded clauses into the discourse and at the same time treat it as presupposed, i.e. as if it was in the common ground. Yet whether factive embedded content is indeed presupposed has been questioned: both its projection behaviour and whether it is seen as backgrounded has been shown to vary considerably depending on the type of embedding verb ([1],[2]), and to a small extent depending on whether the matrix clause or the embedded clause is prosodically focused ([3],[4]). A closely related question is how accessible such content is for anaphoric uptake. We present experimental results on the effect of prosody on the accessibility of factive embedded clause content from a propositional anaphor resolution task in German. Unlike [3]&[4], we holistically manipulated the (potentially) meaningful contours realized on the embedded clause, and crossed this with a manipulation of continuation that matched or mismatched the meaning of the contours. Our study thus contributes to two questions: 1) how does prosody beyond focus marking affect the anaphoric accessibility of clauses/propositions? 2) to what extent does a proposed "meaning" of contours contribute to this?

As contour meanings, we chose "surprise" and "obviousness", because they are an intuitively graspable antonym pair for which a number of predicates with similar lexical meaning exist. They have been broadly used in the description of German intonation, along with proposed candidate contours ([5]-[8], cf. [9] for Spanish), but there is no fully established consensus on how "surprise" and "obviousness" are realized on declaratives in German. We ran a pilot experiment using semantic scales (13 participants) to determine good candidate contours for the main experiment. (1) gives broad GToBI-style approximations for the *contour type* used as prosodic conditions in the main experiment. Besides the prosodic condition, we manipulated the *continuation* to be an equivalent of either *that is not expected* or *that was expected* (cf. (2)) and crossed it with prosody for a 2x3 design on 24 items with 12 different factive matrix verbs. Participants chose which of the two clauses (matrix or embedded) the anaphor *das* in the continuation referred to. Crossing prosody with continuation was intended to decide between two possible accounts: A) If the *meaning* of the contours contributes to propositional accessibility, we might expect an interaction between *contour type* and *continuation*, e.g. increasing resolution to the embedded clause only when there is a meaning match between contour and continuation. B) Alternatively, the non-neutral contours (1b & c) might increase resolution to the embedded clause purely by being prosodically more prominent (cf. [10, 11]), in which case we should not see an interaction with *continuation*.

Besides an effect of *continuation*, statistical results indicate a reliable interaction between prosody and continuation for the *surprising* contour: it increased resolution to the embedded clause more with an *expected* than with an *unexpected* continuation (see Fig. 1). This speaks against the account based purely on prosodic prominence, but is also are not fully in line with the meaning-based account. Our tentative explanation is as follows: 1) the difference due to *continuation* suggests that the content of a factive embedded clause is not per se inaccessible. Instead, an *unexpected* continuation might make it more inaccessible due to pragmatic incoherence: if a speaker just introduced some content as belonging to the common ground, it is then contradictory to follow this up by saying that it is surprising, while saying that it is expected seems more coherent (albeit redundant). 2) the prosodic contribution might come down to a difference in boundary tone: only the *surprising* contour has a final rise/high tone, and such a final rise on a factive embedded clause might counteract its interpretation as presupposed. If these factors converge, accessibility to the embedded clause is highest.

matrix clause	embedded clause
	a) H*L L-% („*neutral” contour*)
(1) Jonas bedauert	b) (LH)* (LH)* (L^H)* H-% („*surprise” contour*)
	c) L*H L*H L^H* L-% („*obvious” contour*)
	dass Sören selten Bier trinkt
Jonas deplores	that Sören rarely beer drinks

(2) a) Das ist überraschend *that is surprising* (2) b) Das war bekannt *that was known*

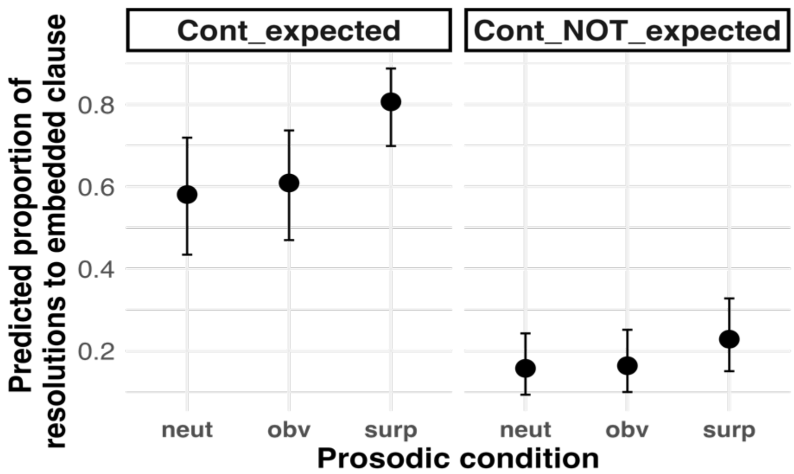


Figure 1. *Predicted resolution to embedded clause per condition combination according to the Bayesian mixed-effect model. Data from 153 participants.*

[1] Degen, J., & Tonhauser, J. (2022). Are there factive predicates? An empirical investigation. *Language*, *98*(3), 552–591.

[2] Xue, J., & Onea, E. (2011). Correlation between Presupposition Projection and At-issueness: An Empirical Study. *Proceedings of the ESSLLI 2011 Workshop on Projective Meaning*, 171–184..

[3] Tonhauser, J. (2016). Prosodic cues to presupposition projection. *Semantics and Linguistic Theory*, 26, 934–960.

[4] Djärv, K., & Bacovcin, H. A. (2020). Prosodic effects on factive presupposition projection. *Journal of Pragmatics*, 169, 61–85.

[5] Grice, M., Baumann, S., & Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, 55-83.

[6] Niebuhr, O. (2008). Perzeption und kognitive Verarbeitung der Sprechmelodie: Theoretische Grundlagen und empirische Untersuchungen. Walter de Gruyter.

[7] Michalsky, J. (2017). Frageintonation im Deutschen: Zur intonatorischen Markierung von Interrogativität und Fragehaltigkeit. De Gruyter.

[8] Zahner-Ritter, K., Einfeldt, M., Wochner, D., James, A., Dehé, N., & Braun, B. (2022). Three Kinds of Rising-Falling Contours in German wh-Questions: Evidence From Form and Function. *Frontiers in Communication*, 7, 838955.

[9] Fliessbach, J. (2023). The intonation of expectations. On marked declaratives, exclamatives, and discourse particles in Castilian Spanish. Language Science Press.

[10] Baumann, S., & Röhr, C. T. (2015). The perceptual prominence of pitch accent types in German. *Proceedings of the 18th ICPhS*, 298.

[11] Röhr, C. T., Baumann, S., & Grice, M. (2022). The influence of expectations on tonal cues to prominence. *Journal of Phonetics*, 94, 101174.

5. Prenuclear variation and nuclear contours in negative and positive polar questions

Johannes Heim[1], Judith Schlenter[2], Sophie Repp[2]
[1]University of Aberdeen, [2]University of Cologne

In pursuit of accounting for question bias, positive polar questions, PQs (1), and negative polar questions, NQs (2), have been compared in their prosodic profiles 000. Key methodological challenges have been the relative infrequency of NQs in corpora of natural conversation, and the limited naturalness of speech in laboratory settings. To approach these challenges and strike a balance between data availability and naturalness, we systematically extracted samples from an American English TV SOAP corpus 0 and tested previous claims about the prosodic characteristics of PQs and NQs regarding nuclear contours and the accentuation of the clause-initial auxiliary, AUX. The existing literature reports variation within and between the two question types. PQs are often reported to have a final rise if information-seeking 0 but they can also occur with falls 0 or plateaus 0. Sometimes these variations are associated with rhetorical or assertive interpretations, but it is an open issue whether falls exclude an information-seeking interpretation 0. Further complication arises from cross-dialectal variation in question rises as they vary in nuclear pitch accents [8]. Accordingly, a wide range of contours are attested for PQs (L* H-H%, L* H-L%, H* H-H%, H* H-L%). NQs are equally heterogenous, albeit with some candidates for potential markers of speaker bias. Depending on where the speaker is leaning toward (*p* or ¬*p*), there may be a correlation of bias with the steepness of a rise 00. The AUX with contracted negation has been found to predominantly occur with a rising L+H* accent [10], which has been interpreted as marking polarity contrast. Yet in PQs, where AUX is often unaccented, there is also polarity contrast 0.

In our paper, we show that previous claims on the prosody of PQs and NQs requires further qualification. In a corpus study of the TV soap *The Bold and the Beautiful*, we identified information-seeking PQs and NQs produced by four actors (2m, 2f). Non-information-seeking candidates were excluded via transcript analysis, probing for an answer (expectation) in the conversation. The final data set (n = 160) contained 20 PQs and 20 NQs per actor. In all questions, the clause-initial AUX (+*n't*) was followed by a subject pronoun. We annotated AUX accents and nuclear contours (MAE-ToBI). Our results for the nuclear contours (Fig. 1a&b), which mostly were rising or contained plateaus, showed no overall differences between PQs and NQs except for somewhat fewer low rises (L* L-H%) in NQs vs. PQs. There were strong individual differences, indicating that individual speakers distinguish between question types, albeit with different preferences: the two male speakers seem to have opposite preferences (Fig 1b). In the PQs, there was overall great contour variation, confirming earlier findings. AUX accentuation (Fig. 2a&b) was low in PQs (very low for male speakers), also confirming earlier findings. However, unlike previous reports, AUX in NQs did not have predominantly L+H*, but showed substantial variation: (i) there were just as many H* and L*+H accents; (ii) we observed an accent previously only reported for German, a late-medial peak (L+H)* 00, which was followed by a fall on the subsequent unstressed syllable; (iii) there was what we mark as L*|H in Fig. 2a&b: a rise whose H tone is realized on the unstressed syllable without perceptible transition, resulting in a tonal *step*; (iv) unaccented AUX. The occurrence of L*|H did not correlate with lexical item, i.e., it did not correlate with the amount of segmental material of AUX. Our findings suggest that there is much more variation in the prenuclear region than was previously discussed for NQs and than is generally assumed for prenuclear accents. Our context analysis indicates that the choice of accent may be related to speaker bias: the steeper or later the rise, the stronger the speaker is biased toward their own belief. For nuclear contours, steepness of the rise and height of the final boundary tone (which in NQs tended to be higher) might serve a similar function. We are currently testing this in a perception study.

(1) Are you exhausted? (2) Aren't you exhausted?

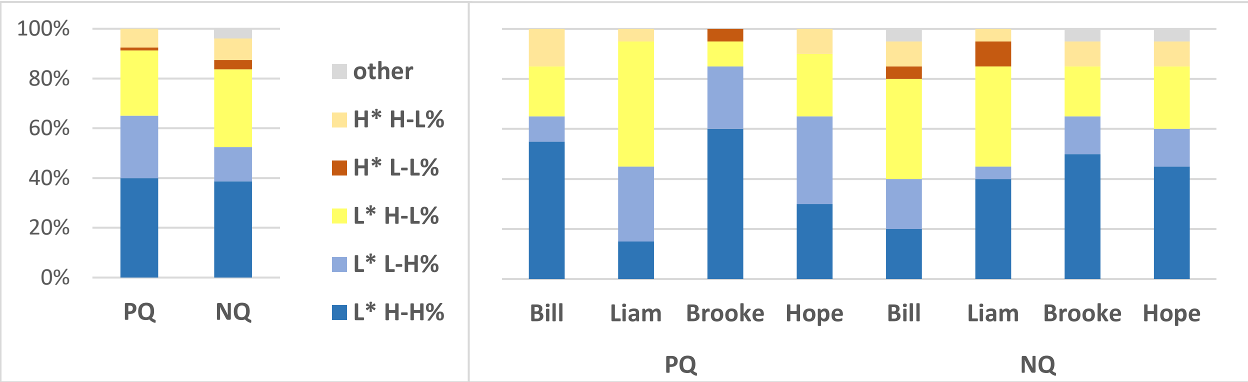


Figure 1a&b. *Proportion of nuclear contours in PQs and NQs: total & by individual*

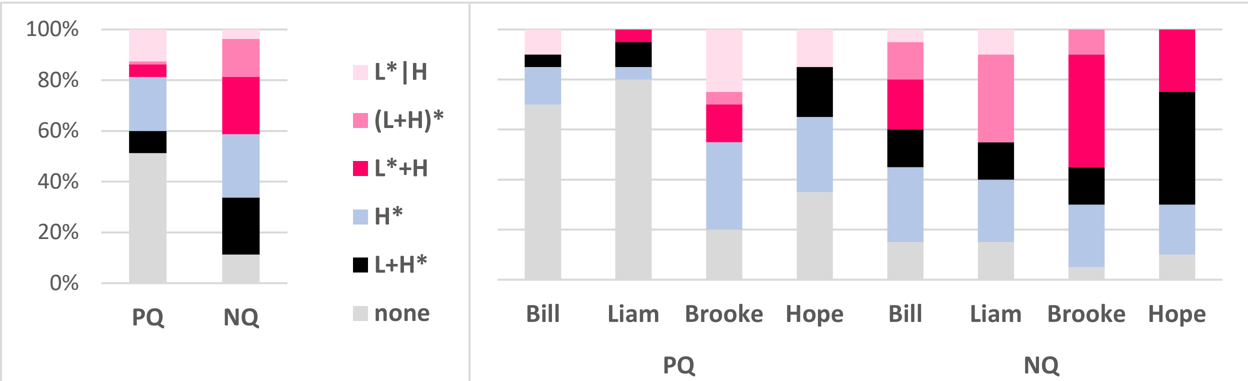


Figure 2a&b. *Proportion of* AUX *accent types in PQs and NQs: total & by individual*

[1] Romero, M. 2024. Biased Polar Questions. *Annual Rev. of Linguistics*, 10(1), 279-302.
[2] Domaneschi, F., Romero, M., & Braun, B. 2017. Bias in polar questions: Evidence from English and German production experiments. *Glossa* 2(1): 26. 1–28.
[3] Arnhold, A., Braun, B., & Romero, M. (2021). Aren't prosody and syntax marking bias in questions? *Language and Speech* 64(1), 141-180.
[4] Davies, M. 2011-. *Corpus of American Soap Operas*. https://www.english-corpora.org/soap/.
[5] Hedberg, N., & Sosa, J. M. 2002. The prosody of questions in natural discourse. In *Speech Prosody 2002, International Conference* (Aix-en-Provence).
[6] Geluykens, R. 1989. R(a)ising questions: Question intonation revisited: A reply to Batliner and Oppenrieder. *Journal of Pragmatics* 13(4), 567-575.
[7] Braun, B., Dehé, N., Neitsch, J., Wochner, D., & Zahner, K. 2019. The prosody of rhetorical and information seeking questions in German. Language & Speech, 62(4), 779–807.
[8] Nilsenová, M. (2006). Rises and falls. Studies in the semantics and pragmatics of intonation. University of Amsterdam.
[9] Armstrong, M., Piccinini, P., & Ritchart, A. (2015). The phonetics and distribution of non-question rises in two varieties of American English. ICPhS.
[10] Zahner-Ritter, K., Einfeldt, M., Wochner, D., James, A., Dehé, N., & Braun, B. 2022. Three kinds of rising-falling contours in German wh-questions: Evidence from form and function. *Frontiers in Communication* 7, 838955.

6. Local and global effects on prosodic prominence relations in German

Stefan Baumann, Janne Lorenzen
University of Cologne

In an exploratory production study on German, we investigate the influence of four semantic-pragmatic and syntactic factors on prosodic prominence, namely referential (r-) and lexical (l-) newness/givenness (Baumann & Riester 2013), grammatical role, and position of a referential target word within a sentence. We collected data from 20 speakers in a listener-directed reading task. Each speaker produced eight stories with five target words in different conditions (see example in (1)). Table 1 shows the predicted impact of the four factors and the resulting prominence weightings of the target words which we compared with their prosodic realizations. We considered several categorical and continuous prosodic correlates of prominence. In the present paper, we will concentrate on the placement of the nuclear accent providing insight into the prosodic strength relations of constituents in an utterance at the syntagmatic level (prosodic annotations by two independent transcribers following DIMA).

The results (Fig.1A) reveal a probabilistic distribution of the target words' accent status which generally confirms our predicted prominence ranking - with the exception that Target5 is produced as prosodically more prominent than Target2. That is, we find an additive influence of the discourse-related and syntactic cues, with lexical newness and initial sentence position showing the strongest boosting effects on a target word's prosodic prominence. In fact, there is evidence from previous studies that initial accents are placed less selectively and mainly for rhythmic reasons (cf. Bolinger's (1986) *accents of power* as well as Baumann et al. 2021), since many referring expressions carry an accent irrespective of being *given* (as in Target3 and Target5 in our dataset). This to some extent 'automatic' placement of accents contributes to the metrical structure of an utterance as a whole and may be ascribed to a *global*, overarching layer of prosodic prominence distribution. In contrast, the three conditions containing Target1, Target2 and Target4 show that the *local* prominence relation between object and verb has an important impact on the rest of the utterance, because it indicates where the nuclear pitch accent will fall, which in turn is decisive for the information-structural interpretation of the sentence.

However, the local prominence relation between object and verb is not only determined by language-specific rules (in German broad focus, the object is commonly more prominent) and the words' levels of givenness but also by their *semantic weight* (e.g., Bolinger 1972). Bolinger showed for American English that the semantic weight of nouns and verbs influences their probability of being accented: the 'heavier' the more informative and thus the more prominent, i.e., the more likely to attract the nuclear accent (see also Uhmann (1991) for German). In our data, this tendency is evident in the variable accentuation patterns of Target2, which is *l-given* and *r-given* and thus generally prone to deaccentuation. Nevertheless, depending on the semantic weight of the following verb, the accentuation rate of the target word may be relatively high (Fig.1B). For example, in the utterance *Dabei hat Sarah die Gardine gesehen* ("In doing so, Sarah saw the curtain") the verb *gesehen* ("saw") is semantically rather 'light', so that the nuclear accent is more frequently placed on the 'heavier' noun *Gardine* ("curtain"). In comparison, the target word *Praline* ("chocolate candy") in the utterance *Dann hat die Kollegin die Praline ausgepackt* ("After that, the colleague unpacked the chocolate candy") typically only receives a prenuclear accent. The nuclear accent falls on the final verb *ausgepackt* ("unpacked") due to its greater weight, or higher level of informativeness. We thus confirm Bolinger's (1972) observation as one of the central results of our study: What counts in order to appropriately reflect the semantic-pragmatic meaning of an utterance is the *relative* prosodic prominence between the elements, in combination with syntactic cues - since it ultimately determines the position of the nuclear accent.

(1) Maria hat im Büro eine Kollegin, mit der sie sehr gut befreundet ist. Fast jeden Tag hat sie ein kleines Geschenk für Maria dabei, das sie ihr in der Mittagspause überreicht. Heute hat die Kollegin **eine Praline (Target1)** mitgebracht. Maria war noch am Arbeiten, als die Kollegin zu ihr an den Schreibtisch kam. Dann hat die Kollegin **die Praline (Target2)** ausgepackt. Sie war von einer edlen Marke und Maria hat sich sehr gefreut. **Die Praline (Target3)** war mit Karamell und Nüssen gefüllt. Als Kind war sie mit Süßigkeiten sehr wählerisch, was auch mit einem Erlebnis auf ihrem achten Geburtstag zusammenhängt. Damals hat Maria **eine Praline (Target4)** gegessen. Sie hat aber nicht gewusst, dass die manchmal mit Alkohol gefüllt sind. Maria hat sich geekelt und danach sehr lange gar keine Schokolade mehr angerührt. Diese Zeiten sind zum Glück vorbei und heute liebt Maria Schokolade aller Art. **Die Praline (Target5)** wird sie gleich als Vorspeise essen.

Maria has a colleague in the office with whom she is very good friends. Almost every day she brings a small gift for Maria, which she presents to her during the lunch break. Today, the colleague brought **a chocolate candy (Target1)**. Maria was still working when the colleague came to her desk. After that, the colleague unpacked **the chocolate candy (Target2)**. It was from a noble brand and Maria was very happy. **The chocolate candy (Target3)** was filled with caramel and nuts. As a child, she was very picky with sweets, which is related to an experience on her eighth birthday. At that time Maria ate **a chocolate candy (Target4)**. But she didn't know that sometimes they are filled with alcohol. Maria was disgusted and didn't touch any chocolate for a very long time afterwards. Fortunately, these times are over and today Maria loves chocolate of all kinds. **The chocolate candy (Target5)** she will eat right away as an appetizer.

	Target1	**Target2**	**Target3**	**Target4**	**Target5**
Referential givenness	New ↑	↓ Given	↓ Given	New ↑	↓ Given
Lexical givenness	New ↑	↓ Given	↓ Given	↓ Given	↓ Given
Grammatical role	Object ↑	Object ↑	↓ Subject	Object ↑	Object ↑
Sentence position	Medial ↑	Medial ↑	↓ Initial	Medial ↑	↓ Initial
Predicted prominence points	●●●●	●●○○	○○○○	●●●○	●○○○

Table 1. *Predicted influence of discourse-related and syntactic factors on prosodic prominence.*

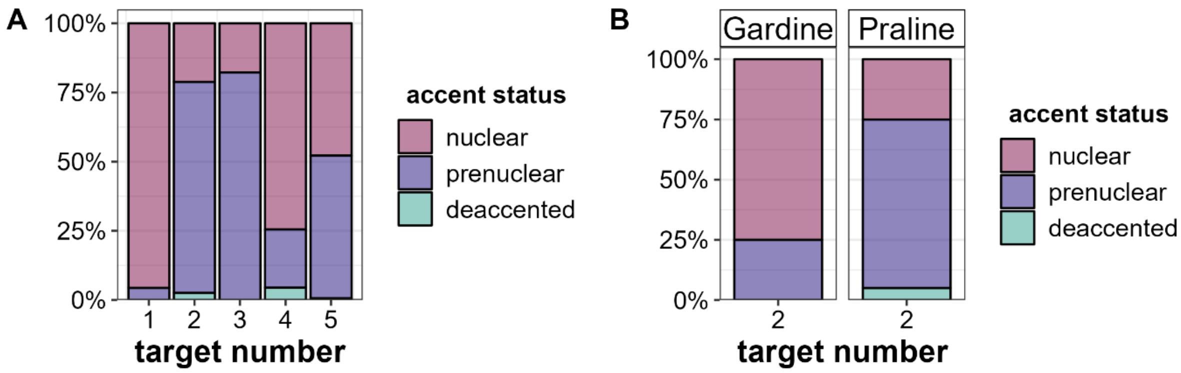


Figure 1. *A: Distribution of accent status across the five target word positions in percent. B: Comparison of accent status distributions in Target2 of Gardine and Praline stories.*

Baumann S., Mertens J., & Kalbertodt J. 2021. The influence of informativeness on the prosody of sentence topics. Glossa 6(1): 95. 1–28. doi: 10.16995/glossa.5871

Baumann, S., & Riester, A. 2013. Coreference, lexical givenness and prosody in German. In Hartmann, J., Radó, J., Winkler, S. (Eds.), Lingua 136, 16–37.

Bolinger D. 1972. Accent is predictable (if you are a mind reader). Language 49, 633–644.

Bolinger, D. 1986. Intonation and Its Parts. Palo Alto: Stanford University Press.
Uhmann S. 1991. Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie. Tübingen: Niemeyer.

7. The role of F0 in cueing sentence stress in Ukrainian

Janina Mołczanow, Beata Łukaszewicz, Anna Łukaszewicz
University of Warsaw

As is well known, F0 is a widely attested acoustic correlate of both word-level and sentence-level prominence. Recent research on Ukrainian ([1], [2], [3]), based on citation forms embedded in a frame, revealed no significant differences in F0 maximum between lexically stressed and unstressed syllables, pointing to increased vocalic duration as the main cue to word-level prominence in accented positions. In addition, mean F0 slopes have been shown to be less steep in lexically stressed syllables than in unstressed syllables, which may be related to the enhanced duration of the former. The failure to detect expected differences in F0 indices between stressed and unstressed syllables may call into question the presence of intonational pitch accents which are associated with a stress-prominent syllable. The present study reappraises the question of whether F0 changes play a role in cueing sentence-level prominence in Ukrainian by (i) zooming in on the F0 shifts across the syllables immediately preceding lexical stress and those in lexically stressed syllables, and (ii) taking into account the existence of polar opposite (rising vs falling) intonational patterns in Ukrainian. The study is based on segmentally identical minimal pairs, which differ only in the position of lexical stress, i.e. σˈσσσ vs σσˈσσ, e.g. [dɔˈbʲiɦatɨ] 'run' (perf.) – [dɔbʲiˈɦatɨ] 'run' (imperf.), occurring in focus position.

14 native speakers of standard Ukrainian (10F, 4M; aged 24-62, M = 39) participated in the experiment. The recordings were done using an H4 Zoom portable recorder set at a sampling rate of 44.1 kHz and an AT897 microphone. In sum, 588 recorded word tokens (14 speakers × 14 words × 3 repetitions) have been obtained. Measurements were conducted in 1176 vowels.

Analyses of F0 max and F0 slope corroborate previous results (Fig. 1). A potential shift in F0 occurring across adjacent syllables is further investigated using the version of Pairwise Variability Index (PVI; [4], [5]; cf. [6]) which expresses the magnitude and direction of local F0 change: $PVI_{F0} = 12*\log_2(F0_k/F0_{k+1}) = ST_k - ST_{k+1}$ (ST stands for 'semitones').

We hypothesise that if focus is expressed by F0 excursions aligned with the position of lexical stress, then different PVI values are expected for the comparison $\boldsymbol{\sigma_1}$ˈ$\boldsymbol{\sigma_2}\sigma_3\sigma_4$ vs $\boldsymbol{\sigma_1\sigma_2}$ˈ$\sigma_3\sigma_4$, but similar PVI values for the comparison $\boldsymbol{\sigma_1}$ˈ$\boldsymbol{\sigma_2}\sigma_3\sigma_4$ vs $\sigma_1\boldsymbol{\sigma_2}$ˈ$\boldsymbol{\sigma_3}\sigma_4$.

Previous descriptions of the Ukrainian intonation [7] as well as a preliminary inspection of the data point to the presence of either a falling or a rising pattern of F0 change in the vicinity of the lexically stressed syllable. We thus divided the data using the k-means classification in SPSS (v. 29), run to fit a two-group model.

The statistical analysis (linear mixed effects models including speaker- and item-specific random slopes) revealed a significantly larger difference in the magnitude of F0 change between the first two syllables in $\boldsymbol{\sigma_1}$ˈ$\boldsymbol{\sigma_2}\sigma_3\sigma_4$ than in $\boldsymbol{\sigma_1\sigma_2}$ˈ$\sigma_3\sigma_4$, both in the falling and in the rising pattern of F0 change ($p < .001$) (Fig. 2). This result indicates that F0 plays a role in signalling sentence-level prominence in Ukrainian. As for the comparison $\boldsymbol{\sigma_1}$ˈ$\boldsymbol{\sigma_2}\sigma_3\sigma_4$ vs $\sigma_1\boldsymbol{\sigma_2}$ˈ$\boldsymbol{\sigma_3}\sigma_4$, no statistically significant difference was observed between words with stress located on the second and the third syllables in the rising intonation group ($p = .09$) (Fig. 3). In turn, the falling intonation group exhibited a small difference of 0.8 ST which turned out to be statistically significant ($p < .05$) (Fig. 3). However, as the difference of 0.8 ST can be considered negligible from the point of view of perception, we conclude that the present findings lend support to the hypothesis that intonational pitch accent is associated with lexically stressed syllables in Ukrainian.

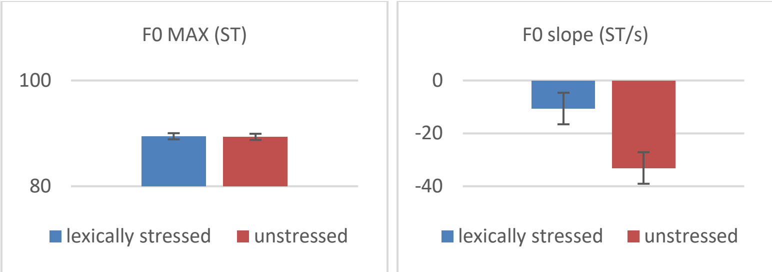


Figure 1. *Mean F0 maximum (left panel) and F0 slope (right panel).*

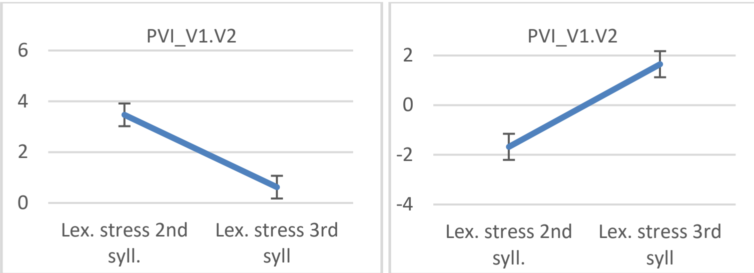


Figure 2. *Mean PVIs of the first two syllables (in* ***σ1*** *ˈ* ***σ2*** *σ3σ4 vs* ***σ1σ2*** *ˈσ3σ4 words) in the group with a falling F0 pattern (left panel) and a rising F0 pattern (right panel).*

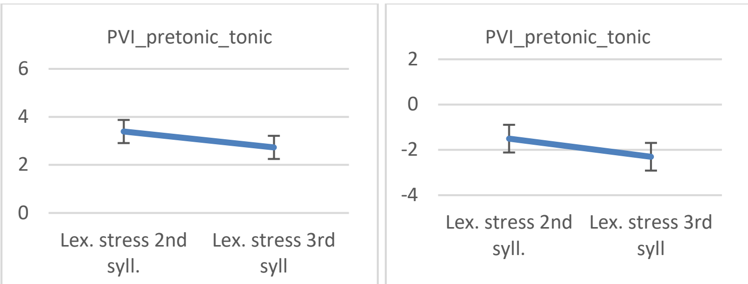


Figure 3. *Mean PVIs of the pretonic and tonic syllables (in* ***σ1*** *ˈ* ***σ2*** *σ3σ4 vs σ1* ***σ2*** *ˈ* ***σ3*** *σ4 words) in the group with a falling F0 pattern (left panel) and a rising F0 pattern (right panel).*

[1] Łukaszewicz, B., & Mołczanow, J. (2025). *Issues in Metrical Phonology: Insights from Ukrainian.* Cambridge: Cambridge University Press.

[2] Łukaszewicz, B., & Mołczanow, J. 2018. Rhythmic stress in Ukrainian: acoustic evidence of a bidirectional system. *Journal of Linguistics* 54, 367-388.

[3] Łukaszewicz, B., & Mołczanow, J. 2018. The role of vowel parameters in defining lexical and subsidiary stress in Ukrainian. *Poznan Studies in Contemporary Linguistics* 54, 355-375.

[4] Arciuli, J., Simpson, B. S., Vogel, A. P., & Ballard, K. J. 2014. Acoustic changes in the production of lexical stress during Lombard speech. *Language and Speech* 57, 149-162.

[5] Łukaszewicz, B. 2018. Phonetic evidence for an iterative stress system: the issue of consonantal rhythm. *Phonology* 35, 115-150.

[6] Low, E. L., Grabe, E., & Nolan, F. 2000. Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43, 377-401.

[7] Nikolaeva, T. M. 1977. *Frazovaja Intonacija Slavjanskix Jazykov*. Moscow: Izdatel'stvo Nauka.

8. Variations of Tone 3 sandhi in spontaneous speech: a comparison between Beijing Mandarin and Taipei Mandarin

Yuxin Lu, Yu-Hsiang Tseng, R. Harald Baayen
Quantiative Linguistics, University of Tübingen

One of the best-known tone sandhi rules in Mandarin Chinese is the Tone 3 sandhi. When Tone 3 (T3, dipping tone) is followed by another T3, the first T3 is pronounced similarly to Tone 2 (T2, rising tone). Thus, disyllabic words with T3-T3 tone pattern, such as 水果 (shui3guo3, 'fruits'), are realized as shui2guo3 in connected speech. Previous research has reported that T3-T3 is incompletely neutralized with T2-T3, with varying degrees of difference between T2-T3 and T3-T3 depending on variety of Mandarin being studied (Tian et al., 2022).

The current study investigates the tonal realization of Tone 3 sandhi in spontaneous speech, by comparing Beijing Mandarin and Taipei Mandarin. The data for Taipei Mandarin were drawn from the Taiwan Mandarin Spontaneous Speech Corpus (Fon, 2004). The data for Beijing Mandarin come from a corpus of conversational speech (Ruan et al., 2018). We selected disyllabic words with T2-T3 and T3-T3 tone patterns. The resulting dataset contains 1667 tokens representing 26 word types for Beijing Mandarin, and 1946 tokens representing 26 word types for Taipei Mandarin. Using Generalized Additive Mixed Models (GAMMs, Wood, 2017), we modeled log-transformed fundamental frequency (f0) with several predictors, including normalized time, speaking rate, word position in the utterance, tonal context, speaker, and word. We fitted GAMMs to four datasets separately: female Beijing, male Beijing, female Taipei, and male Taipei, to control for the effect of gender.

The statistical analyses show that the main effects of normalized time show the similar excursion across four datasets, with greater f0 range in speakers of Taipei Mandarin (see Figure 1). The GAMMs analyses did not support differences between the partial effects of tone pattern (T2-T3/T3-T3), for female Beijing and male Taipei, indicating complete neutralization. The difference was significant differences for male Beijing (p=0.0003) and female Taipei (p=0.0004).

These results contrast with previous research based on carefully controlled laboratory speech, which reported incomplete neutralization. This discrepancy might be attributed to differences in register, as our findings are based on freely spoken Mandarin.

However, another reason for incomplete neutralization having been reported is that previous studies did not take into account the fact that pitch contours in Mandarin are often modulated in word-specific ways. Without controlling for the effect of word, the tone sandhi in our data also appears incomplete for all cases. Figure 2 shows the partial effect of word on pitch contours, as detected by the GAMMs models. These word-specific components are driven by word's semantics (Chuang et al., 2024; Lu et al., 2024) and collocational preferences. For instance, 有点 (*you3dian3*, 'a bit') is mostly preceded by pause and 是 (*shi4*, 'to be') in both Beijing Mandarin and Taipei Mandarin. In contrast, 可以 (*ke3yi3*, 'can') is mostly preceded by *pause*, 也 (*ye3*, 'also'), and 你 (*ni3*, 'you') in Beijing Mandarin, but mostly by *pause* in Taipei Mandarin.

An important direction for future research is to utilize contextualized embeddings to represent words' meanings more precisely, further exploring the relationship between fine phonetic realizations and context-specific semantics.

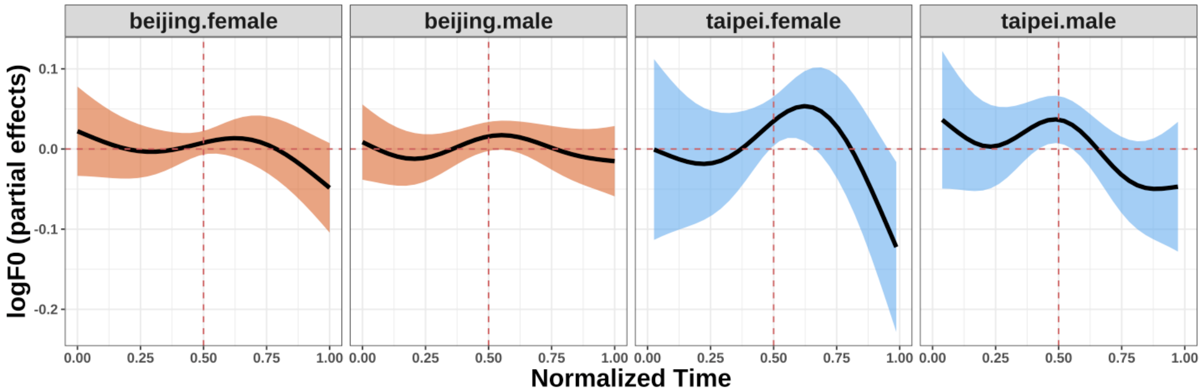


Figure 1. *The main effect of normalized time for T3-T3 tone pattern. The orange curves indicate Beijing Mandarin, and blue curves indicate Taipei Mandarin.*

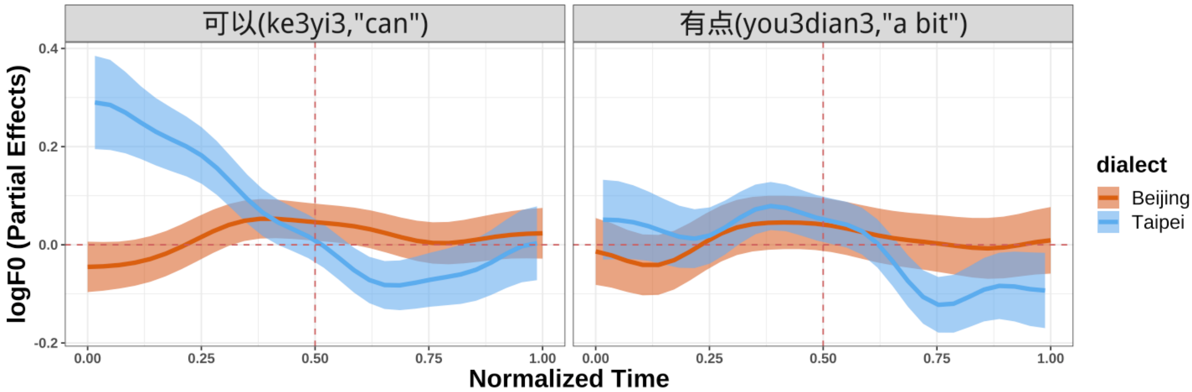


Figure 2: *Selected word-specific components of pitch contours, estimated by factor smooth.*

Chuang, Y.-Y., Bell, M. J., Tseng, Y.-H., and Baayen, R. H. (2024). Word-specific tonal realizations in mandarin. *arXiv preprint arXiv:2405.07006*.

Fon, J. (2004). A preliminary construction of Taiwan Southern Min spontaneous speech corpus. Technical report, Tech. Rep. NSC-92-2411-H-003-050, National Science Council, Taiwan.

Lu, Y., Chuang, Y.-Y., and Baayen, R. H. (2024). Form and meaning co-determine the realization of tone in Taiwan Mandarin spontaneous speech: the case of tone 3 sandhi. *arXiv preprint arXiv:2408.15747*.

Ruan, F., Song, Q., Li, K., and Hao, Y. (2018). Definition of corpus, scripts, standards and specifications of environment/speaker coverage for Mandarin languages. Technical report, Beijing Haitian Ruisheng Science Technology Ltd.

Tian, Z., Dong, X., Gao, F., Wang, H., and Lin, C.-J. C. (2022). Mandarin tone sandhi realization: Evidence from large speech corpora. In *INTERSPEECH*, pages 5273–5277.

Wood, S. N. (2017). Generalized additive models: an introduction with R. CRC press.

9. How do French listeners perceive English intonation? The case of the fall-rise contour

Antoine Régis[1], Sophie Dufour[2], Sophie Herment[3], Amandine Michelas[4]
[1, 2, 3, 4]*Aix-Marseille University, CNRS, LPL, UMR 7309, Aix-en-Provence, France*

While the influence of the native language on the perception of segments is well established, the influence of the native language on the perception of suprasegmental features such as intonation contours is less known. Previous studies have suggested that the processing of intonation would be less sensitive to native language experience than segmental processing because it would be based mostly on general auditory mechanisms [1]. Grabe and colleagues [1] compared the ability of English, Spanish and Chinese listeners to discriminate between several types of falling and rising English intonation contours. While all groups of listeners were able to distinguish between falls and rises, no differences were observable between different rises. By contrast, a more recent study [2] showed that listeners can discriminate between different types of rises provided that these rises have a linguistic function in their native language. The perception of intonation would thus be more influenced by the existing inventory of contours in the native language than on general auditory mechanisms.

In this study, we provided a more in-depth examination of the hypothesis according to which the existing inventory of intonation contours in the native language influences the perception of intonation. To do so, we investigated how French and English speakers discriminate two types of rising contours: a simple rise (L*H-H%) and a fall-rise (H*L-H%, see [3, 4, 5]). While the simple rise exists both in English and in French, the fall-rise contour (also called tone 2 in Halliday's system [6]) is common in both British and American English to indicate doubt or disapproval ([6]) but it does not exist in the French inventory of contours.

Two groups of participants, a group of native speakers of American English and a group of French native speakers performed an ABX task. In this task, participants heard nonwords created by a concatenation of two English words (ex. the nonword *maridy* coming from *mariner* and *melody*) produced by three different American English speakers and had to indicate whether X was identical to A or B. The A and B stimuli varied in their intonation contour and were produced as either a simple rise (L*H-H%) or a fall-rise contour (H*L-H%, Figure 1).

The results showed that French listeners had significantly more difficulties than American English listeners to perform the ABX discrimination task based on intonation contours (Figure 2). Because the fall-rise contour exists in the American English inventory of contours but not in the French inventory, the difficulties observed for the French listeners argue in favour of the claim that the existing inventory of intonation contours in the native language influences the perception of intonation. In a more general way, our results constitute a further demonstration that the inventory of contours in the native language shapes the way listeners perceive intonation contours [2], just as the inventory of phonemes shapes the perception of phonemes [7].

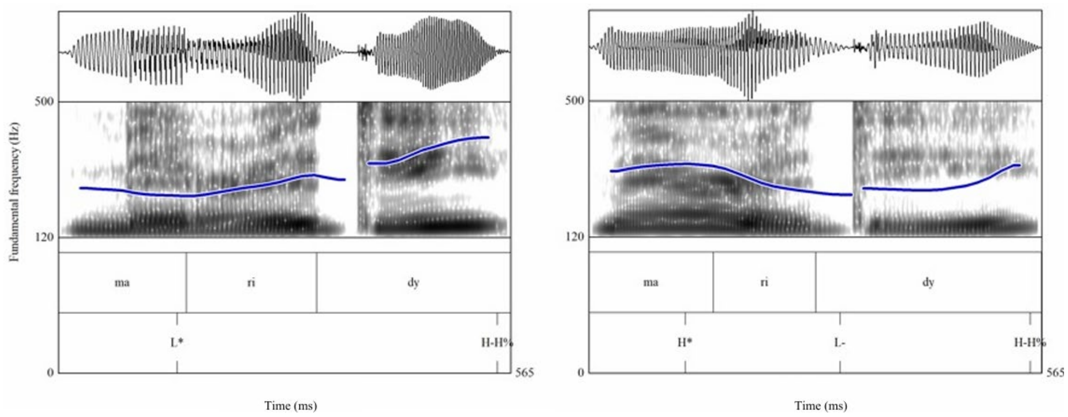


Figure 1. *Phonemic and prosodic profile for the nonword maridy produced by a female voice with a simple rising contour (left panel) and with a fall-rise contour (right panel).*

Figure 2. *Percentage of correct responses for English and French listeners. Error bars represent standard error.*

[1] Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, *46*(4), 375-401.

[2] Schmidt, E., Kung, C., Post, B., Yuen, I., & Demuth, K. (2016). L1 experience shapes the perception of intonational contours. *Proc. Speech Prosody 2016*, 242-246.

[3] Silverman, K. E., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P., Pierrehumbert, J., Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *ICSLP,* (2), 867-870.

[4] Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* [Doctoral dissertation, Bloomington/Indiana University Linguistics Club].

[5] Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, *17*(2), 143-185.

[6] Halliday, M. A. K. (1970). *A course in spoken English: intonation*. London: Oxford University Press.

[7] Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109*(2), 775-794.

10. Variability and task effects in the intonation of fronted Wh-questions in Basque Spanish

Carolina González, Lara Reglero, Susan Cox
Florida State University

Our study examines the intonational variation of wh-questions in Basque Spanish, and specifically the effect of task type on the realization of initial pitch accents and nuclear configurations. Across Spanish dialects, wh-questions usually involve fronting of the wh-word to sentence-initial position. Intonationally, they remain understudied (Henriksen 2014). Neutral wh-questions reportedly show a 'slide' pattern, with an initial rise and a gradual declination (Navarro Tomás 1944, Quilis 1993, RAE 2011). Variation occurs across dialects (Sosa 1999, Hualde & Prieto 2015). For Basque Spanish, Gaminde et al. (2014) reports a high (H*) initial pitch accent and a prevalence of final rises, particularly for Basque-dominant speakers. Delgado (2024) also finds widespread final rises independently of language dominance. Both studies are based on reading tasks, which tend to involve more careful speech that might differ in several ways from the intonation of more casual speaking styles (Face 2003, Henriksen 2009).

Our study compares the intonation of neutral, fronted wh-questions in Basque Spanish in reading and elicitation tasks. A total of 190 fronted questions from 19 Spanish/Basque bilinguals was analyzed in Praat following Spanish ToBI conventions (Aguilar et al. 2024). In the reading task, consistent with Gaminde et al. (2014) and Delgado (2024), most questions (88%) involve final rises (typically L* H%). Final falls are attested in Spanish-dominant participants only. On the other hand, in the elicitation task final falls are prevalent (63%, typically L*L%), and only 37% of questions showed a final rise. Our results also show wider variation in the realization of nuclear pitch accents in the elicitation task, with L+H* present in 15% of the questions analyzed. In contrast, in the reading task L+H* only occurred in 3% of cases.

In addition, the initial pitch accent was mostly realized as a delayed rise in the reading task (L+<H*; 77% of cases). A high pitch accent (H*) occurred in 19% of cases, most typically in Spanish-dominant speakers and monosyllabic wh-words (such as 'qué' *what*). In contrast, in the elicitation task delayed rises were less prevalent (55%), with high and non-delayed rising accents also common (23%, 19%). Our results attest to more variability in the intonation of wh-questions in Basque Spanish compared with previous studies. Our findings are also in line with studies on Spanish (Face 2003, Henriksen 2009) and other languages (Cruttenden 2007, Savino 2012), which show that the specific task used in data collection has an impact on the intonational contours obtained.

Aguilar, L.: Roseano, P.; Vanrell-De-la-Mota, M., & P. Prieto. 2024. *Sp_ToBI Traning Materials*. https://sp-tobi.upf.edu

Cruttenden, A. 2007. Intonational diglossia: a case study of Glasgow. *Journal of the International Phonetic Association* 37. 257-274.

Delgado, N. 2024. The intonation of Spanish and Basque in Gipuzkoa and French and BASQUE in Labourd: an exploratory study. Ph.D. dissertation, Florida State University.

Dorta, J. 2000. Entonación hispana: Interrogativas no pronominales Delgado, Nerea. 2014. *The intonation of Spanish and Basque in Gipuzkoa and French and Basque in Labourd: An exploratory study*. PhD dissertation, FSU.

Face, T. 2003. Intonation in Spanish Declaratives: Differences Between Lab Speech and Spontaneous Speech. *Catalan Journal of Linguistics 2*. 115-131.

Gaminde, I., Romero Andonegi, A., Garay Ruiz, U., & Etxebarria, A. 2014. Los tonos de frontera de las oraciones interrogativas pronominales producidas por hablantes bilingües precoces en vasco y español. *Zeitschrift für romanische Philologie* 130. 1-16.

Henriksen, N. 2009. Wh-question intonation in Peninsular Spanish: Multiple contours and the effect of task type. *Journal of Portuguese Linguistics* 8. 47-74.

Hualde, J. I., & P. Prieto. 2015. Intonation variation in Spanish: European and American varieties. In R. Frota and P. Prieto (eds.), *Intonation in Romance.* 350-391. Oxford: OUP.

Navarro Tomás, T. 1944. *Manual de entonación española.* México D.F.: Colección Málaga.

Quilis, A. 1993. Tratado de fonología y fonética españolas. Madrid: Gredos.

RAE. 2011. Nueva Gramática de la Lengua Española. Fonética y fonología. Madrid: Espasa.

Savino M. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42. 23-48.

Sosa, J. M. 1999. La entonación del español: Su estructura fónica, variabilidad y dialectología. Madrid: Cátedra.

11. Perception of Japanese consonant length contrasts by advanced learners from Mandarin-, Mongolian- and Vietnamese-speaking backgrounds

Kimiko Tsukada[1], Yurong[2] and Trang Le[3]
[1]The University of Melbourne, [2]Inner Mongolia University, [3]Waseda University

Japanese is a so-called quantity language that uses durational variation contrastively for both vowels (e.g., *shi**i**te* 'doing' vs *shi**ii**te* 'spreading, laying out') and consonants (e.g., *shi**t**e* 'doing' vs *shi**tt**e* 'knowing'). It is widely acknowledged that length contrast is difficult for non-native speakers from diverse L1 (first language) backgrounds [1]. The main question to be addressed in this study is the extent to which the Japanese consonant length (i.e., short/singleton vs long/geminate) contrasts are processed accurately by highly advanced learners of Japanese from Mandarin, Mongolian and Vietnamese backgrounds. Unlike Japanese, consonant length is non-contrastive in the learners' L1s. However, Mongolian and Vietnamese, but not Mandarin, permit various consonants in coda position, which leads to a high incidence of fake/concatenated gemination (as in English *one nail*, *unnoticed* [2]). Further, contrastive use of vowel duration is systematic in Mongolian [3], limited to a few lexical pairs in Vietnamese [4] and non-existent in Mandarin [5].

Twelve Japanese word pairs (e.g. *kato-katto*, *kako-kakko*) were audio-recorded by six (3 males, 3 females) L1 Japanese speakers and presented to the participants via a two-alternative forced-choice AXB discrimination task. Only tokens with voiceless stops (/t k/) were considered in this study, because voiced geminates are disfavoured [6, 7] and bilabial geminates are limited to loan words and onomatopoeic words in Japanese. On average, the closure durations were 96 ms and 262 ms for singletons and geminates, respectively. The geminate-to-singleton ratios were 2.7 for alveolars (/t/-/tː/) and 2.8 for velars (/k/-/kː/), respectively. These values approximately match with the ratios reported in earlier research [8].

Three groups of advanced learners and a control group of native Japanese speakers participated in the study. The learner groups consisted of 10 Mandarin speakers ($M_{\text{age}} = 24.9$), 10 Mongolian speakers ($M_{\text{age}} = 23.3$) and 13 Vietnamese speakers ($M_{\text{age}} = 28.6$), all of whom had passed the highest level (N1) of the Japanese Language Proficiency Test. However, this does not guarantee that the proficiency level was perfectly matched across groups. The Mandarin and Mongolian groups participated in the study in Hohhot, Inner Mongolia Autonomous Region, China and the Vietnamese group participated in the study in Tokyo, Japan. A control group of 10 Japanese speakers ($M_{\text{age}} = 21.0$) participated in the study in Eugene, Oregon, USA. All native Japanese speakers were born and spent most of their life in Japan. Their mean length of stay in the US was 0.4 years (*sd* = 0.22) at the time of participation. None of these Japanese speakers participated in the audio-recording sessions. According to self-report, all four groups of participants had normal hearing. The speech materials (/(C)V**C(C)**V/ tokens) were arranged in 200 unique triads (e.g. $ha\underline{\mathbf{k}}a_2$ – $ha\underline{\mathbf{k}}a_1$ – $ha\underline{\mathbf{kk}}a_3$, where the subscripts indicate different speakers). The presentation of the stimuli and the collection of perception data were controlled by the PRAAT program [9]. Excluding the first 8 trials for practice, the triads contained 96 singleton or 96 geminate tokens intervocalically (underlined and bolded).

The overall mean discrimination accuracy was 86%, 96%, 91% and 99% for the Mandarin, Mongolian, Vietnamese and Japanese groups, respectively (Figure 1). Levene's test for homogeneity of variance was significant [$F(3, 39) = 4.6$, $p < .01$] when all four groups were included, but non-significant [$F(2, 30) = 2.9$, $p = .07$] when the control Japanese group was excluded, suggesting that homogeneity of variance can only be assumed for the three learner groups. A non-parametric Kruskal-Wallis test with Group reached significance [$\chi^2(3) = 20.3$, $p < .001$, $\eta^2 = .44$]. Bonferroni-adjusted post-hoc tests showed that the control Japanese group outperformed the Mandarin and Vietnamese groups [$p < .05$], but not the Mongolian group.

The Mongolian group was significantly more accurate than the Mandarin group [$p < .05$], but not the Vietnamese group. These results are consistent with previous research which reported L1 influence in Mandarin tone learning for advanced learners from Vietnamese vs American English backgrounds [10] and suggest that L1 influence may continue to affect the cross-linguistic processing of Japanese consonant length even for highly advanced learners. Pedagogical interventions targeting on specific areas of L1-L2 misalignment may be additionally beneficial for learners familiar with vowel and/or consonant length contrasts [10].

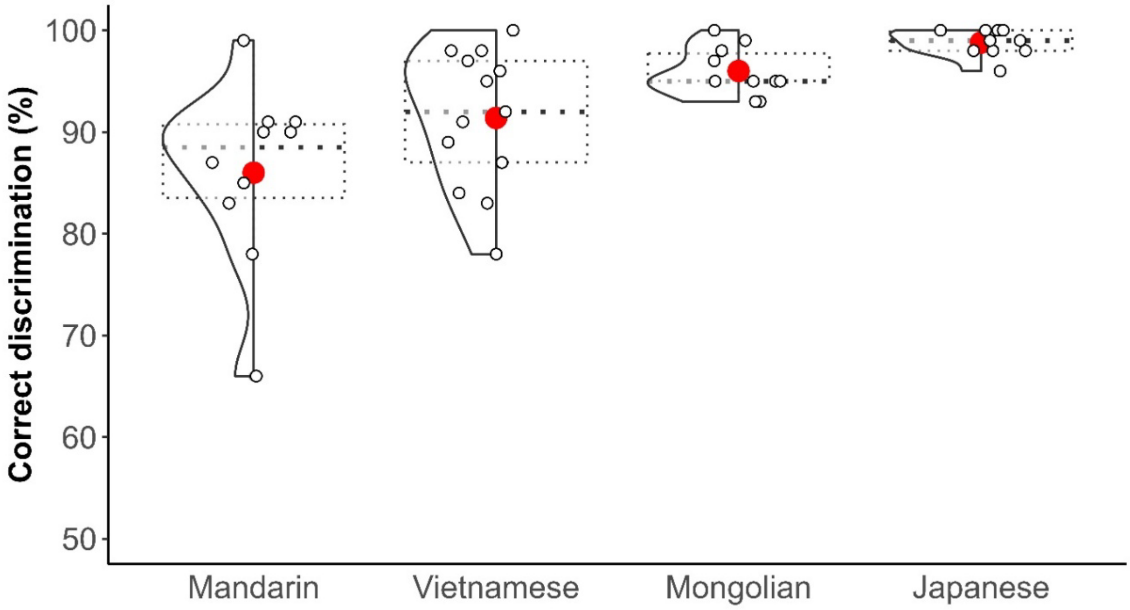


Figure 1. *Overall discrimination accuracy (%) for the four groups of participants. The horizontal line and the red circle in each box indicate the median and mean, respectively. The bottom and top of the box indicate the first and third quartiles.*

[1] Hirata, Y. 2015. L2 phonetics and phonology. In Kubozono, H. (Ed.), *Handbook of Japanese Phonetics and Phonology*. De Gruyter Mouton, 719–762.

[2] Oh, G. E., & Redford, M. A. 2012. The production and phonetic representation of fake geminates in English. *Journal of Phonetics* 40, 82–91.

[3] Iivonen, A., & Harnud, H. 2005. Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt. *Journal of the International Phonetic Association* 35, 59–71.

[4] Kirby, J. P. 2011. Illustrations of the IPA: Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association* 41, 381–392.

[5] Feng, Q., & Busà, M. G. 2022. Mandarin Chinese-speaking learners' acquisition of Italian consonant length contrast. *System* 111, 192938.

[6] Hussein, Q., & Shinohara, S. 2019. Partial devoicing of voiced geminate stops in Tokyo Japanese. *Journal of the Acoustical Society of America* 145, 149–163.

[7] Kawahara, S. 2015. The phonetics of sokuon, or geminate obstruents. In Kubozono, H (Ed.), *Handbook of Japanese Phonetics and Phonology.* De Gruyter Mouton, 43–78

[8] Han, M. S. 1992. The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica* 49, 102–127.

[9] Boersma, P., & Weenink, D. 2016. *Praat: doing phonetics by computer*. Computer program.

[10] Pelzl, E., Liu, J., & Qi, C. 2022. Native language experience with tones influences both phonetic and lexical processes when acquiring a second tonal language. *Journal of Phonetics* 95, 101197.

12. The prosodic status of color ideophones in Kriol: an exploratory study

Gabriela Braga[1], João Imbatene[2], Flaviane Fernandes-Svartman[2], Márcia Oliveira[2]
[1]University of Lisbon, [2]University of São Paulo

Ideophones are linguistic elements that constitute an open lexical class of marked words that depict sensory imagery [1]. Each ideophone only matches one host word and depends on this host word to compose the meaning. The ideophone itself can never appear alone in the sentence, and its host word can be a verb, a noun, an adjective, or an adverb. Literature about ideophones usually defines them as a linguistic mechanism with a similar function to the intensifier adverbs, although an adverb can appear alone in the sentence and present a meaning by itself.

In Kriol, a language spoken in Guinea-Bissau, there are three color ideophones, 'fandan', 'wak', and 'nok', that can only follow, respectively, the colors white ('branku'), red ('burmedju'), and black ('pretu') [2]. Our goal is to investigate if the color ideophones in Kriol constitute or not independent prosodic words.

To reach our goal, we elaborated a corpus of 243 declarative utterances with different phrasing to test: 1) which prosodic cues can be found associated with the ideophones in several prosodic boundaries; and 2) if there are any differences between ideophones and the intensifier adverb 'mal' (a lot, very) from a prosodic perspective. The data was collected through a Reading Task with four Guinean speakers (two men and two women), aged between 24-26 years old, native speakers of Kriol among other autochthonous languages (Papel, Manjak, Mankanya) and Guinean Portuguese. The audio material was segmented and analyzed with Praat [3]. All data analysis used the theoretical framework of Prosodic Phonology [4], and Intonational Phonology [5], and the P-ToBI annotation [6].

In Kriol, all prosodic words (PW) head of phonological phrases (PhP) present pitch accents associated. Besides, the presence of phrasal accents (L^{-}) associated with the right edge of internal PhPs, and lengthening of the stressed vowel of the PW head of PhP are cues of the boundary of this prosodic constituent [7].

Our results so far show pitch accents associated with both the prosodic word that conveys the meaning of the color ($PW_{[color]}$) and the ideophone that follows that $PW_{[color]}$, regardless of whether the right edge of the ideophone matches the boundary of a PhP (Fig. 1) or the IP boundary. Furthermore, the pitch accent associated with the ideophone usually presents the L*+¡H configuration, being the peak of the intonational contour.

Thus, the pitch accent associated with the ideophone together with the stressed vowel lengthening of the $PW_{[color]}$ that precedes it are cues that ideophones may constitute independent PhPs from the PWs they established a complementary relation with. The possibility that the ideophone may be being produced with prosodic focus is a hypothesis yet to be investigated, but that also reinforces that the ideophone constitutes a PhP by itself. Furthermore, ideophones show similar intonational behavior such as the one found for the intensifier adverb that appears in the same position within the utterance and in the same complementary relation with the $PW_{[color]}$ they follow, which seems to confirm the hypothesis that color ideophones, although unable to form an utterance alone, behave as a prosodic word.

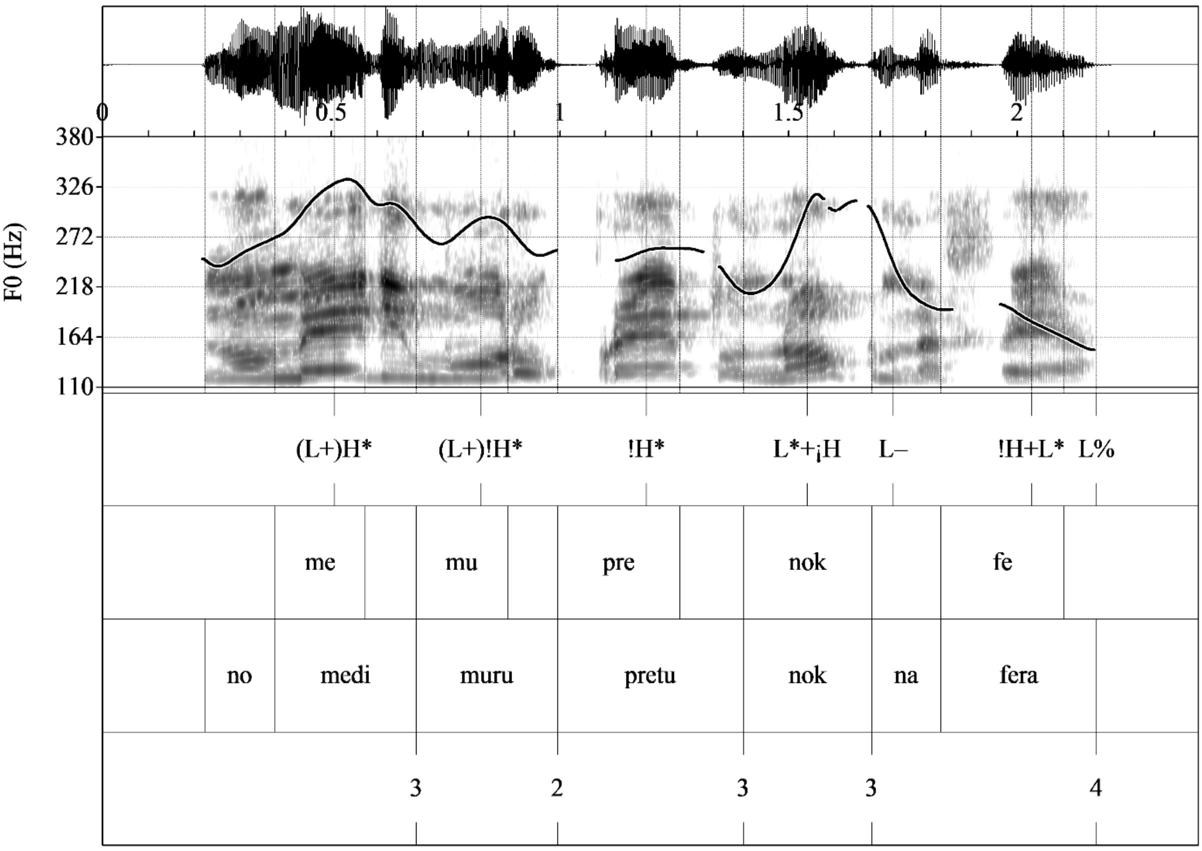


Figure 1. *Sentence "No medi muru pretu nok na fera" (We are afraid of the darker "black" sorcerer from the market), showing the prosodic cues associated with the color ideophone.*

[1] M. Dingemanse, "'Ideophone' as a comparative concept," in *Ideophones, Mimetics and Expressives*, K. Akita and P. Pardeshi, Eds., in Iconicity in Language and Literature. John Benjamins Publishing Company, 2019, pp. 13–34. doi: 10.1075/ill.16.02din.

[2] J. E. Imbatene, "O sistema ideofônico da língua guineense," Master Thesis, Universidade de São Paulo, 2024. doi: 10.11606/D.8.2024.tde-01102024-122826.

[3] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. (2014). [Windows]. Available: http://www.praat.org

[4] M. Nespor and I. Vogel, *Prosodic Phonology: With a New Foreword*. Berlin/ New York: Mouton de Gruyter, 2007. doi: 10.1515/9783110977790.

[5] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press, 2008. doi: 10.1017/CBO9780511808814.

[6] S. Frota, P. Oliveira, M. Cruz, and M. Vigário, *P-ToBI: tools for the transcription of Portuguese prosody*. Lisboa: Laboratório de Fonética, CLUL/FLUL, 2015. Accessed: Dec. 28, 2023. [Online]. Available: http://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/

[7] G. Braga, "A prosódia do guineense (kriol) e do português falado na Guiné-Bissau: análise de tipos frásicos," Thesis (Doctorate of Philosophy), Universidade de São Paulo/ Universidade de Lisboa, São Paulo/ Lisboa, 2023. doi: 10.11606/T.8.2023.tde-22112023-155220.

13. Segmental pre-nuclear peak alignment in two Mexican Spanish ethnolects

Gilly Marchini, Jeremy Steffman
University of Edinburgh

This paper employs a bottom-up, data-driven approach via clustering to examine variation in pre-nuclear peak alignment across two Mexican Spanish ethnolects: Afro-Mexican Spanish (AFS), and Altiplateau Mexican Spanish (AMS). Whilst pre-nuclear peak alignment varies according to focus in AMS (tonic peaks in broad focus, post-tonic peaks in narrow) [1], pilot analysis of AFS highlights variability according to the segmental string: broad focus peaks align at the end of vowel + /n/ sequences regardless of whether segments are hetero- or tautosyllabic [2], [3]. We expand this analysis, examining peak offset according to focus, the segmental string and dialect. We consider whether patterns are unique to AFS and the nasal, or can be expanded to other similar sounds, e.g., /l/.

11 participants from each community read aloud 12-syllable sentences containing target words in broad and narrow focus. The target syllable contained the tonic vowel /a/ preceding /l/, /n/, or /s/ (C1). Syllable aperture was manipulated through the inclusion of /t/, e.g., manto (/ˈman.to/, 'cloak') vs mano (/ˈma.no/, 'hand'). Target words received the 3rd and penultimate pitch accent. F0 was sampled at 20 time-normalised intervals within each syllable, totalling 40 across the tonic and post-tonic syllable. F0 trajectories were z-scored within speaker and processed by K-Means clustering for time-series data [4], [5]. Clusters with a rising shape were subset from the data and passed through the clustering model again to examine the factors conditioning pitch shape and peak alignment in rises (AFS, $N = 737$; AMS, $N = 943$).

For both dialects, K-Means identified two rises: Cluster A with a tonic peak, and Cluster B with a post-tonic peak (Figure 1). Descriptively, segmental effects emerge for both dialects. For AFS, Cluster A represents 58.53% of peaks in broad focus /Vl.t/ and Cluster B 58.82% of peaks in /V.lV/. For AMS, Cluster B accounts for the majority of rises in /Vl.t/ and /Vn.t/ (66.66% & 56.25% respectively) in broad focus. For both dialects, clusters were evenly distributed across /Vs.t/ and /V.sV/ environments (Figure 2). Interestingly, these effects failed to reach statistical significance in either dialect. Instead, significant syllabic effects emerged for AFS, with Cluster A more likely in closed syllables than open across C1 ($t = -2.230$, $p < .05$). Cluster A was also more likely in narrow focus than broad for both dialects ($p < .001$).

Experimental data reveals that whilst segmental effects emerge, these are not statistically significant in either dialect. Instead, dialects differ in the role of syllable, with tonic peaks more likely in closed syllables than open for AFS, but not AMS. Both dialects exhibit significant focus effects, with tonic peaks more likely in narrow than broad focus. In this way, AFS shows a double effect: peak alignment is affected by both syllable aperture and focus. We postulate that this is attributable to differences in syllable duration between dialects. It may be that closed syllables are longer than open in AFS, such that they provide sufficient phonetic material for the peak to continue rising [6], [7]. These differences may not be as pronounced in AMS, hence invariant peak alignment patterns occur. We propose further analysis of the durational properties of vowel and C1 segmental units to test this, with further GAMM and FPCA analysis (as in [8], for FPCA) to ascertain where precisely on the segmental string peaks align, and whether significant differences emerge between the clusters. We further highlight the need for speaker-specific clustering, since statistical models point to speaker-specific partitioning of trajectories between the clusters which may be overriding segmental trends. Regardless, preliminary data provides promising evidence of previously undocumented interactions between syllable organisation and peak alignment in an overlooked variety of Spanish.

.

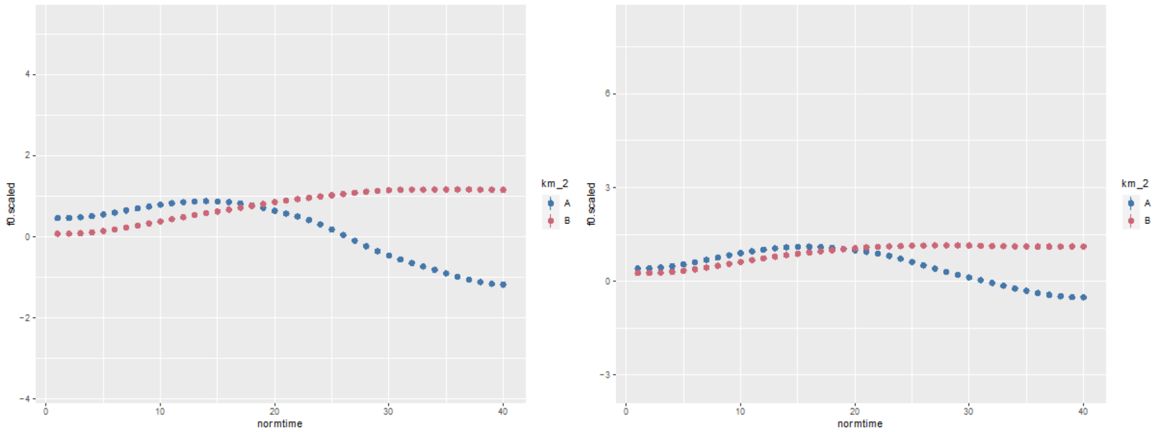


Figure 1. *Clustering output for AFS (left) and AMS (right).*

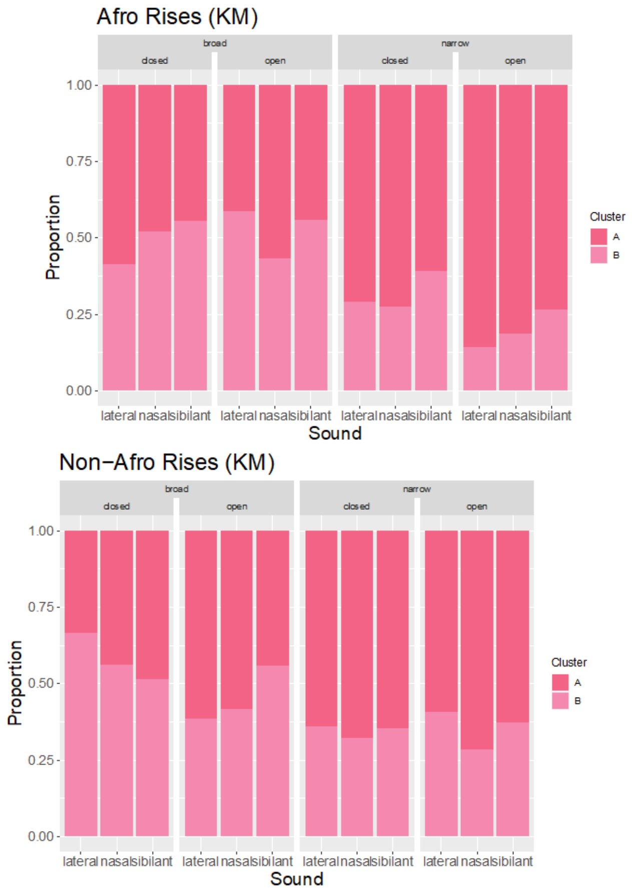


Figure 2. *Distribution of Cluster A and Cluster B across segmental, syllabic and focus conditions for AFS (left) and AMS (right).*

[1] de la Mota, C., Butragueño, P. M., Prieto, P., and Fabra, P. 2011. Mexican Spanish Intonation Mexican Spanish Intonation. In Prieto, P. and Roseano, P. (Eds), Transcription of Intonation of the Spanish Language, Munich: Europa, 319–359.
[2] Marchini, G. 2024a. An exploration of pitch in Afro-Mexican Spanish. ISSP Conference Proceedings (Grenoble)
[3] Marchini, G. 2024b. Peak Alignment in Afro-Mexican Spanish: an exploratory analysis. Speech Prosody 2024

[4] Genolini, C., & Falissard, B. 2010. KmL: k-means for longitudinal data. Computational Statistics, 25(2), 317-328.
[5] Steffman, J., Cole, J., & Shattuck-Hufnagel, S. 2024. Intonational categories and continua in American English rising nuclear tunes. Journal of Phonetics, 104, 101310.
[6] Barnes, J., Veilleux, N., Burgos, A., and Shattuck-Hufnagel, S. 2010. The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation, Speech Prosody 2010
[7] Barnes, J., Veilleux, N., Burgos, A., and Shattuck-Hufnagel, S. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. Laboratory Phonology 3(2), 337-383.
[8] Arvaniti, A., Katsika, A., & Hu, N. (2024). Variability, overlap, and cue trading in intonation. Language, 100(2), 265-307.

14. L1-Specific cue-weighting and adaptation in L2 prosodic processing: Evidence from Dutch and Chinese learners' processing of English prominence

Yiran Ding, Yiya Chen
Leiden University Centre for Linguistics, and Leiden Institute for Brain and Cognition, Leiden University

To comprehend speech, listeners must establish stable speech representations while also flexibly adapting to variability in the speech signal (Holt & Lotto, 2006; Baese-Berk et al., 2022). However, it remains unclear how listeners' first language (L1) influences these processes in the context of second language (L2) prosodic processing. This study addresses this question by comparing how Dutch and Chinese learners of English perceive and adaptively process prosodic prominence in English, investigating both their L2 speech representations and perceptual adaptation mechanism.

We employed a cue-weighting paradigm with two forced-choice identification tasks (Jasmin et al., 2023). Twenty-four Chinese learners and twenty-nine Dutch learners of English took part in two tasks. The first task assessed the listeners' baseline perceptual weights of fundamental frequency (F0) and duration for English prosodic prominence. Participants were presented with stimuli manipulated and synthesized using STRAIGHT (taken from Jasmin et al., 2021), each comprising seven steps of both F0 and duration to create a two-dimensional speech continuum. The second task investigated learners' perceptual adaptation strategies. Participants underwent the sequential exposure of two distinct listening blocks: a canonical block reflecting naturalistic prosodic patterns of English prominence, a reversed block where inherent F0-duration correlation of English prominence was flipped to create an artificial "accent".

Bayesian hierarchical modelling was implemented using Bambi (Capretto et al., 2020) to analyse the response data. The model incorporated both group-level and individual-level parameters with weakly informative priors. Our results firstly revealed distinct language-specific patterns in L2 prosodic processing. Chinese learners predominantly relied on pitch cues when judging English prosodic prominence (mean = 0.531, HDI [0.355, 0.703]), likely due to the lexical tonal nature of their native language. In contrast, Dutch learners showed greater sensitivity to duration cues than the Chinese group (mean = 0.112, HDI [0.002, 0.219]), suggesting a preference for temporal cues in prominence judgment.

Furthermore, our findings indicate an influence of L1 on adaptive cue-weighting mechanisms. Chinese learners partially replicated the cue re-weighting patterns of native English listeners by down-weighting the duration in response to accented speech, suggesting that their pitch-related processing experience in L1 may support the development of English native-like adaptation mechanisms in L2 prosody. Conversely, Dutch learners displayed a different adaptation pattern by down-weighting the pitch cue when encountering acoustic variability. This suggests they were developing a cue re-weighting strategy to cope with input variability, likely aligned with their L1 preference for duration cues.

Additionally, individual variation was observed within both language groups, with some learners exhibiting more native-like adaptation patterns than others, highlighting individual differences in L2 prosodic learning.

Conjointly, our findings demonstrate possible influence of L1 on L2 prosodic prominence processing. The consistent patterns observed - Chinese learners' reliance on pitch cue and Dutch learners' preference for duration cue - highlight how listeners with different L1s may develop different initial cue weights in their L2 speech representations, which further affects the adaptive re-weighting of cues during online L2 speech processing.

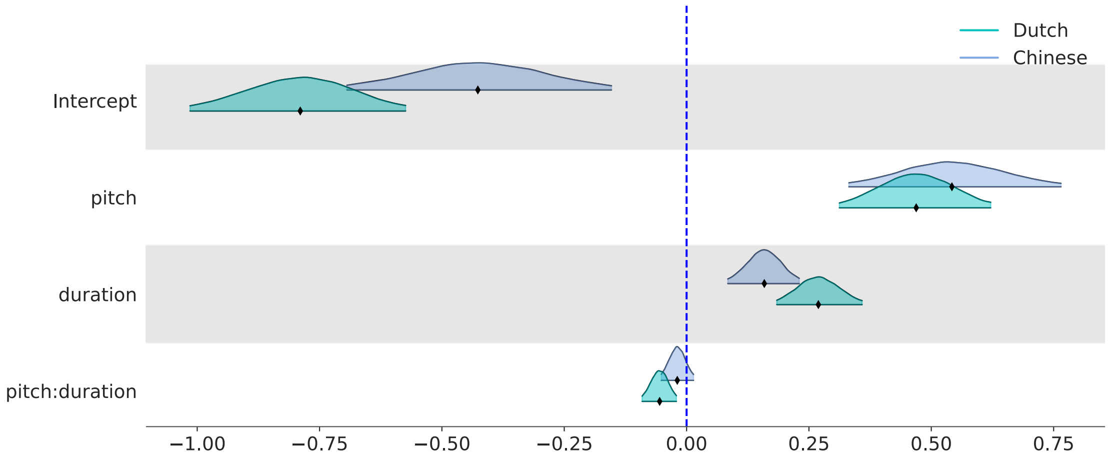


Figure 1. *Main result of task 1: Forest plot of posterior probability of each parameter in Chinese and Dutch groups.*

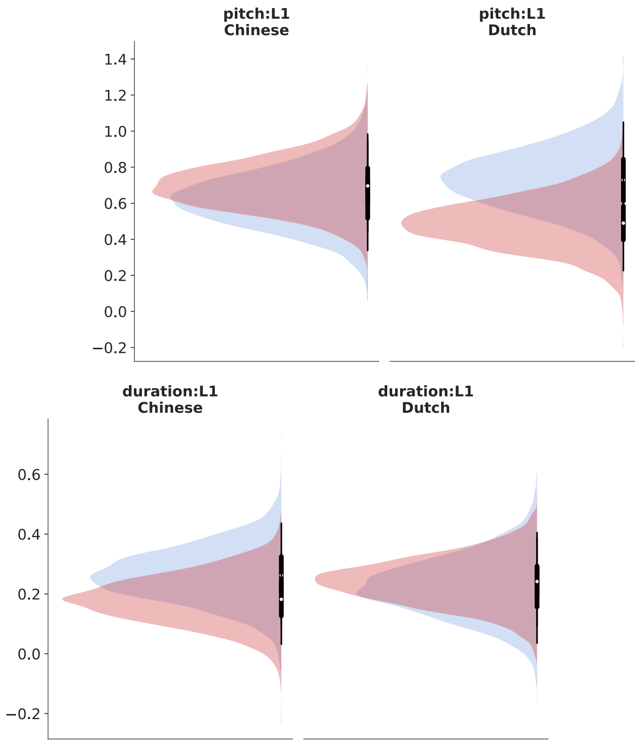


Figure 2. *Main result task 2: Violin plots show the (shift) difference of posterior probability of pitch and duration across different exposure blocks (Canonical Block in blue, Reversed Block in red) and L1s (Chinese and Dutch groups).*

[1] Baese-Berk, M. M., Chandrasekaran, B., & Roark, C. L. (2022). The nature of non-native speech sound representations. The Journal of the Acoustical Society of America, 152(5), 3025–3034.

[2] Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. The Journal of the Acoustical Society of America, 119(5), 3059–3071.

[3] Jasmin, K., Tierney, A., Obasih, C., & Holt, L. (2023). Short-term perceptual reweighting in suprasegmental categorization. Psychonomic Bulletin & Review, 30(1), 373-382.

[4] Jasmin, K., Dick, F., & Tierney, A. T. (2021). The multidimensional battery of prosody perception (MBOPP). Wellcome Open Research, 5, 4.
[5] Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2020). Bambi: A simple interface for fitting Bayesian linear models in Python. *arXiv preprint arXiv:2012.10754*.

15. Perception of English fricative contrasts by L1 Spanish and L1 Portuguese learners

Ellen Deemer Pérez, Miquel Simonet
University of Arizona

This study examines the perception of English fricatives /s z ʃ/ among L1 Portuguese and L1 Spanish speakers, focusing on the relative perceptual sensitivity of ESL learners to the voicing (/s/-/z/) and place of articulation (/s/-/ʃ/) contrasts. While Portuguese shares with English its fricative inventory, Spanish lacks both /z/ and /ʃ/. This study explores the extent to which overlap between the phonemic inventories of the L1 and the L2 enhances (or hinders) perceptual development and word learning and whether learners are more sensitive to place than to voicing distinctions, or *vice versa*, when both are novel.

40 L1 Spanish speakers from Mexico, 39 L1 Portuguese speakers from Portugal, and 36 L1 English speakers from the United States (as "controls") were recruited online. Participants completed an ABX perceptual discrimination task and a two-alternative forced choice perceptual identification task with real words. Stimuli consisted of bi-syllabic minimal pairs with fricatives in word-initial (*I say sink/zinc again*) or word-final positions (*I say ice/eyes again*). Stimuli were recorded and played in the context of a constant carrier phrase: *I say __ again*. This ensured that all target fricative sounds, whether word-initial of word-final, were intervocalic. Contrast conditions included /s/-/ʃ/, /s/-/z/, and /s/-/f/ (as control). Participants also completed three English proficiency assessments: two vocabulary-size tasks and a cloze task ([1], [2]). This allowed us to correlate their English proficiency with their phonetic behaviour as captured by our offline perceptual experiments.

Results revealed that L1 Portuguese listeners accurately discriminated fricative contrasts, showing high sensitivity to both voicing and place distinctions. Performance in the word identification task was also highly accurate. Interestingly, these findings suggested that both perceptual discrimination and word learning were robust in this learner sample, at least when measured with these offline tasks. L1 Spanish listeners, on the other hand, struggled with the /s/-/z/ contrast, with accuracy only slightly above chance, but were more accurate with the /s/-/ʃ/ contrast. This was true for both the discrimination and identification tasks, and it suggests that obstructed perceptual development may hinder word learning.

The performance difference between the learner groups suggests that the phonological inventory of the L1 shapes how listeners perceive fricatives in an L2, including not only discrimination but also word learning. At this juncture, this is hardly surprising, as these findings come to corroborate many similar ones with various populations and phonemic contrasts. Most relevantly, however, the findings also suggest that learners who lack both /z/ and /ʃ/ may have a harder time with one than with the other. For this population, acquisition of the voicing contrast seems harder than that of the place contrast. In our discussion, we consider several potential explanations for our finding, and we obviously conclude that not all novel phonemes or phonemic contrasts are equally challenging.

Follow-up research being conducted as of the writing of this abstract focuses on the perceptual exploitation of these perceptual patterns during online lexical processing. An auditory lexical decision task with sound substitutions is being run. Nonwords were created by swapping one phoneme with another critical one: *pre*[s]*ident* rather than *pre*[z]*ident*, *ba*[z]*ic* rather than *ba*[s]*ic*, *fa*[s]*ion* rather than *fa*[ʃ]*ion*, and *ma*[ʃ]*ive* rather than *ma*[s]*ive*. Inaccurate word-endorsement (accepting a nonword as word) rates and response times are to be obtained and measured. The results of this task will corroborate the extent to which perceptual and lexical development are interdependent. We postulate that even L1 Portuguese learners will experience major difficulties with this task (e.g., [3]). While we intend to report on all three experiments, we can confirm that, as of the writing of this abstract, we have perceptual data for two offline experiments (ABX, 2AFC) and a total of 115 listeners.

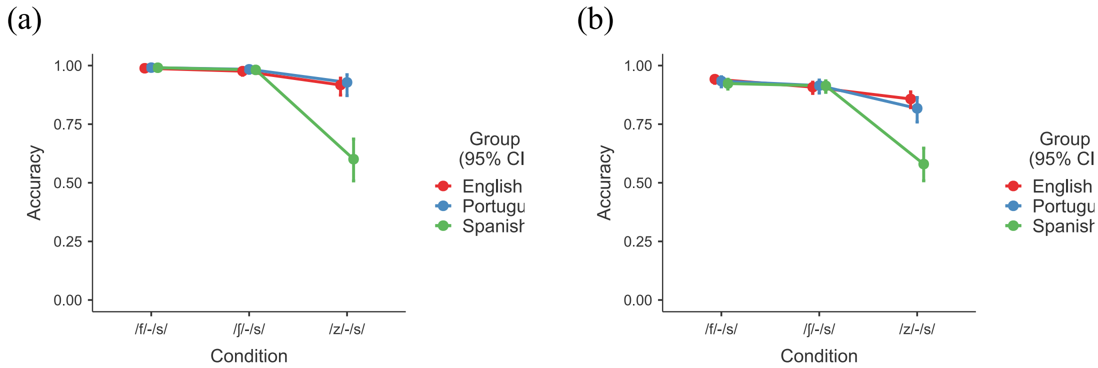


Figure 1. *(a) Estimated marginal means (means and 95% CI) from multilevel logistic regression model with Accuracy (proportion correct) as response and Condition and L1 Group as predictors, data from* <u>*2AFC task*</u>*; (b) Estimated marginal means (means and 95% CI) from multilevel logistic regression model with Accuracy (proportion correct) as response and Condition and L1 Group as predictors, data from* <u>*ABX task*</u>.

[1] K. Lemhöfer and M. Broersma, "Introducing LexTALE: A quick and valid lexical test for advanced learners of English," *Behav. Res. Methods*, vol. 44, pp. 325–343, 2012.

[2] A. Tremblay, "Proficiency assessment standards in second language acquisition research: 'Clozing' the gap," *Studies in Second Language Acquisition*, vol. 33, pp. 339–372, 2011.

[3] M. Amengual, "The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals," *Appl. Psycholinguist.*, vol. 37, pp. 1221–1251, 2016.

16. Investigating the neural processing of regional and foreign accents

Darya Klymenko[1], Nadege Bault[1], Julien Besle[1], Jeremy Goslin[1], Nina Kazanina[2], Caroline Floccia[1]
[1]University of Plymouth, UK, [2]University of Geneva, Switzerland

The aim of this pre-registered fMRI study is to investigate whether regional (native) and foreign (non-native) accents in English are processed differently by the brain of native English speakers. Regional and foreign accents are different in nature, with foreign accents having greater within-speaker variability in the way vowels are realised (Laturnus, 2020), as well as often introducing phonemes unfamiliar to the listener (Wells, 1982). To date, ERP evidence points to a qualitative difference in the neural mechanisms for the processing of regional and foreign accents (Goslin et al., 2012; Thomas et al., 2020). A proposed explanation for this is that while the unfamiliar features of a regional accent are "filtered out" before lexical selection, foreign-accent processing relies on top-down mechanisms from the lexical level (Goslin et al., 2012).

fMRI studies on perception of regional (eg. Adank et al., 2011; 2012; Hernandez et al., 2019) and foreign (Callan et al., Yi et al., 2014) accents found partial overlap in active brain regions. However, due to methodological differences, a between-study comparison of activation for regional vs. foreign accents does not yield a conclusive answer. Therefore, distinct neural processing is plausible but should be investigated with a direct comparison. We collected fMRI data from 30 English-speaking monolinguals from the South-West of England with no reported everyday exposure to other accents. These participation criteria were chosen to control for participants' familiarity with the accents used in the experiment. In an event-related design, participants heard grammatically correct sentences read either by speakers for whom English was not a first language (*foreign* condition), native speakers from outside of the South-West of England (*regional* condition), and native speakers from the South-West of England (*home* condition used as a control). Each condition was represented by sentences spoken by ten speakers, each with a different accent in the FOREIGN and REGIONAL conditions. Participants engaged in a go/no-go semantic task based on the content of the sentences to ensure attention. We also collected online ratings of the stimuli from 30 additional participants from the South-West of England, measuring subjective intelligibility, objective intelligibility, strength of accent and prestige associated with the accent. Differences between the foreign-accent and regional-accent stimuli were found for all these dimensions. Controlling for these differences, we predict that the REGIONAL – HOME and the FOREIGN – HOME contrasts will show activity in partially overlapping, but distinct brain networks. Based on previous literature, we expect to see increased activity in the temporal lobe regions, such as the STG or the STS, for the REGIONAL – HOME contrast, since those areas are associated with phonological processing (Adank et al., 2011; Hickock & Poeppel, 2007). On the other hand, we expect the FOREIGN- HOME contrast to show increased activity in the areas associated with more complex top-down processing (Hickock & Poeppel, 2007), such as the frontal lobe (Yi et al., 2014) and potentially motor regions (Yi et al., 2014; Callan et al., 2014**).**

The pre-registration of the study can be found on OSF: https://osf.io/x9ejp

[1] Adank, P., Davis, M. H., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, *50*(1), 77–84. https://doi.org/10.1016/j.neuropsychologia.2011.10.024

[2] Adank, P., Noordzij, M. L., & Hagoort, P. (2011). The role of planum temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping*, *33*(2), 360–372. https://doi.org/10.1002/hbm.21218

[3] Callan, D., Callan, A., & Jones, J. A. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00275

[4] Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102. https://doi.org/10.1016/j.bandl.2012.04.017

[5] Hernández, M., Ventura-Campos, N., Costa, A., Miró-Padilla, A., & Ávila, C. (2019). Brain networks involved in accented speech processing. *Brain and Language*, *194*, 12–22. https://doi.org/10.1016/j.bandl.2019.03.003

[6] Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

[7] Laturnus, R. (2020). Comparative Acoustic Analyses of L2 English: The Search for Systematic Variation. *Phonetica*, *77*(6), 441–479. https://doi.org/10.1159/000508387 [4] Antolik, T.K., & Fougeron, C. 2013. Consonant distortions in dysarthria due to Parkinson's disease, Amyotrophic Lateral Sclerosis and Cerebellar Ataxia. *Proceedings of Interspeech 2013* (Lyon, France), 2152-2156.

[8] Thomas, T., Martin, C. D., & Caffarra, S. (2022). An ERP investigation of accented isolated single word processing. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2022.07.07.499100

[9] Wells, J. C. (1982). *Accents of English: An Introduction.* https://www.amazon.com/Accents-English-Introduction-JWells/dp/0521297192

[10] Yi, H. G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00768

17. Incidental learning of non-native tone sandhi by tonal and non-tonal L1 speakers

Xinbing Luo, Brechtje Post, John Williams
Department of Theoretical and Applied Linguistics, University of Cambridge

Tone sandhi, the phonologically and morphosyntactically context-dependent alteration of tones [1], is a prominent feature in tonal languages like Mandarin. Acquiring these tonal alternations can be particularly challenging for second language (L2) learners, as it requires both accurate tone perception and the ability to apply phonological rules that are not always explicitly taught. While previous studies have shown that native speakers of tonal languages and non-tonal languages may differ in tone discrimination, categorisation, and production (e.g., [2], [3]), it is unclear whether this distinction extends to the learning of tone sandhi patterns (for production study see [4]). Moreover, the role of phonetic motivation—whether certain sandhi patterns are more learnable due to their natural phonetic grounding—remains an open question (see [5] for a review). To address these gaps, the current study investigates whether incidental exposure alone is sufficient for both tonal and non-tonal speakers to acquire Mandarin tone sandhi patterns and whether prior tonal experience and phonetic motivation influence the success of this learning.

Forty-four native Cantonese speakers and forty-four native English speakers were recruited and randomly assigned to the Cantonese/English Experiment Groups (CExp/EExp) and Cantonese/English Control Groups (CC/EC). They first completed a citation tone AXB discrimination task (including 36 trials), which was designed to test if both groups could accurately distinguish relevant tonal contrasts. Following this, participants in the experiment groups proceeded to an incidental exposure phase (including 60 trials) in which they were presented with pseudowords observing Tianjin Mandarin tone sandhi rules (phonetically motivated: T3T2, T3T4; phonetically unmotivated: T3T3) without being informed of these rules as well as sandhi-irrelevant pseudowords (T2T4, T4T2), mimicking natural language exposure and reducing conscious focus on tone manipulation, while the control groups skipped the exposure phase. This exposure phase was designed to simulate real-world incidental learning conditions, in which L2 learners encounter novel linguistic patterns without explicit instruction. After exposure, participants completed a two-alternative forced-choice (2AFC) task, where they selected the likely correct pronunciation between two tonal options, assessing their ability to apply tone sandhi rules to both familiar and novel word pairs (60 trials in total). To gauge the level of conscious awareness of the rules, participants were asked to attribute their responses to one of four categories: guess, intuition, memory, or rule. This attribution approach provided a means to distinguish between implicit and explicit knowledge, as higher accuracy in responses attributed to intuition or guess would indicate implicit learning [6].

The results indicated that both tonal and non-tonal L1 speakers successfully distinguished the citation forms of the target tones, with no observed correlation between their performance in the AXB task and the subsequent 2AFC task. Both tonal and non-tonal L1 speakers were able to incidentally acquire tone sandhi knowledge, with native Cantonese speakers achieving significantly higher overall accuracy and higher accuracy on sandhi-irrelevant items (see Figure 1). While both implicit and explicit knowledge of sandhi rules emerged, explicit awareness did not substantially boost accuracy (see Figure 2). However, phonetic motivation—whether a pattern was phonetically grounded—had minimal influence on learnability, indicating that perceptual salience may not universally impact tonal learning as expected. These findings offer insights into the cognitive mechanisms of L2 tonal learning, underscoring the role of incidental exposure in the acquisition of complex phonological rules. This study suggests that incidental learning enables effective sandhi acquisition across different language backgrounds, with potential applications for immersive, meaning-based approaches to teaching tonal languages.

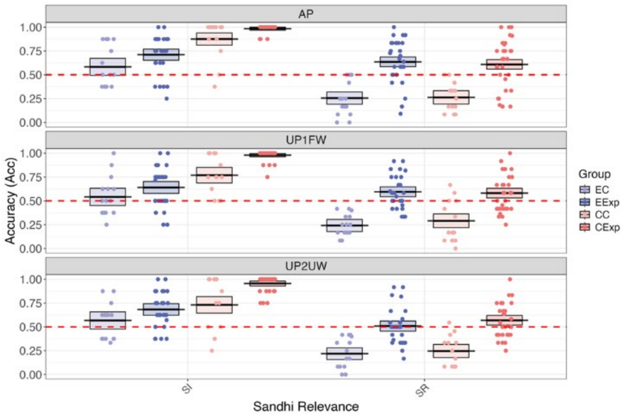


Figure 1. *Accuracy of the 2AFC Task by Group, Condition (Attested Phrase, Unattested Phrase with One Familiar Word, and Unattested Phrases with two Unfamiliar Words), and Sandhi Relevance. Bold bars indicate group means, rectangles show 95% confidence intervals, individual data points represent participant means, and red dotted lines show the chance level (0.5).*

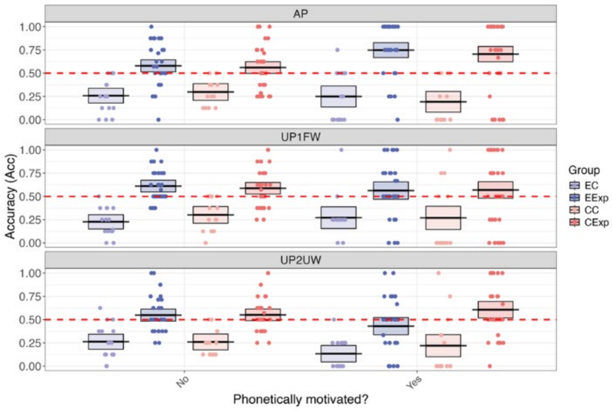


Figure 2. *Accuracy of the 2AFC Task (Sandhi Relevant Items) by Group, Condition, and Phonetic Motivation. Bold bars indicate group means, rectangles show 95% confidence intervals, individual data points represent participant means, and red dotted lines show the chance level (0.5).*

[1] Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

[2] Laméris, T. J., Llompart, M., & Post, B. (2024). Non-native tone categorization and word learning across a spectrum of L1 tonal statuses. *Bilingualism: Language and Cognition*, *27*(4), 729-743.

[3] Laméris, T. J., Li, K. K., & Post, B. (2023). Phonetic and Phono-Lexical Accuracy of Non-Native Tone Production by English-L1 and Mandarin-L1 Speakers. *Language and Speech*, *66*(4), 974-1006.

[4] Chen, S., He, Y., Wayland, R., Yang, Y., Li, B., & Yuen, C. W. (2019). Mechanisms of tone sandhi rule application by tonal and non-tonal non-native speakers. *Speech Communication*, *115*, 67-77.

[5] Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning, part II: Substance. *Language and linguistics compass*, *6*(11), 702-718.

[6] Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological research*, *69*, 338-351.

18. *L'occitan de l'avenidor*. Towards a description of the New Speaker variety of Occitan

Daniela Müller[1], Rafèu Sichel-Bazin[2], Arvèi Lieutard[3]
[1]LMU Munich, [2]Universitat Joan Jaurès Tolosa, [3]Universitat Pau Valèri Montpelhièr

In the near future, Occitan in France will be spoken primarily by individuals who were raised in French-speaking households and who have acquired Occitan through school or adult education. The project *L'occitan de l'avenidor* aims to analyse New Speakers of Occitan and the emerging Occitan variety they speak.

Occitan is a Gallo-Romance language historically spoken across much of what is now Southern France. Its rapid decline in the 20th century followed several centuries of increasing bilingualism across the region.

The project is currently in its pilot phase. For this stage, we recruited 15 participants who are native speakers of French and learned Occitan as a second language (L2) at the *Departament d'occitan* of the *Universitat Pau Valèri Montpelhièr*. Participants completed picture descriptions and story narration tasks in both Occitan and French. Recordings were transcribed and pre-segmented in Praat [1] and further processed in R [2] using the *emuR* package [3]. All recordings were segmented into sounds, syllables, and words, and annotated at the segmental, syllabic, lexical, and metrical levels. In addition, we collected accent judgement data on participants' French productions from a large sample of native French speakers, and we plan to collect comparable data for the Occitan recordings.

To understand the relationship between the participants' L1 French and their L2 Occitan, we analyse both language varieties:

French L1: Our analysis of the speakers' L1 French data aim at (1) determining the progress of accent levelling towards standard French in our speaker sample, and (2) identifying the weight of previously described characteristics in the perception of Southern French. Our analysis of participants' L1 French focuses on (1) assessing the extent of accent levelling towards Standard French, and (2) evaluating the perceptual salience of linguistic features previously described for Southern French [4]. The Southern French accent largely results from phonological transfer from Occitan, which was the majority language in the region's historically rural population. In our analysis, we examine both features associated with Occitan substrate influence—such as (a) regular realisation of *e muet* (schwa) and its potential backing, (b) partial orality of nasal vowels and their consonantal endings, (c) conditioned alternation of mid vowels, and (d) the presence of lexical stress—and indicators of levelling, such as (e) the increasing aspiration of plosives in various word positions.

Occitan L2 or New Speaker Occitan: For participants' L2 Occitan, we expect phonetic interference and/or phonological transfer from French, mirroring the historical transfer process that gave rise to the Southern French accent. To date, no phonetic studies of New Speaker Occitan exist. Our goals are therefore to (1) describe this emerging variety for the first time, including its internal variation, and (2) assess the direct influence of participants' L1 French on their Occitan productions. In particular, we analyse: (a) the centralisation of final unstressed [ɔ], etymologically and functionally related to word-final French *e muet*, (b) nasalised vowels before nasal consonants, which should be oral in Occitan, (c) the transfer of allophonic mid-vowel alternation despite a phonological contrast between mid-high and mid-low front vowels in stressed syllables, (d) the realisation of word-level stress and its potential shift to phrasal stress, as in French, and (e) the aspiration of plosives as in French.

Together with sociolinguistic interviews conducted with our participants, we also aim to gain insight into the sociophonetic factors conditioning the emergence of this particular New Speaker variety. In doing so, the project seeks to contribute to a deeper understanding of language acquisition processes in contexts of asymmetrical language contact.

[1] Boersma, P. & Weenink, D. (2025). *Praat: doing phonetics by computer* [Computer program]. Version 6.4.35, retrieved 15 June 2025 from http://www.praat.org/.

[2] R Core Team (2025). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[3] Jochim, M., Winkelmann, R., Jaensch, K., Cassidy, S., & Harrington, J. (2024). *emuR: Main Package of the EMU Speech Database Management System*. doi:10.32614/CRAN.package.emuR https://doi.org/10.32614/CRAN.package.emuR, R package version 2.5.0, https://CRAN.R-project.org/package=emuR.

[4] Coquillon, A. & Turcsan, G. (2012), « An overview of the phonological and phonetic properties of Southern French. Data from two Marseille surveys". In Gess, R., Lyche, Ch.& Meisenburg, T. (eds.), *Phonological Variation in French: Illustrations from three Continents*. Amsterdam, Philadelphia: John Benjamins; pp. 105-128.

Poster session 1b

1. A longitudinal study of perception and production of L2 Spanish stops and the role of individual differences

Xiaotong Xi[1], Christine Shea[2], Peng Li[3]
[1]*Shandong University of Finance and Economics,* [2]*University of Iowa,* [3]*Basque Center on Cognition, Brain and Language*

Flege's Speech Learning Model [1] suggests that increased second language (L2) experience can improve perception and production of non-native sounds, but findings on the role of study abroad (SA) in L2 sound learning are mixed, with some studies showing positive effects (e.g., [2]) and others reporting null effects (e.g., [3]). Furthermore, most research has examined the role of SA at the group level, yet individual differences may largely account for the effects of SA on L2 sound learning, especially concerning fine-grained phonetic details at the sub-phonemic level. One such case is the voice onset time (VOT) of stops. While Mandarin stops contrast in aspiration, Spanish stops contrast in voicing. Mandarin-speaking learners of Spanish tend to assimilate Spanish voiced stops to unaspirated stops, even among advanced learners with long-term residence in Spain [4]. This study explores whether SA improves learners' perception and production of Spanish stops (RQ1), and investigates how individual differences in auditory processing ability for duration (AP) and L2 use mediate the SA effects (RQ2).

Thirty Mandarin-speaking female learners of Spanish ("Chinese students") were recruited. They had received an average of 4.17 years of formal Spanish instruction in China. Upon their arrival in Spain for one-year master's programs instructed in Spanish, their proficiency ranged from intermediate to advance levels according to the DELE test. Perception and production of Spanish stops were tested at two time points: upon their arrival (pre-SA) and after one year when they were to finish the master program (post-SA). To provide baseline data, 19 female native speakers of Peninsular Spanish ("Spanish natives") completed the same perception and production tasks once. In the perception task, participants identified stops in three continua (/b/-/p/, /d/-/t/, and /g/-/k/), with the VOT ranging from −60 ms to 60 ms in 13 steps. Each stimulus was repeated four times. For the production task, participants read short dialogues embedding six Spanish stops in sentences in 2 position (sentence-initial vs. non-initial) × 2 speech prominence (prominent vs. non- prominent) × 2 stress (stressed vs. unstressed) conditions. As individual measures, at T1, Chinese students completed an auditory duration discrimination task [5]; and at T2, they reported their Spanish use in listening and speaking throughout the immersion period [6].

The key findings are as follows. On the group level, SA led to a shallower perceptual categorization slope for the /d-t/ contrast at post-SA than at pre-SA but did not show significant effects on the /b-p/ or /g-k/ contrasts (Fig. 1). Besides, we did not observe significant changes in the perceptual boundary. Regarding production, SA did not show a significant change in VOT values for any of the three contrasts, leading to significant differences from Spanish natives at both tests (Fig. 2). While Spanish natives could produce voiced and voiceless stops with distinct VOT values, Chinese students assimilated the two. On the individual level, for perception, we found that the AP of duration significantly mediated the SA effects. Specifically, Chinese students with better AP showed steeper perceptual slopes for all three stop contrasts after immersion (Fig. 3). In terms of production, we found that Chinese students with greater L2 use related to speaking produced reduced VOT values for voiced stops after immersion.

These findings suggest that, at group level, experienced L2 learners show non-native perception and production patterns, and SA experience does not facilitate their L2 sound acquisition. However, the effects of SA are mediated by aptitude and experiential factors at the individual level. Specifically, domain-general auditory processing abilities in subtle acoustic signals (e.g., duration) mediates the SA effects on establishment of new phonemic contrasts. Moreover, the quantity of L2 input during SA can mediate the SA effects on the fine-grained phonetic details in speech production. The current results call for the need to revisit previous findings on the role of SA in the development of speech sounds by focusing on individual-level factors.

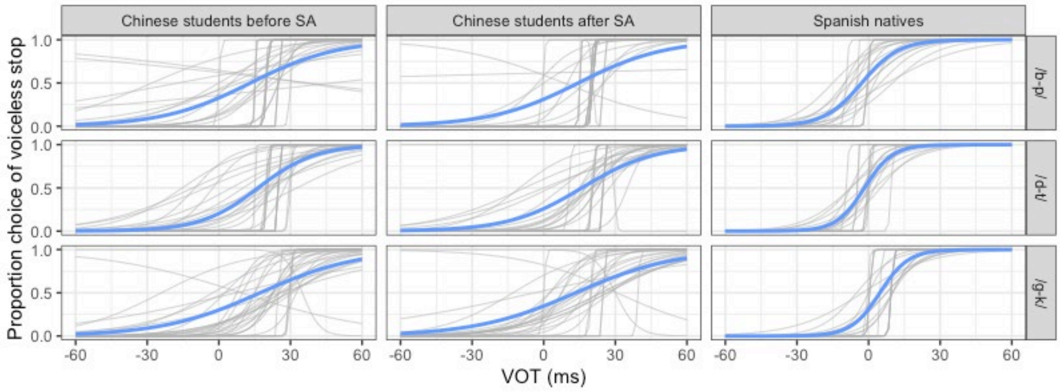


Figure 1. *Perception: Chinese students' stop categorization on three VOT continua (/b-p/, /d-t/, and /g-k/) before and after SA, in comparison with Spanish natives. Blue lines show group-level means, while grey lines show individual data.*

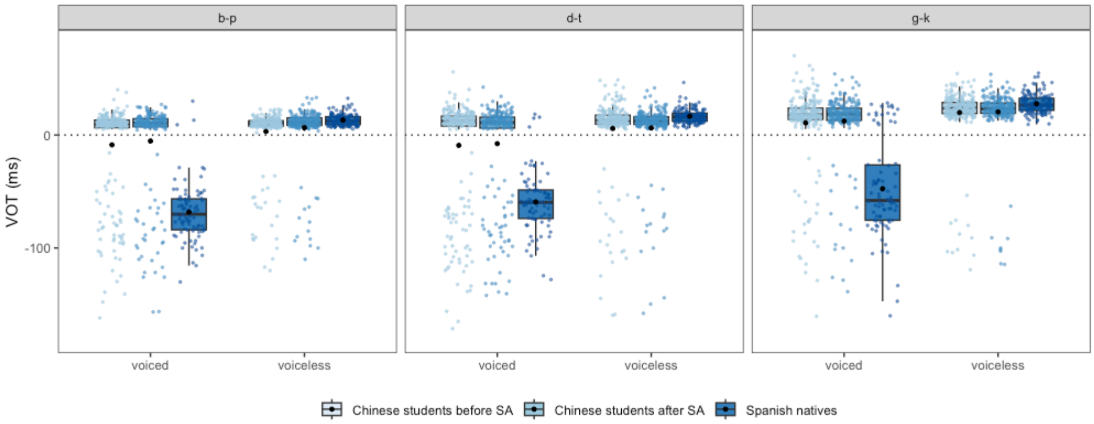


Figure 2. *Production: Boxplots showing Chinese students' VOT for three stop contrasts (/b-p/, /d-t/, and /g-k/) before and after SA, in comparison with Spanish natives.*

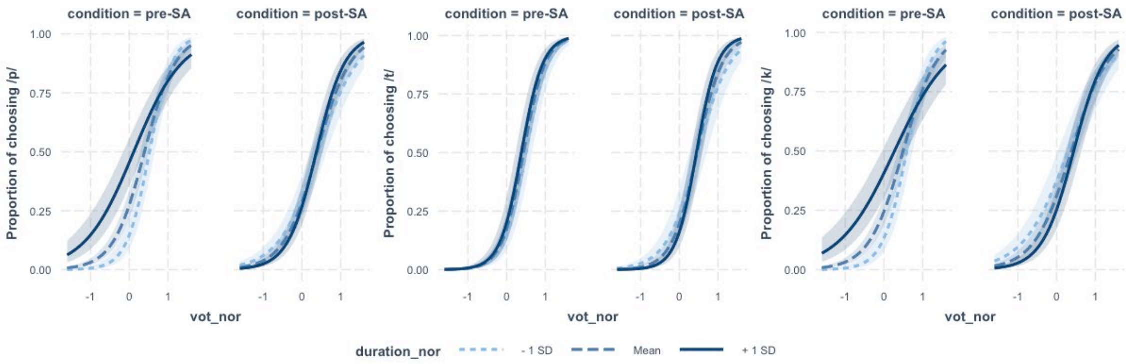


Figure 3. *Effects of individual differences in auditory processing ability for duration (normalized) on perception slopes for /b-p/ (left panel), /d-t/ (middle panel), and /g-k/ (right panel) at pre-SA and post-SA.*

[1] J. E. Flege, "Second language speech learning theory, findings, and problems," in *Speech perception and linguistic experience: Issues in cross-language research*, York Press, 1995.
[2] J. Turner, "The role of L2 input in developing a novel L2 contrast phonetically and phonologically: Production evidence from a residence abroad context," *Second Language Research*, pp. 1–31, 2023.
[3] P. Avello and A. R. Lara, "Phonological development in L2 speech production during study abroad programmes differing in length of stay," in *Language Acquisition in Study Abroad and Formal Instruction Contexts*, C. Pérez-Vidal, Ed., John Benjamins Publishing Company, 2014.
[4] X. Xi and P. Li, "Spanish stops and their allophones produced by proficient Mandarin learners of Spanish," in *Proceedings of O-COCOSDA 2022*, Hanoi, Vietnam: IEEE, Nov. 2022, pp. 1–6.
[5] I. Mora-Plaza, K. Saito, Y. Suzukida, J.-M. Dewaele, and A. Tierney, "Tools for Second Language Speech Research and Teaching." 2022.
[6] P. Li, S. Sepanski, and X. Zhao, "Language history questionnaire: A Web-based interface for bilingual research," *Behavior Research Methods*, vol. 38, no. 2, pp. 202–210, 2006.

2. Are rhythmic differences relevant in the perception of L2 Portuguese?

Marisa Cruz, Chao Zhou
University of Lisbon

Rhythmic patterns may change as a function of L2 proficiency (e.g., [1]). However, such change in L2 rhythm production may not be perceptually identified by native listeners of the target language ([2, 3]), questioning the relevance of speech rhythm at the interplay of prosodic cues in L2 speech.

In this study, we assessed whether native listeners of European Portuguese (EP), which displays a mixed rhythmic nature ([4]), can perceptually detect the rhythmic difference in L2 Portuguese spoken by L1-Mandarin (syllable-timed – [5]) learners with different proficiency levels (intermediate – INT, and advanced - ADV). Although a previous production study reported that the L2 Portuguese rhythm by L1-Mandarin learners develop from more syllable-timed patterns to more target ones ([6]), it remains unclear whether EP native listeners are sensitive to such L2 rhythmic development.

An AX discrimination task was set up following [7] and performed by 25 EP native listeners (aged 19-65). L2 speech samples were first selected from a corpus by [8], who recorded 61 L1-Mandarin learners (30 advanced and 31 intermediate) reading the Portuguese version of “The North Wind and the Sun”. We recorded the same text by 10 EP native speakers (NAT). These speech materials were then segmented into IPs and low-pass filtered at a frequency of 400Hz. The F0 contour was flattened in order to eliminate intonational cues from the signal. EP listeners were told that they would be listening to acoustically modified sentences of 3 different exotic languages: Vedda (L1 Portuguese), Waigali (L2-INT Portuguese), and Zazaki (L2-ADV Portuguese). After the familiarization phase in which they listened to 4 sentences of each group, isolated and in same/different pairs, and had to perform on 14 discrimination trials with feedback, the participants performed on 30 trials without feedback (6 same trials, 12 intermediate-native and 12 advanced-native trials) in the test phase.

As illustrated in Figure 1, Portuguese native listeners were better at discerning the rhythmic difference in the intermediate-native group than that in the advanced-native group. This was confirmed in a mixed-effects linear regression on D-prime scores (b=0.44, 95%CI[0.27 , 0.61], p<.001). We further explored which rhythmic metrics (%V - proportion of vocalic intervals and deltaC - duration variability of consonantal intervals – [9] –, and their rate-normalized versions VarcoV and VarcoC – [10]) account for our perceptual results. Mixed-effects logistic regressions revealed a significant effect of VarcoV and VarcoC on EP native listeners' discrimination accuracy.

Different from prior perceptual studies ([2, 3]), these results show that native listeners of the target language are sensitive to the development of L2 speech rhythm, and their rhythmic discrimination is predicted by rate-normalized durational variability in L2 Portuguese produced by learners' different proficiency levels.

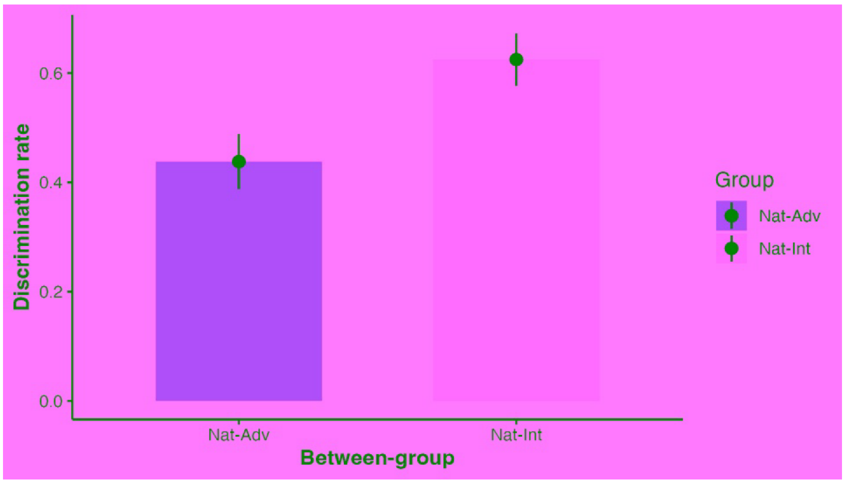


Figure 1. *Proportion of correct responses per condition: Native-Advanced L2 and Native-Intermediate L2.*

[1] Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, *42*, 244-257. doi:10.1016/j.system.2013.12.004

[2] Ordin, M & Polyanskaya, L. (2015). Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology, 6*, 316. doi: 10.3389/fpsyg.2015.00316

[3] van Maastricht, L., Zee, T., Krahmer, E., & Swerts, M. (2021). The interplay of prosodic cues in the L2: How intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Communication*, *133*, 81–90.

[4] Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus, 13*, 247-273.

[5] Lin, H., & Wang, Q. (2007). Mandarin rhythm: An acoustic study. *Journal of Chinese Linguistics and Computing*, *17*(3), 127-140.

[6] Zhou, C., Cruz, M., & Frota, S. (2017). O ritmo da interlíngua na produção do Português Europeu por falantes chineses. *Revista da Associação Portuguesa de Linguística, 3*, 423-435. DOI: https://doi.org/10.26334/2183-9077/rapln3ano2017a22

[7] Cruz, M. (2013). *Prosodic variation in European Portuguese: phrasing, intonation and rhythm in central-southern varieties*. Unpublished PhD Dissertation. University of Lisbon, Lisbon. https://repositorio.ul.pt/handle/10451/9893

[8] Zhou, C. (2021). L2 speech learning of european portuguese /l/ and /ɾ/ by L1-mandarin learners: experimental evidence and theoretical modelling. Unpublished PhD Dissertation. University of Lisbon, Lisbon. https://repositorio.ul.pt/handle/10451/55523

[9] Ramus, F., Mehler, J., & Nespor, M. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292.

[10] Dellwo, V. (2006). Rhythm and speech rate: a variation coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.), *Language and language processing* (pp. 231-241). Peter Lang.

3. Learning second language lexical tone discrimination: how is incremental cue training modulated by task difficulty?

Yanyu Li, Laurence White, Ghada Khattab
Newcastle University

Learners' first languages (L1) affect their perceptual sensitivity to acoustic cues relevant to meaningful second language (L2) contrasts [1]. Some L2 training methods attempt to change listeners' perceptual cue weighting, e.g., using exaggeration of acoustic dimensions to direct attention to underweighted cues [2]. Such methods have largely focused on segmental cues: here we test their efficacy in the tonal domain, specifically for learning lexical tone in languages where *F0* direction is a primary cue (e.g., Mandarin). Speakers of non-tonal languages (e.g., English) have been shown to be less perceptually sensitive than L1 speakers to *F0* direction cues (e.g., in identification tasks) [3]. Here we test the effect of incrementally exaggerating *F0* direction in tone discrimination training. Previous segmental studies [4, 5] found advantages of cue exaggeration over natural stimuli in difficult training contexts: thus we also attempted to manipulate task difficulty by varying *F0* height for one training group, but not the other.

We trained and tested 166 L1 English listeners on four pseudo-tones synthesised based on Mandarin, and primarily contrasting on *F0* direction (Fig. 1). Using a 2x2 design, we varied both Training Type (**Incremental** vs **Fixed**) and F0 Height (No variability **Nv** vs With height variation **Wv**). In the **Nv** condition (N=86), all tonal stimuli had the same tonal *F0* onset values per speaker; in the **Wv** condition (N=80), onset *F0* values for each tonal stimulus varied both within and between trials, whilst tonal slopes remained unchanged. Participants in the two *F0* height variability conditions were randomly assigned into incremental and fixed training groups. The **Incremental** group was trained on tonal stimuli (*high-rising 45* vs *high-falling 41*) where *F0* movement relevant to *rising/falling* was exaggerated at training outset and reduced blockwise (Fig. 1). The **Fixed** group was trained on baseline tonal stimuli of the same tone pair (i.e., the same pseudo-tones as the final training block for the **Incremental** group: Fig. 1). In pretest and posttest, both groups were tested on an untrained tonal contrast *level 44 vs falling-rising 425*. Each testing/training trial consisted of an ABX task (Instruction: *Is Sound X the same as Sound A or Sound B?*), with visual feedback (correct/incorrect) in training blocks only.

Test and training data were analysed separately. Response accuracy was analysed through mixed-effect logistic regression models, with fixed effects of Training Type (**Fixed** & **Incremental**), *F0* Height (**Nv** & **Wv**), and an ordered factor of Block (training 1, 2, 3; pretest & posttest). There were by-participant and by-syllable random intercepts, and a block-by-participant random slope. Effects were tested via likelihood ratios in backwards model comparison. For training blocks (Fig. 2 left), the **Incremental** group were more accurate than the **Fixed** group, $\chi2(1) = 21.77$, $p < .001$. Additionally, the effect of *F0* Height, $\chi2(1) = 14.15$, $p < .001$, indicated lower accuracy for the variability group (**Wv**). There was also an effect of Block, $\chi2(2) = 10.86$, $p = .004$, and a Training Type x Block interaction, $\chi2(2) = 34.25$, $p < .001$. Pairwise tests showed that accuracy decreased over blocks for the **Incremental** group, but increased for the **Fixed** group. For test blocks (Fig. 2 right), there was likewise lower accuracy for the *F0* Height variability (**Wv**) group, $\chi2(1) = 28.19$, $p < .001$. Accuracy improved by Block (pretest to posttest), $\chi2(1) = 28.19$, $p < .001$. Finally, model coefficients showed a *F0* Height x Block interaction over the test blocks, $\chi2(1) = 28.19$, $p < .001$, was due to a smaller improvement in the **Wv** group, $\beta = -0.25$, $SE = 0.1$, $p = .01$.

Both tone training methods (**Fixed** vs **Incremental**) improved discrimination, but the perceptual boost in Incremental training was not sustained when exaggeration was removed. Moreover, there was no boost to incremental training with greater *F0* Height variability. Ongoing work is examining whether the immediate perceptual boost from cue exaggeration can be reinforced and sustained given multiple separate training sessions.

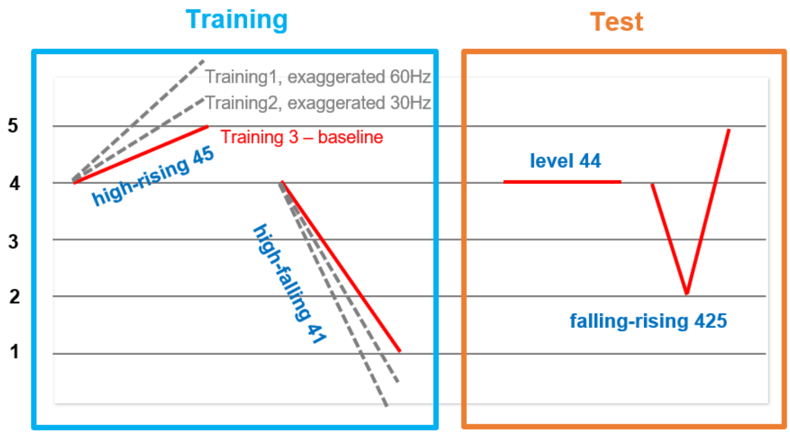


Figure 1. *The four pseudo-tones represented in Chao-digits [6]. High-rising 45 and high-falling 41 (shown with levels of exaggeration for the Incremental group) were used in training blocks; Level 44 and falling-rising 425 in test blocks.*

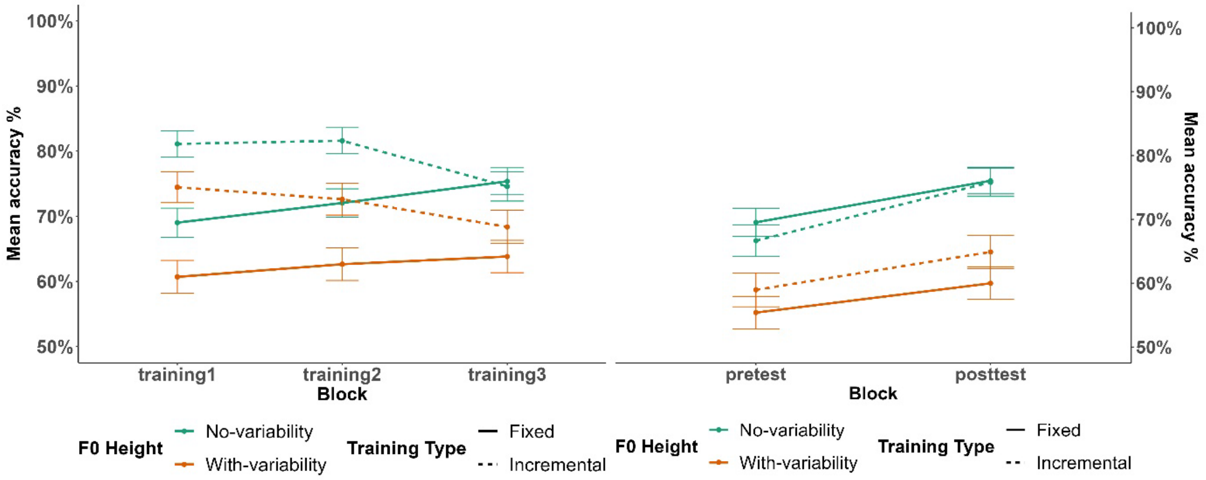


Figure 2. *Raw data on average accuracy (%) by Training Type (incremental & fixed) and F0 Height (Nv & Wv) in training blocks (training 1, 2, 3) (left) and test blocks (pretest, posttest) (right). Error bars indicate 95% confidence interval for the sample mean.*

[1] P. K. Kuhl and P. Iverson, ‘Linguistic experience and the “perceptual magnet effect”’, *Speech Percept. Linguist. Exp. Issues Cross-Lang. Res.*, pp. 121–154, 1995.

[2] P. Iverson, V. Hazan, and K. Bannister, ‘Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults’, *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 3267–3278, 2005.

[3] A. L. Francis, V. Ciocca, L. Ma, and K. Fenn, ‘Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers’, *J. Phon.*, vol. 36, no. 2, pp. 268–294, 2008.

[4] E. M. Ingvalson, L. L. Holt, and J. L. McClelland, ‘Can native Japanese listeners learn to differentiate /r–l/ on the basis of F3 onset frequency?’, *Biling. Lang. Cogn.*, vol. 15, no. 2, pp. 255–274, Apr. 2012, doi: 10.1017/S1366728911000447.

[5] H. Pashler and M. C. Mozer, ‘When does fading enhance perceptual category learning?’, *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 39, no. 4, pp. 1162–1173, 2013, doi: 10.1037/a0031679.

[6] Y.-R. Chao, ‘A system of tone letters’, *Maître Phon.*, 1930.

4. Obstruent Coarticulation Patterns of Ukrainian Speakers' L1 and L2 Production: An Ultrasound Study

Maiia Bulakh, Daniel Pape
McMaster University, Canada

This project aims to examine how Ukrainian speakers differ in their obstruent coarticulation patterns comparing the languages Ukrainian (L1), Russian (a language almost all Ukrainians are highly proficient and use daily) and L2 English. This study is a continuation of [1] where we showed that L2 English /g/ is produced more fronted compared to both Ukrainian and Russian /g/, and that L2 English /h/ is produced with significantly lower tongue positions compared to native counterpart Ukrainian /ɦ/. In this paper we present coarticulation comparisons, using articulatory ultrasound imaging, for /h x g/ in high /i u/ and low /a/ vowel contexts. The aim is to show (1) how different Ukrainian speakers vary in their consonantal tongue body positions when influenced by different following vowel contexts and (2) to compare the three described languages by speaker.

Method: We recorded Ultrasound articulation data (MicroUltrasound with convex 2-4Mhz transducer) and combined acoustics (Sennheiser ME63, Focusrite Scarlett) for the posterior consonants /ɦ x ɡ/ (Ukrainian), /x g/ (Russian) and /h g/ (English) for seven Ukrainian speakers (6 native, 1 heritage [HP]) for initial consonant position in /a i u/ vowel contexts in three repetitions in the soundproof booth at the Phonetics lab. The three languages were presented in blocks to the speakers (Ukrainian first, Russian last) on a computer screen with corresponding orthography prompts. We used the language background questionnaire, so speakers self-assessed their language skills and proficiency. Tongue contours were recorded using AAA software (v. 221.0.1) and traced using the DLC spline algorithm MobileNet 1.0 (v1.1.0)) at the articulatory phoneme midpoint with a following manual (fan shape) correction for algorithm errors and hyoid/mandible correction. Data was then further analyzed and plotted using polar SSANOVAs [2].

Results: All seven speakers showed consistency for producing all obstruents in a more fronted place of articulation when followed by the high front /i/ vowel for all languages. However, the high back vowel /u/ context did not lead to a more posterior place of articulation (for none of the speakers and languages in all /x h g/ productions), see Figure 1 for two examples.

With respect to differences in obstruent identity and language, we observe, for all three languages, that the voiced velar plosive /g/ is significantly less coarticulated compared to both (velar and glottal) fricatives /x h/. When comparing languages, it was observed for each speaker that there was no difference in coarticulation patterns for their L2 English /h/ compared to the native /ɦ/, for none of the three vowel contexts. The same consistent result (i.e. no difference between languages) was found when comparing Russian versus Ukrainian /x/ phonemes.

Finally, with respect to speaker variability, we find that three of the seven speakers (MB, NT, OG) do not show any coarticulatory differences for Ukrainian and L2 English /g/, however they show consistent coarticulation differences for their Russian /g/ productions (see Figure 2).

Conclusion: We found strong speaker variability for the coarticulation patterns for the velar stop, but no observable differences for the /h x/ fricatives comparing the three languages for each speaker. Furthermore, all Ukrainian speakers show consistent anterior articulation patterns when obstruents are followed by high front vowels, but no consistent posterior coarticulation for high back vowels. One possible reason for this difference could be that /u/ in Ukrainian is described as more centralized /ʉ/ position (see formant plots in [3]), similar to some American English dialects and/or sociophonetic varieties [4].

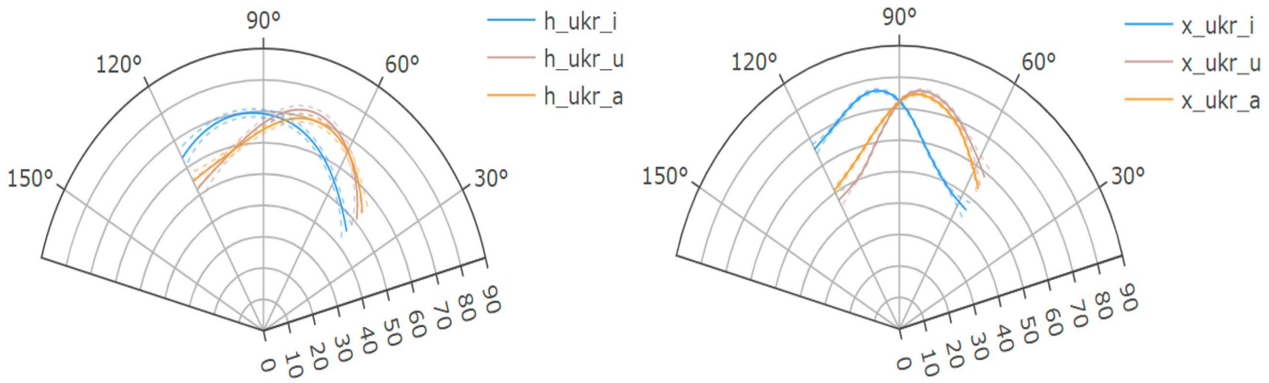


Figure 1: *Coarticulation patterns for the Ukrainian phoneme /ɦ/ for speaker HP (left), and Ukrainian /x/ for speaker NT (right). Vowel context is coded in colours (blue = /i/. purple = /u/, orange = /a/)/.*

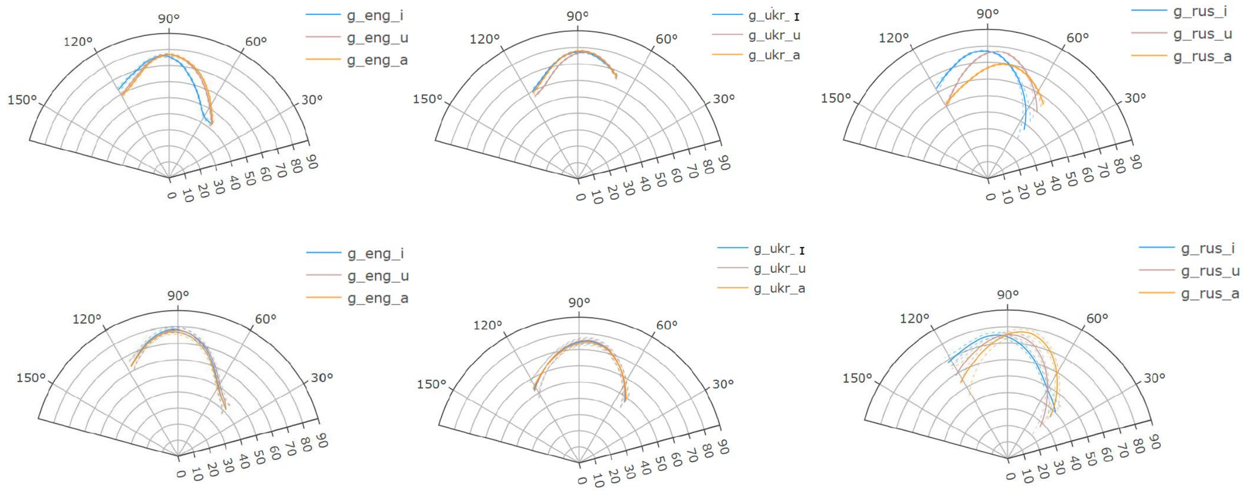


Figure 2: */g/ coarticulation patterns for speakers NT (upper panels) and OG (lower panels). English, Ukrainian and Russian language productions are shown in the left, middle and right columns, respectively. Vowel context is coded in colours (blue = /i ɪ/ and, purple = /u/, orange = /a/)/.*

[1] Bulakh M., Pape D. 2024. Ukrainian speakers' articulation of posterior consonants comparing Ukrainian, Russian and L2 English: an Ultrasound study. In Wilson, Mizoguchi, Perkins, Villegas, & Yamane (Eds), *Ultrafest XI: Extended Abstracts* (p.140–142), Aizu, Japan.

[2] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120(1), 407-415.

[3] Pompino-Marschall, B., Steriopolo, E, Żygis, M. 2017. Ukrainian. *Journal of the International Phonetic Association*, 47(3): 349-357. doi:10.1017/S0025100316000372.

[4] Hillenbrand J., Getty L., Clark M., Wheeler K. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97 (5): 3099–3111.

5. (Dis)fluency across languages and in L2: The case of Italian learners of German

Loredana Schettino[1], Simona Sbranna[2], Marta Maffia[3]
[1]Free University of Bozen, [2]University of Cologne, [3]University Orientale of Naples

In the CEFR [1], disfluencies are considered fluency-disrupting and their type and frequency distinguish proficiency levels. Research findings are instead more nuanced. For example, [2] reported that hesitation types vary across proficiency levels, [3] observed no difference in types of disfluencies but in their frequency and position, while [4,5] indicated that advanced learners tend to use more filled pauses than beginners. Nevertheless, most research has only addressed individual aspects of disfluencies, overlooking the fact that their functions may even enhance fluency supporting discourse organization [6]. Indeed, speakers may be disfluent to revise previous statements or use pauses, fillers, and prolongations to take time [7]. To provide a more complete picture of disfluency phenomena in L2, we conduct an exploratory, but comprehensive analysis of Italian learners of German in their L1 and L2 across proficiency levels, and compared to their target language.

We analysed 10 Map Task dialogues [8, from 9]: four by Italian students of L2 German in their L1 and L2 (two beginners – BEG, and two advanced learners – ADV) and two by native Germans. We annotated disfluencies contextually [10] including type (Insertion, Deletion, Substitution, Repetition, Silent Pause, Prolongation, Filled Pause, Lexicalized Filled Pause), macro-functions (Backward-Looking or Forward-Looking disfluencies, henceforth BLD and FLD) and specific functions of FLD (Word Searching, Structuring, Focusing, Interactional or Hesitative). We compared across groups the frequency of disfluencies, their type, their macro- and specific functions. We tested the results with LLM with speakers as random effect.

Disfluencies are significantly more in learners L2 as in both L1s independently of proficiency, even though Germans seem to produce more disfluencies than Italians (ADV: L1=217; L2=433; BEG: L1=233; L2=428; GER L1=362). Coming to disfluency types and macro-functions (Fig.1), the two L1s have a similar distribution of BLD and FLD as opposed to learners who have more FLD than both L1s. The L1s show only non-significant differences in the distribution of types of disfluencies fulfilling these functions. Comparing the L1s to the L2 groups, both L1s show more variety as in L2 lexicalised filled pauses are almost absent ($p=0.004$), and substitutions, insertions and deletions are highly reduced. Moreover, beginners show more silent pauses ($p=0.01$) than advanced learners and less prolongations ($p=0.007$), whereas advanced learners have a similar number of silent pauses as in both L1s. Regarding the specific function of FLD (Fig.2), the L1s differ a lot in the proportion of the hesitative function, while they are similar in the number of word-searching functions. Learners seem to transfer their L1 FLD pattern to the L2 but with some differences: increased word searching ($p=0.0007$) and reduced focusing ($p<0.0001$) and structuring ($p<0.0001$) disfluencies, especially in beginners.

This exploratory dataset reveals that, while disfluencies are more frequent in German than Italian, both L1s display similar distributions in type and function (the notable difference in hesitative tokens may reflect dyad-specific preferences and should be viewed as preliminary due to the small sample). Proficiency levels show even higher disfluency frequency than L1 German, with no notable differences in disfluency type and function, except for silent pauses, suggesting that fluency improvements do not align with greater L2 knowledge. Neither learner group uses lexicalized filled pauses frequently, likely due to unfamiliarity with target-language-specific tokens. Non-nativeness is also evident in more frequent word searches and fewer tokens for focus and structure – advanced communicative and discursive features. Overall, in classroom settings, learners fail to reach native-like fluency or adopt disfluency patterns similar to those in either the native or target language, challenging CEFR expectations and prior studies. A larger participant sample is needed to further validate these findings.

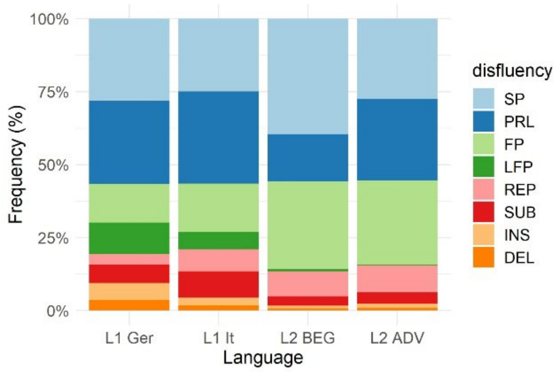


Figure 1: *Frequency of disfluency phenomena per language and proficiency. Macro-functions are colour-coded: the greens and the blues display forward-looking disfluencies, while the reds and oranges show backward-looking disfluencies*

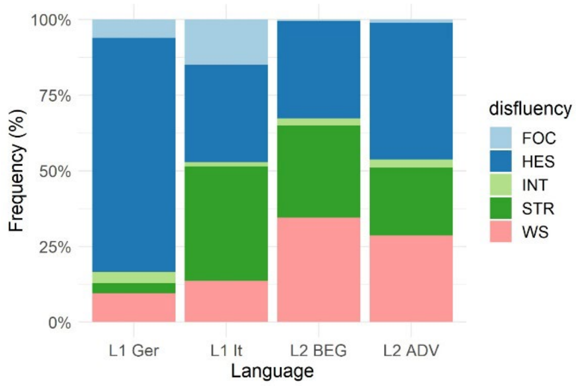


Figure 2: *Frequency of specific functions of forward-looking disfluencies per language and proficiency*

[1] Council of Europe (2020), Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume, Council of Europe Publishing, Strasbourg.

[2] Ishikawa, T. (2014). Distribution of different types of hesitation phenomena and their patterns of correlations. *創価女子短期大学紀要*, (45), 169-192.

[3] Wang, Y. B. (2021). A Study on the Use of Hesitation Markers in Varied-Level EFL Learners' L2 Speaking Process. *Open Journal of Modern Linguistics*, 11, 823-840.

[4] García-Amaya, L. (2015). A longitudinal study of filled pauses and silent pauses in second language speech. *Papers presented at Diss*, 23-27.

[5] Liao, J. (2023). Disfluency and self-repair in presentational and interpersonal speech modalities. *Foreign Language Annals*, *56*(2), 401-427.

[6] Mutta, M., Lintunen, P., & Peltonen, P. (Eds.). (2020). *Fluency in L2 Learning and Use.* Multilingual Matters.

[7] Ginzburg, J., Fernandez, R., & Schlangen, D. (2014). Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics*, 7, 9–1.

[8] Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., ... & Weinert, R. (1991). The HCRC map task corpus. *Language and speech*, 34(4), 351-366.

[9] Sbranna, S. (Forthcoming). Prosody and interactional fluency of Italian learners of German. (EuroSLA Studies). Berlin: Language Science Press.

[10] Schettino, L. (2022). The role of disfluences in Italian discourse. Modelling and speech synthesis applications. PhD Thesis, University of Salerno, Italy.

6. Structuring discourse with pitch contours – intonation contours of discourse particles in bilingual heritage speakers' two languages

Marlene Böttcher, Margaret Zellers
Kiel University

This exploratory study investigates heritage speakers' prosodic realization of different kinds of particles occurring at boundaries between broader units of discourse, where they are interpreted as indicating frame shifts [1]. Heritage speakers (HSs) are bilingual speakers of a minority or heritage language (HL), spoken in informal familial settings, as well as a majority language (ML), dominant in different areas of everyday life, who have been reported to sound native like and show similar phonology, while also showing discrepancies from monolingual native speakers of their languages in prosody and pragmatics [2]. Filler particles (FPs) like *uhm*, discourse markers (DMs) like *yeah* and discourse connectors (CONs) like *and* share discourse pragmatic functions related to discourse structuring, while being syntactically not integrated and semantically bleached or empty [3, 4]. Prior research on bilingual speakers' use of these kinds of discourse particles has found differences in frequency and form compared to monolinguals [5] and has identified a possible influence of the dominant language on the HL (e.g. [6]). Additionally, [7] found evidence for different particles in different communicative situations, with more FPs in formal and more DMs in informal contexts. This study explores the prosodic similarities and differences of these particles in heritage speakers' discourse.

Data analysis included narrations from the RUEG corpus [8] by 10 female HSs of Russian in Germany and in the U.S. in both their languages (ML: English/ German, HL: Russian). For comparison, narrations by 5 monolingual female speakers of each of the languages were also analyzed. The narrations in the corpus are produced in two different scenarios differing in formality and are annotated for major discourse boundaries differentiating openings and closings from the main narration. The pitch contour of particles occurring at such a discourse boundary was analyzed using a contour clustering script [9].

The produced pitch contours of discourse particles clustered in two basic patterns: level and falling intonation. The distribution of these intonation contours is presented in Figure 1 below. Across HSs groups and monolinguals the most frequent contour is level (n = 89) while falling is less frequent (n = 27). In the formal narrations FPs are more frequently produced across all groups and languages predominantly with a level contour. In the informal narrations DMs and CONs are more frequently produced at discourse boundaries. In this position these particles share prosodic features with FPs with mostly level intonation. The bilingual HSs in the U.S produce falling contours more frequently in their Russian than in their English, while HSs in Germany also produce this contour in their German. Overall there does not seem to be an influence in pitch contour from one of the bilingual speakers' languages on the other based on the analyzed data.

These results show that the prosodic form of discourse pragmatic markers is comparable across languages and across mono- and bilingual speakers and add to an understanding of bilingual HSs as native speakers of both their languages [10]. Future analyses will address the variation in the pitch contour which could be modulated by their lexical form and the immediate discourse context. The choice of the particle in our data is influenced by the situational context (formal vs. informal narrations) showing more variation in bilingual HSs. This study provides further evidence for the use of filler particles across languages similarly to other discourse pragmatic markers at discourse boundaries [11] and adds an investigation and comparison of their prosodic form.

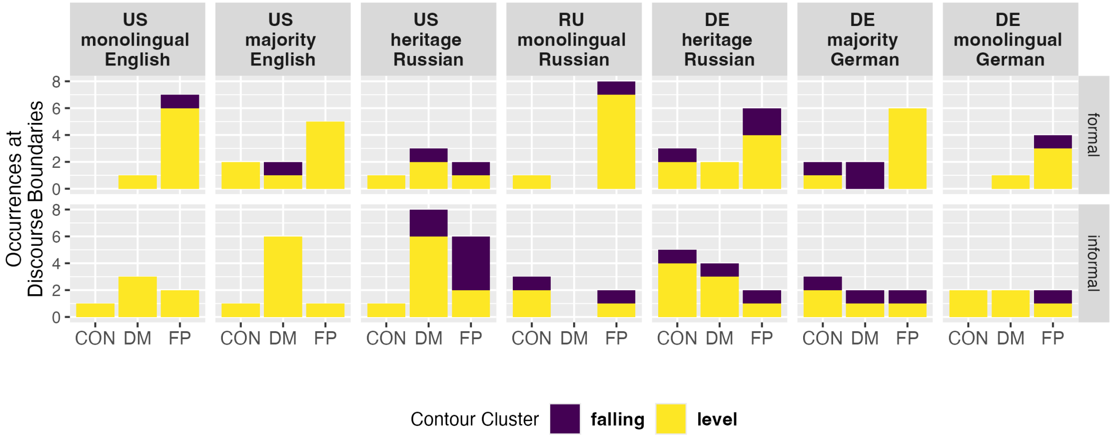


Figure 1. *Distribution of intonation contour clusters across formal and informal narrations, in different languages and speaker groups in the U.S., in Russia (RU) and in Germany (DE).*

[1] Maschler, Y., & Schiffrin, D. (2015). Discourse markers language, meaning, and context. The handbook of discourse analysis, 189-221.

[2] Polinsky, M. (2018). Heritage languages and their speakers (Vol. 159). Cambridge University Press.

[3] Swerts, M. (1998). Filled pauses as markers of discourse structure. Journal of pragmatics, 30(4), 485-496. ISO 690

[4] Crible, L., & Bunt, H. (2016). Discourse markers and disfluencies: Integrating functional and formal annotations. In Proceedings of the LREC 2016 Workshop ISA-12 (pp. 38-45)..

[5] Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference. Pragmatics and corpus linguistics: A mutualistic entente, 2, 119-149

[6] Lo, J. J. (2020). Between Äh (m) and Euh (m): The distribution and realization of filled pauses in the speech of German-French simultaneous bilinguals. Language and speech, 63(4), 746-768

[7] Böttcher, M., & Zellers, M. (2024). When "uhm","and" and "yeah" sound the same—prosodic aspects of discourse pragmatic markers in American English. In *Proc. SpeechProsody 2024* (pp. 906-910)]

[8] Wiese, H., Alexiadou, A., Allen, S., … Zuban, Y. (2019). RUEG Corpus. Zenodo. https://doi.org/10.5281/zenodo.3236069

[9] Kaland, C. 2023. Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association* 53(1), 159-188. doi:10.1017/S0025100321000049

[10] Wiese, H., Alexiadou, A., Allen, S.,... Zuban, Y. (2022). Heritage speakers as part of the native language continuum. *Frontiers in psychology*, *12*, 717973.

[11] Kosmala, L., & Crible, L. (2022). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. Language and Speech, 65(1), 216-239.

7. The relationship between acoustic features of second language speech and holistic intelligibility ratings

Wai Ling Law, Graham England
The Hong Kong University of Science and Technology

Despite the significance of speech intelligibility in second language (L2) acquisition, teaching and assessments, the notion remains largely elusive: there is to date no consensus on the definition of speech intelligibility or how it should be measured [1, 2, 3]. In its narrow sense, some researchers have defined speech intelligibility as the extent to which a speaker is understood by a listener, differentiating intelligibility from accentedness. However, the notions of low intelligibility and heavy accents are often confounded in listeners' ratings [4]. Specifically, listeners' familiarity with accented speech has been found to influence intelligibility ratings [5, 6]. Therefore, the present study examines what acoustic-phonetic properties of speech may be associated with higher and lower intelligibility, as rated holistically in the context of a standardized university speaking test, to reduce the potential impact of listeners' familiarity with the L2 accent on intelligibility ratings.

Twenty de-identified samples were drawn from the database of a criterion-referenced and performance-based English speaking test at a university in Hong Kong, a test designed to assess research postgraduate students' English speaking proficiency and their qualifications to conduct classroom teaching in English. In the test, examinees respond to questions asked by the examiner(s). Their responses are evaluated by university language instructors on four areas, one of which is speech intelligibility, on a scale from 0 to 7. Examinees are required to obtain an overall score of at least 4 to be eligible for classroom teaching. All the responses are digitally recorded in a quiet testing room. The present study analyzed 10 samples of higher intelligibility from examinees obtaining intelligibility scores from 5 to 7 and 10 samples of lower intelligibility from those attaining intelligibility scores from 2 to 3. The duration of the samples ranged from 412 to 561 seconds. To control for the effect of L2 accent, all examinees were native speakers of Chinese. All the examiners were experienced university English instructors teaching Chinese students, allowing us to control for their familiarity with Chinese influence on English pronunciation. Speech samples were measured acoustically on vowel durations, vowel space, pitch (f0) range, and nine measures on speech fluency, all of which have been found to be associated with intelligibility ratings in previous studies [7, 8].

Consistent with previous findings, results indicated that more intelligible speakers produced significantly higher mean f0 range than the less intelligible ones. However, contrary to previous findings, more intelligible speakers produced significantly shorter vowel durations than the less intelligible ones, presumably because of their faster speech rate. There was also no significant difference between the two groups in vowel space. As for the measures on speech fluency, the more intelligible group spoke with a significantly higher speech rate and produced significantly more syllables per run than did the less intelligible group. Interestingly, in investigating the occurrence of pauses, the two groups differed in the production of silent pauses only, but not filled pauses (e.g., *hmm*). The more intelligible group spoke with significantly shorter mean silent pause time and lower silent pause ratio than did the less intelligible group. However, the two groups did not differ significantly in the number of filled pauses produced, mean filled pause time, and filled pause ratio. Taken together, these results seem to suggest that at least for trained listeners, acoustic measures related to suprasegmental properties may be more relevant to speech intelligibility than those related to segmental properties. These findings may imply that L2 pronunciation instruction should prioritize fluency to help learners enhance their speech intelligibility in L2.

[1] Munro, M. J. 2008. Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193 – 218). Amsterdam / Philadelphia: John Benjamins Amsterdam.

[2] Pickering, L. 2006. Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics* 26, 219 – 233.

[3] Sewell, A. 2010. Research methods and intelligibility studies. *World Englishes* 29(2), 257 – 269.

[4] Munro, M. J. & Derwing, T. M. 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 49(s1), 285 – 310.

[5] Gass, S. M. & Varonis, E. M. 1984. The effect of familiarity on the comprehensibility of non-native speech. *Language Learning* 34, 65 – 89.

[6] Wingstedt, M. & Schulman, R. 1987. Comprehension of foreign accents. In W. Dressler (Ed.), *Phonologica 1984: Proceedings of the Fifth International Phonology Meeting, Eisenstadt, 25 – 28 June 1984* (pp. 339 – 344). Cambridge: Cambridge University Press.

[7] Bradlow, A. R., Torretta, G. M. & Pisoni, D. B. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20(3), 255 – 272.

[8] Liu, H.-M., Tseng, C.-H. & Tsao, F.-M. 2000.Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Clinical Linguistics & Phonetics* 14(6), 447 – 464.

8. Phonetic similarity and acquisition challenges of Spanish rhotics among Chinese L2 learners: A theoretical and analytical approach

Dingning Zhang, Wendy Elvira-García
Universitat de Barcelona

The acquisition of Spanish rhotic consonants, including the alveolar tap and trill, presents significant challenges for Chinese learners and often requires prolonged effort to master [1]. According to the Speech Learning Model (SLM) proposed by Flege [2] and the Perceptual Assimilation Model for Second Language Acquisition (PAM-L2) developed by Best et al. [3], these difficulties are could be explained by cross-linguistic phonetic similarities between the rhotics in the two languages. While SLM focuses on advanced L2 learners and bilinguals on production, PAM-L2 emphasizes perception among late L2 learners, functional monolinguals and naïve L2 listeners [4].

This study conducts a comprehensive review of 42 studies on Chinese and Spanish rhotics, summarizing the phonetic features of all rhotics in Chinese, [6], [7], [8], [9], [10], [11], [12], [13], [14], [15] and Spanish [9], [16], [17], [18], [19], [20], [21]. It then examines cross-linguistic phonetic similarities in three key rhotic correspondences: (1) Chinese syllable-initial /ɻ/ and Spanish tap /ɾ/, (2) Chinese syllable-initial /ɻ/ and Spanish trill /r/, and (3) Chinese syllable-final /ɚ/ and Spanish tap /ɾ/. These comparisons are analyzed through five frameworks: IPA Comparison (IC), Acoustic Difference (AD), Feature Redeployment (FR), Articulatory Similarity (AS), and Phonological Similarity (PS) [5]. The findings are interpreted within the frameworks of SLM and PAM-L2 to provide a comprehensive understanding of how Chinese L2 learners acquire Spanish rhotics [4], [5].

IC and AD analyses reveal low similarity between these rhotic pairs, considering the Spanish rhotics as “new” sounds within the SLM framework. This suggests that successful acquisition depends on learners’ increasing ability to distinguish L1-L2 contrasts with greater L2 exposure. In contrast, FR, AS, and PS analyses show that the rhotics are categorized as either “equally distant” or “equally uncategorized” in PAM-L2, leading to poor discrimination and significant learning difficulties. Regarding transfer effects, this study identifies negative transfer from Chinese lateral, rhotic, and "er" coda sounds, influencing the production of Spanish rhotics across various syllabic contexts. Intriguingly, tap accuracy correlates with Spanish proficiency, while no such relationship is observed for trills, but once trills are acquired, their accuracy tends to be more stable and more closely aligned with native speakers’ production compared to taps. As for the order of acquisition, our study identifies this predictable acquisition sequence: lateral → tap → trill. The findings justify that PAM-L2, provides a more effective framework for predicting the L2 rhotic acquisition than SLM. Moreover, factors such as task complexity, phonological environment, regional Chinese dialects, the orthographic representation of Pinyin, and English experiences have predominantly exerted a negative influence on the L2/L3 realization of Spanish rhotics.

[1] Jiao, H., & Li, Y. 2021. Difficulties and solutions for Chinese people in learning big tongue sound: Taking Spanish as an example. *Time Report*, 106–107.
[2] Flege, J. E. 1987. The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47-65.
[3] Best, C.T., Tyler, M.D., Bohn, O., & Munro, M.J. 2007. Nonnative and second-language speech perception : commonalities and complementarities.
[4] Yang, Y., & Chen, X. 2019. Within-organ contrast in second language perception: The perception of Russian initial /r-l/ contrast by Chinese learners. *J. Acoust. Soc. Am.* 1. 146 (2): EL117–EL123. https://doi.org/10.1121/1.5120549

[5] Yang, Y., Chen, X., & Xiao, Q. 2022. Cross-linguistic similarity in L2 speech learning: Evidence from the acquisition of Russian stop contrasts by Mandarin speakers. *Second Language Research*, 38(1), 3-29. https://doi.org/10.1177/02676583221063548
[6] Smith, J.G. 2010. Acoustic Properties of English /l/ and /ɻ/ Produced by Mandarin Chinese Speakers. [Doctoral dissertation]. University of Toronto.
[7] Lee, W.-S. 1999. An articulatory and acoustical analysis of the syllable-initial sibilants and approximants in Beijing Mandarin. *In Proceedings of the International Congress on Phonetic Sciences* (pp. 413–416).
[8] Chen, S. & Mok, P. 2021. Articulatory and Acoustic Features of Mandarin /ɹ/: A Preliminary Study, *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Hong Kong, 2021, pp. 1-5, doi: 10.1109/ISCSLP49672.2021.9362070.
[9] Chen, S., Whalen, D. & Mok, P. 2024. What R Mandarin Chinese /ɹ/s? – acoustic and articulatory features of Mandarin Chinese rhotics. *Phonetica*, 81(5), 509-552. https://doi.org/10.1515/phon-2023-0023
[10] Patience M. 2018. Acquisition of the Tap-Trill Contrast by L1 Mandarin–L2 English–L3 Spanish Speakers. *Languages*. 3(4):42. https://doi.org/10.3390/languages3040042
[11] Chen, S., Mok, P., Tiede, M., Chen, W., & Whalen, D.H.. 2017. Investigating the production of Mandarin rhotics using ultrasound imaging. *Ultrafest VIII*, Potsdam, Germany.
[12] Xing, K. 2022. Phonetic and phonological perspectives on rhoticity in Mandarin [Doctoral dissertation]. University of Manchester.
[13] Liu, H. 2019. Acoustic experiment and error analysis of erhua among British and American primary Chinese learners [Master's thesis], Hebei Normal University. Retrieved from CNKI.
[14] Wang, X. 2013. Experimental and Comparative Study on R-sound between American English and Standard Chinese [Master's thesis] Yanbian University. Retrieved from CNKI.
[15] Lee, W.-S. 2005. A phonetic study of the "er-hua" rimes in Beijing Mandarin. *Proceedings of Interspeech 2005*, 1093–1096. https://doi.org/10.21437/Interspeech.2005-433.
[16] Rodriguez, L.V. 2020. The perception of tap and trill in Spanish-English bilinguals (2022). Open Access Theses & Dissertations. https://scholarworks.utep.edu/open_etd/3581
[17] Hualde, J. I. 2013. *Los sonidos del español: Spanish Language edition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511719943
[18] Quilis, A. 1993. Tratado de fonología y fonética españolas. Editorial Gredos.
[19] Falgueras, B. B. 2002. Las vibrantes del español: Manifestaciones acústicas y procesos fonéticos. [Doctoral dissertation] Universitat Autònoma de Barcelona.
[20] Massone, M. 1988. Estudio acústico y perceptivo de las consonantes nasales y líquidas del español. Universidad Nacional de La Plata.
[21] Almeida, M. & Dorta, J. 1993. Datos acústicos de las líquidas españolas. In C. Díaz Alayón (Ed.), *Homenaje a José Pérez Pidal* (pp. 97-110).

9. Intonational marking of focus in L2 Italian and French and the markedness differential hypothesis

Bianca Maria De Paolis
Università di Torino

L2 prosody learning models typically consider the degree of similarity between a learner's L1 and L2. However, other factors, such as markedness and learning direction, have not been systematically examined in this context. The limited application of the Markedness Differential Hypothesis (MDH) [1] in prosody may stem from the complexity of defining 'markedness' in relation to sentence prosody. According to pioneering works [2, 3], a markedness scale can also be conceptualized for prosody: in this framework, structurally determined prosody is considered less marked than pragmatically determined prosody (see Fig. 1). To establish a language's position on the markedness scale, it is necessary to observe its accentuation, intonation, and discourse-level prominence marking systems. French is often characterized by high variability in the placement of prosodic prominence, due to the 'syncretism' between these features [4, 5, 6]. Italian, in contrast, exhibits a more regular and structurally determined prosodic and accentual system [7, 8, 9]. Based on these observations, Italian is classified as a language with structurally determined sentence prosody, while French occupies an intermediate position between structurally and pragmatically determined systems [2]. For the purposes of this study, this distinction implies that French is more marked than Italian. Following the MDH, it is predicted that francophone learners of L2 Italian will exhibit greater proficiency in implementing focus-marking prosody, compared to italophone learners of L2 French. This prediction aligns with the hypothesis that markedness influences L2 acquisition asymmetrically, making structurally simpler systems easier to acquire. Practically, this means we expect a higher frequency of successfully realized focus-marking prosodic contours in the L2 Italian group, while the L2 French group is anticipated to exhibit fewer instances of focus-related intonational marking.

To test these hypotheses, we conducted a study involving 15 francophone learners of L2 Italian, 15 francophone learners of L2 French, 15 native monolingual Italian speakers, and 15 native monolingual French speakers. Data were elicited using a protocol adapted from Gabriel [10], which targeted focus subjects under broad, identification, and correction focus conditions within SVO utterances. Each participant produced six target utterances, resulting in a dataset of 360 tokens covering all focus types. Prosodic patterns were analyzed using the Polytonia algorithm [11], ensuring neutral, language-independent annotation.

The results from native speakers revealed that broad focus subjects were realized without prosodic prominence in both Italian and French (see Fig. 2). Under narrow focus conditions, Italian speakers marked subjects prosodically in 50% of cases, regardless of focus subtype (identification or correction). In contrast, French speakers employed prominent contours in 20% of identification and 40% of correction contexts, showing significant functional differentiation. French speakers also exhibited great variability in f0 peak alignment, consistent with the characterization of French prosody as more variable and pragmatically driven [4, 5, 6]. Among L2 speakers, French learners of Italian approximated native Italian focus-marking patterns, using prominent f0 contours in 39% of identification and 60% of correction contexts. Conversely, Italian learners of French struggled with prosodic prominence, with rates not exceeding 21% in either narrow-focus condition. Unlike their French counterparts, they failed to express functional differentiation between identification and correction focus. Notably, Italian learners of French also exhibited an overall underuse of prosodic marking for focus constituents. This pattern aligns with our initial hypothesis, rooted in MDH: Italian prosody is more accessible for L2 learners, while French's system poses greater challenges. The consistent directionality of the asymmetry suggests that markedness plays a key role. Our findings highlight the importance of markedness in L2 acquisition and in designing effective teaching strategies for languages with divergent prosodic systems.

Keywords: L2 prosody, focus, Markedness Differential Hypothesis, French, Italian

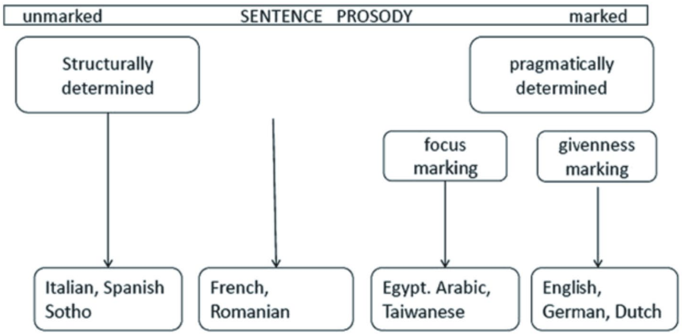


Figure 1. *Markedness scale of languages (from [3])*

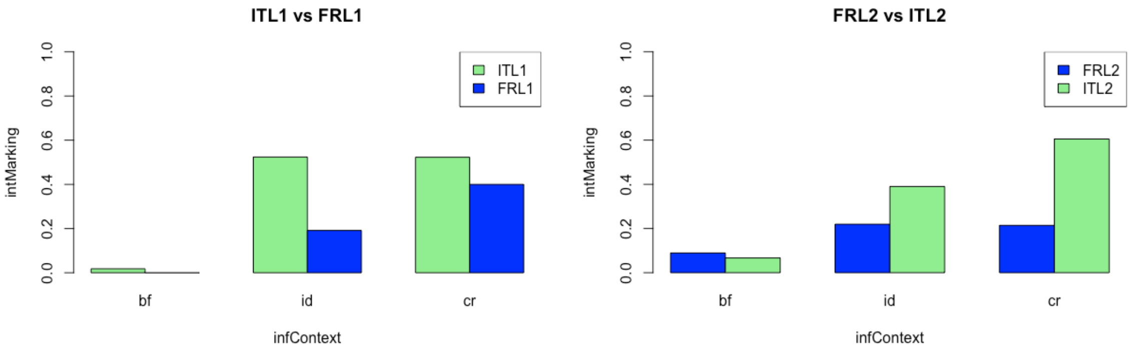


Figure 2. *Marked contours in broad, identification, correction focus utterances for target subjects: L1 groups on the left and L2 groups on the right*

[1] F. R. Eckman, “Markedness and the contrastive analysis hypothesis,” *Language Learning*, vol. 27, 2, 1977.
[2] L. Rasier and P. Hiligsmann, “Prosodic transfer from L1 to L2. Theoretical and methodological issues,” *Nouveaux cahiers de linguistique française,* vol. 28, 2007.
[3] S. Zerbian, “Markedness considerations in L2 prosodic focus and givenness marking,” in *Prosody and Language in Contact. Prosody, Phonology and Phonetics*, E. Delais-Roussarie, M. Avanzi, and S. Herment, Eds. Berlin, Heidelberg: Springer, 2015.
[4] J. Vaissière, “Cross-linguistic prosodic transcription: French vs. English,” in *Problems and methods of experimental phonetics*, Eds. St Petersburg State University Press, 2002.
[5] M. Rossi, “Le français, langue sans accent?” in *L’accent en français contemporain*, I. Fonagy and P. P. Leon, Eds. Montréal: Didier, 1980.
[6] P. Martin, “L’intonation du français : le vilain petit canard parmi les langues romanes ?” *Language design: journal of theoretical and experimental linguistics*, 2008.
[7] P. Garde, *L’accent*. Paris, Presses universitaires de France, 1968.
[8] A. L. Lepschy & G. Lepschy, *The Italian Language today*. London: Hutchinson & Co., 1977.
[9] P. M. Bertinetto, *Strutture prosodiche dell’italiano*. Firenze: Accademia della Crusca, 1981.
[10] C. Gabriel, “On Focus, Prosody, and Word Order in Argentinian Spanish. A Minimalist OT Account,” In *Revista Virtual de Estudos da Linguagem*, vol. Special issue 4, 2010.

[11] P. Mertens, “Polytonia : a system for the automatic transcription of tonal aspects in speech corpora,” *Journal of Speech Sciences*, vol. 4(2), 2014.

10. Individual differences in predictive prosodic processing in French

Caterina Petrone[1], Christelle Zielinski[1,2], Maud Champagne-Lavau[1]
[1]*Aix-Marseille Univ, CNRS, LPL, Aix-en-Provence, France,* [2]*Aix-Marseille Univ, ILCB, Aix-en-Provence, France*

Listeners use their knowledge about the association between accentuation and information structure to predict the words they are about to hear [1, 2]. However, the use of accentuation to infer contrastive meaning may differ across listeners [3], which could lead to different strategies when relying on these cues for lexical prediction.

In this study, we explored whether, in French, individual differences in the predictive use of accentuation exist. French is a language with fixed stress, in which speakers signal a contrast between two words, e.g., via syntactic or prosodic phrasing, and via accentuation [4, 5]. An emphatic accent can occur at the beginning of the words, depending on the speaker's choice [5]. The focal region can be preceded by "pre-focal" acoustic cues, such as compressed pitch range or lower pitch [5]. It is unclear, though, whether listeners rely on the emphatic accent for lexical prediction, despite the high level of variability to mark emphasis.

In a question-response game, we monitored eye movements to lexical competitors (**ca**nard/**ca**non, "duck/cannon") during question comprehension as well as latencies of speech responses. Thirty-three French participants replied to 24 trials consisting of a sequence of two pre-recorded questions. Questions were produced by a French prosodist and controlled for their intonation. The first question asked whether the location of one object was above or below a geometric shape (*Est-ce que le canard est au dessus du rond?*, "Is the duck above the circle?"), and the following question (*Et est-ce que le canard/canon est en dessous du carré?* "And is the duck/cannon below the square?") asked the location either of the previously mentioned object ("given" condition) or of a new object ("contrastive" condition). The object name could either carry an emphatic accent on the word-initial syllable or not. If the emphatic accent is used for lexical prediction, its presence should lead listeners to predict a "contrastive" target, while the lack of the accent should induce participants to predict a "given" target. Predictive processing of the emphatic initial accent should be visible at around 200-300 ms from the word-initial syllable. Effects before 200 ms would reflect the predictive use of pre-focal cues [5].

We found a bias for fixating images that were already referred to in the first question. Individual fixation proportions revealed two patterns of behavior (Figure 1). While a subgroup of participants did not use accentual information (the PROSODY- subgroup, n = 18), the other participants (the PROSODY+ subgroup, n = 15) showed an effect of prosody: they looked at a previously mentioned object more often when the initial syllable of the target word was deaccented than when it was accented. The effect of prosody was assessed within each subgroup and for each condition using cluster-based permutation analysis on two time windows: W1 (from -200 to 200 ms after word onset) and W2 (from 200 to 600 ms after word onset). While there was no effect on the PROSODY- subgroup, the statistical analysis confirmed an effect for the PROSODY+ subgroup. In the given condition, differences in fixations proportions between the accented and the deaccented condition started at 250 ms after word onset, indicating that listeners used the emphatic initial accent before they processed the complete sound pattern of the word (summedT = -49.6, $p = 0.003$). In the contrastive condition, the effect of prosody on fixations on the competitor occurred before word onset, suggesting a predictive use of the pre-focal cues (summedT = -47.01, $p < 0.001$). No differences were found in response latencies across the two subgroups. Our findings support the idea that predictive processing is an "optional" mechanism during language comprehension [6]. In line with [7], the use of predictive prosodic cues in French may be the result of an optimal choice by listeners, after balancing the reliability of the cues (the emphatic accent and the pre-focal cues) and top-down expectations (about the intonational form-meaning mapping), for instance [7].

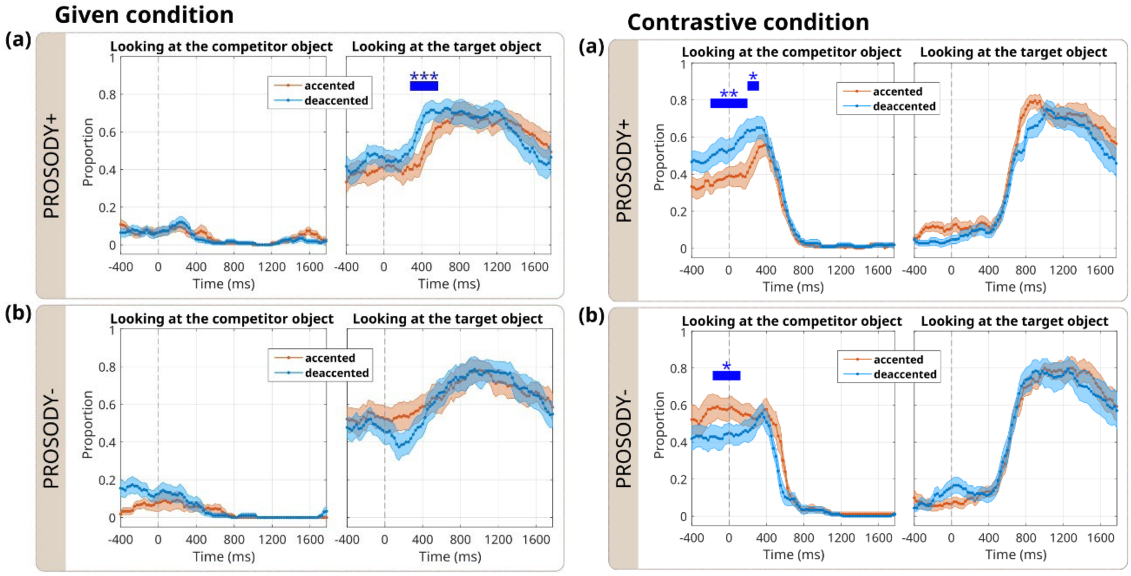


Figure 1. *Fixation proportions for the PROSODY+ (a) and the PROSODY- (b) subgroups in the given (left) and contrastive conditions (right). Within each graph, fixations proportions are shown for the competitor (left panel) and the target object (right panel), averaged in adjacent time bins of 25 ms across trials and participants and aligned with the onset of the target word. The averages (in bold) are surrounded by the standard error values (lighter color). Fixation proportions are split by the presence vs. absence of the emphatic initial accent on the initial syllable of the target word (red = accented; blue = deaccented).*

[1] Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314.

[2] Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573.

[3] Esteve-Gibert, N., Schafer, A.J., Hemforth, B. *et al.* (2020). Empathy influences how listeners interpret intonation and meaning when words are ambiguous. *Memory and Cognition* 48, 566–580.

[4] Jun, S.-A., & Fougeron, C. (2000). A Phonological Model of French Intonation. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 209–242). Springer.

[5] Turco, G., Dimroth, C., & Braun, B. (2013). Intonational Means to Mark Verum Focus in German and French. *Language and Speech*, *56*(4), 461–491.

[6] Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31.

[7] Roettger, T. B., & Franke, M. (2019). Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation. Cognitive Science, 43(7), Article e12745.

11. L2-L1 and L1-L2 influence in vowels of late Salento Italian-French bilinguals

Marie Fongaro[1], Barbara Gili Fivela[2], Maud Pélissier[3]
[1]University of South Bohemia in České Budějovice, [2]University of Salento, [3]University Paris Cité

L2-L1 and L1-L2 influence have been well studied at the phonetic level, but few authors have yet examined (1) the link between the two, addressing both the L1 and L2 of bilinguals [1,2], and (2) L2-L1 influence in other language combinations than those where English is L1 or L2 of bilinguals [3]. this study concerns both the L1 and L2 of bilinguals by focusing on an as yet unexplored L1-L2 combination, i.e., Salento Italian (L1) and French (L2) of Italians from Salento (South of Italy) who started to learn French mostly in adolescence, moved to the Paris region as adults and have lived there for varying length of time (hereafter late Salento Italian-French bilinguals - SIF). French and Salento Italian interestingly differ in their vowel systems: The distinction between open and close mid vowels is not found in Salento Italian, but it is in French (here, the variety of French spoken in the Paris region) [4,5], and the opening of /a/ is more marked in Salento Italian than in French [6,7]. This study therefore compares vowels produced by SIF in their L1 and L2 speech with those produced in L1 speech of two control groups, i.e. monolingual speakers from Salento and from the Paris region. In the hypothesis of the assimilation and/or dissimilation (SLM, SLM-r [8]) of vowels of SIF as related to length of residence in L2 country (LOR), we expect mainly the F1 of their L1 and/or L2 /a/, /e/, /ɛ/, /o/ and /ɔ/ to be affected, pointing to (1) a possible assimilation of L2 to L1 categories and the lack of L2-L1 influence in the case of short LOR, and (2) a possible dissimilation of L2 from L1 categories and the L2-L1 influence in case of long LOR.

We recorded the L1 and L2 of 15 SIF (9M, 6F; mean age = 41.13 y.o.; SD = 10.39, mean LOR = 14.4 years, SD = 9, min LOR = 1 year, max LOR = 33 years) and the L1 of 15 Salento Italian and 15 French controls, matched for age, sex and education level. To elicit speech, we created two sets of pictures evoking mostly two-syllable Italian and French words, with target vowels placed in stress-controlled position. In the first set, each Italian and French target vowel (/a/, /ɛ/, /e/, /i/, /ɔ/, /o/, /u/, /y/, /ø/, /œ/, /ə/) occurred in two words. In the second set, the French and Italian words were orthographically and phonologically similar (e.g., *certificat* in French and *certificato* in Italian), French and Italian /a/, /i/ and French /y/ occurred in two words each, French and Italian /e/, /ɛ/, /o/, /ɔ/, /u/ in four or five words, and the words formed 'easy pairs' or 'challenging pairs' depending on whether the expected target vowel in the similar words was the same in French and Italian or not (e.g., easy pair: /tɛte/ in French and /tɛsta/ in Italian, challenging pair: /pɔz/ in French and /poza/ in Italian). A target word X, shown on the PC screen, was produced by speakers in the carrier sentence 1) 'I say X', and 2) 'I put X next to Y. I moved X' to describe the way the experimenter was moving X on the screen. Automatic labelling and segmentation obtained using BAS Web Services [9] were corrected manually in PRAAT [10] where the vowel F1 and F2 were automatically measured via script (mean value from the middle third of the vowel). The values manually corrected were automatically normalised [11]. Results were obtained by performing in R a set of linear mixed-effects models (see Eq. 1), the comparison of estimated means was carried out with *emmeans* [12]. Significant results ($p < .05$; Fig. 1) reveal for SIF 1) Not French monolingual-like F1 in L2 production of /e/, /ɛ/ and /o/ (significant for 2nd set of pictures only), with an assimilation of F1 of French /e/ and /ɛ/ to their L1 /e/, French /o/ to their L1 /o/; 2) Not French monolingual-like F2 in L2 production of /e/ and /y/ (sig. for both sets), of /a/, /u/, /ə/ (sig. for 1st set only) and of /o/ (sig. for 2nd set only) with an assimilation of F2 of French /e/ to their L1 /e/, French /y/ to their L1 /u/, French /a/ to their L1 /a/, French /u/ to their L1 /u/, French /ə/ to their L1 /e/ and French /o/ to their L1 /o/; 3) Salento Italian monolingual-like L1 production of all vowels: Thus, no L2-

L1 influence was found; 4) No straightforward relationship between LOR and L1-L2 and L2-L1 influence.

(1) lmer(F1~(group*set*language) + (1+set|speaker) + (1|word), data)
lmer(F2~(group*set*language) + (1+set|speaker) + (1|word), data)

Equation 8. *Linear mixed-effects models used for the statistical analysis.*

Note: Fixed effect *language* and *set,* and random slopes *set* inserted into the model when relevant, *set*=set of pictures (1st or 2nd), 'data'=data set for a single target vowel.

(2)

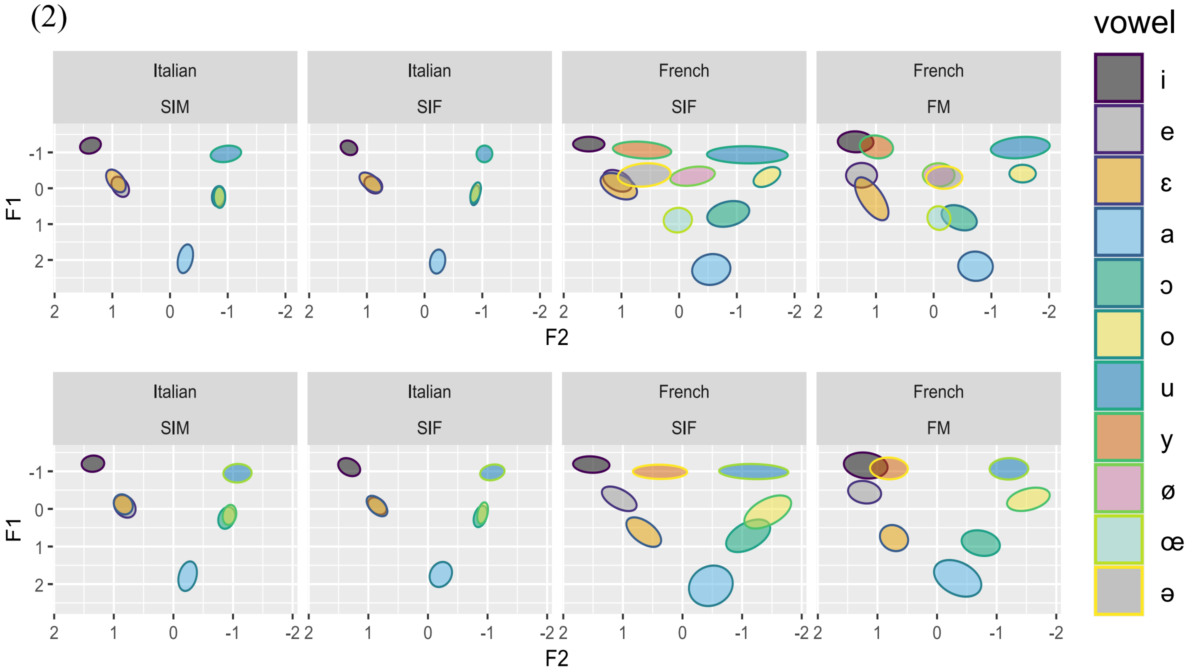


Figure 9. *Vowels of 15 SIF, SIM and FM plotted in normalised F1/F2 plane.*

Note: The ellipses indicate 50% of analysed data. SIM=Salento Italian monolinguals, FM=French monolinguals, upper plots=vowels from the 1st set of pictures, lower plots=vowels from 2nd set of pictures.

[1] Zampini, M. L. 2008. L2 speech production research: Findings, issues, and advances. In Hansen Edwards, J. G. & Zampini, M. L. (Eds.), *Phonology and Second Language Acquisition.* Amsterdam: John Benjamins B.V, 219-250.

[2] Kornder, L., Reubold, U., Mayr, R., & Mennen, I. 2023. The relationship between L2 acquisition and L1 attrition in the phonetic domain. *Proceedings of ICPhS 2023.* (Prague, Czech Rep.), 2780-2784.

[3] De Leeuw, E. 2019. Phonetic attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition.* Oxford, New York: Oxford University Press, 204–217.

[4] Grimaldi, M. & Calabrese, A. 2017. Metaphony in Southern Salento: New analysis and new data. In D'Alessandro, R. & Pescarini, D. (Eds.), *Advances in Italian dialectology: Sketches of Italo-Romance Grammars.* Leiden: Brill, 253-29.

[5] Munot, P., & Nève, F.-X. 2002. *Une introduction à la phonétique*. Liège: CEFAL.

[6] Escudero, P., Sisinni B. and Grimaldi M. 2014. The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels, *Journal of Acoustical Society of America*, 135 (3), 1577-1584
[7] Tubach, J. P. 1989. *La parole et son traitement automatique*. Paris, Milan, Barcelona: Masson.
[8] Flege, J. E., & Bohn, O.-S. 2021. The revised speech learning model (SLM-r). In Wayland, R. E. (Ed.), *Second Language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press, 3–83.
[9] Kisler, T., Reichel U. D. and Schiel, F. 2017. Multilingual processing of speech via web services, *Computer Speech & Language* 45, 326–347.
[10] Boersma, P., & Weenink, D. 2024. *Praat: doing phonetics by computer*. Computer program.
[11] Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49 (2B), 606–608.
[12] Lenth, R. V. 2021. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.4.

12. Production of English /p/ versus /b/ by Hindi-English and Portuguese- English bilinguals

Suzanne V.H. van der Feest[1], Yuga Kothari[1], Yasuaki Shinohara[2], Katarina Antolovic[1], Theresa Bloder[3], Aline Durney[1], and Valerie L. Shafer[1]
[1]The Graduate Center, City University of New York, [2]Waseda University, [3]Catholic University Eichstätt-Ingolstadt.

Recently more studies have recognized the value of examining the phonetic patterns of speech in multilingual individuals from less commonly studied language groups, with different ages of acquisition, use and proficiency levels of their various languages (e.g. Chen, Best & Antoniou, 2020). Assimilation of sounds in a later-acquired language ("L2") into categories formed in an earlier-acquired language ("L1") in perception and production is a major factor contributing to a speaker's 'accentedness' in their L2. Recent research indicates that how exactly assimilation unfolds depends on several factors, including phonological categories and acoustic-phonetic details (e.g. Flege & Bohn, 2021). However, much work reamins to understand how various factors contribute to L2 learning (and subsequent L3, etc.) categories.

The aim of this study is to examine productions of phonemes contrasting in laryngeal features in Hindi-English and Portuguese-English bilinguals. Hindi makes use of two laryngeal features: (1) Voice Onset Time (VOT, the [voice] feature), describing the onset of glottal pulsing, and (2) spread glottis, or After Closure Time (the [spread glottis] or [ACT] feature), which reflects the open versus closed status of the glottis following burst release (Mikuteit & Reese, 2007). Hindi [b] and [b^h] are realized with pre-voicing (lead or negative (VOT), whereas [p] and [p^h] are realized as voiceless (short-lag VOT). In addition, [b^h] and [p^h] are both realized with ACT (long-lag VOT after burst release). In contrast, English and Portuguese make use of only one laryngeal feature. English distinguishes between /b/ (as in 'bear', phonetically [p] - voiceless and without ACT) and /p/ (as in 'pear', phonetically [p^h] - voiceless with ACT). These phonetically match Hindi /p/ and /p^h/ respectively, albeit with quite substantial realization of aspiration in the English [p^h] as compared to Hindi (Antony et al. 2024). In contrast, Portuguese distinguishes /p/ and /b/ on the basis of voicing only, and these phonetically match Hindi [p] and [b]. (See Table 1 for a summary of featural representations under different accounts (e.g. Kager et al., 2007) in English, Portuguese and Hindi.) We hypothesized that early bilingual learners of English would adjust the phonetic realization to compromise between the systems. For example, Portuguese-English bilinguals would produce English /p/ voiceless and aspirated (with ACT), but might produce the English /b/ as (pre)voiced (and without ACT) because it is allophonic in English (Davidson, 2018). In contrast, *late* Portuguese-English bilinguals were predicted to produce English /p/ as voiceless unaspirated (without ACT), while both early and late Hindi-English bilinguals should map English /p/ to Hindi /p^h/ and English /b/ to Hindi /p/.

We analyzed English productions from 30 participants in the three groups (target n=16 per group). We recorded 10 /b/- and /p/-initial words in isolation and 5 of each sound in sentences (30 tokens per target per participant), read from a script. Preliminary findings measuring VOT using PRAAT reveal Portuguese-English bilinguals who learned English early (before age 10), produce English /p/ as voiceless and with ACT. Both late and early bilinguals, whether Portuguese or Hindi speakers, produced English /b/ as prevoiced. Unexpectedly, the Hindi-English bilinguals patterned with late Portuguese-English bilinguals, producing English /p/ as voiceless but without ACT (patterns illustrated in Figure 1). We discuss several different explanations, including the possibility that Hindi listeners focus on different cues for voicing categorization (e.g., F0) in perception and production (cf. Antony et al., 2024). Alternatively, we discuss whether sociolinguistic factors, e.g. related to English as an important language in India, or phoneme-letter correspondence, might underlie the Hindi-English bilinguals' different realization for the English laryngeal contrast.

Table 1: *Laryngeal contrasts and features (under a binary versus private featural account)*

	[pʰ]	[p]	[b]	[bʰ]
English (binary features) (privative features)	/p/ [-voice],[+ACT] [ø], [ACT]	/b/ [+voice],[+ACT] [voice], [ø]		
Portuguese (binary features) (privative features)		/p/ [-voice], [-ACT] [ø], [ø]	/b/ [+voice], [-ACT] [voice], [ø]	
Hindi (binary features) (privative features)	/pʰ/ [-voice],[+ACT] [ø], [ACT]	/p/ [-voice], [-ACT] [ø], [ø]	/b/ [+voice], [-ACT] [voice], [ø]	/bʰ/ [+voice], [+ACT] [voice], [ACT]

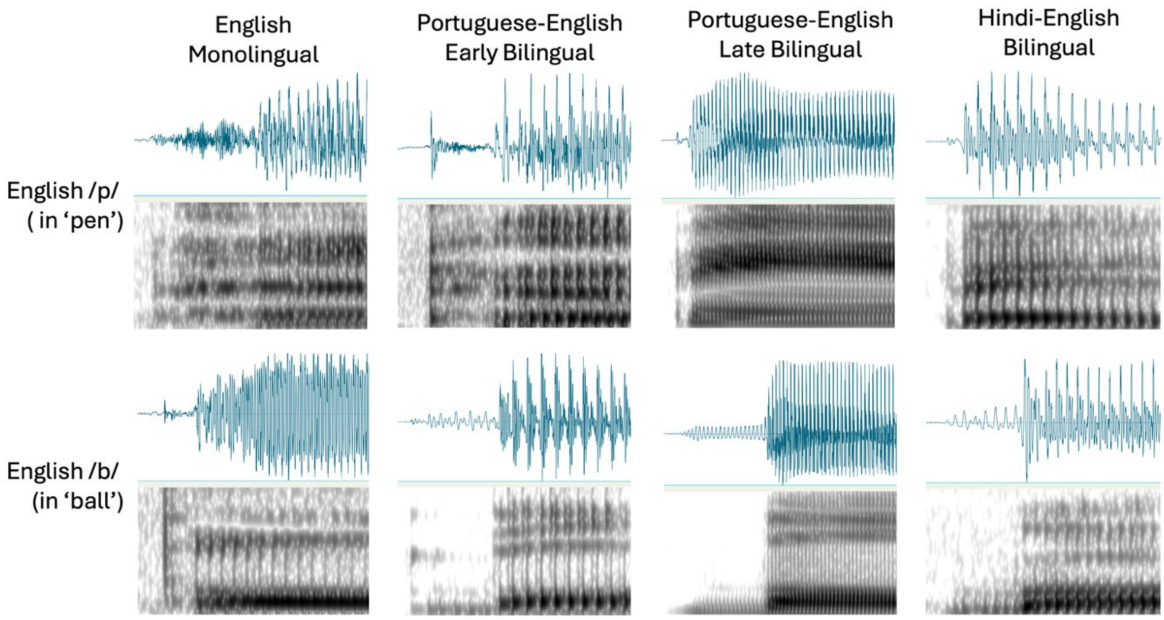


Figure 1: *Illustrations of English /p/ and /b/ productions*

Antony, R., Martin, B., Shafer, V.S., Behrens, S. (2024). Perception of Voicing and Aspiration in Hindi, American English, and Tamil Listeners in Quiet and in Background Noise. *Journal of Speech Language Hearing Research. 67*(7). 2367-2393.

Chen, J, Best, C.T, Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners, *Journal of Phonetics, 83*. doi:10.1016/j.wocn.2020.101013.

Davidson, L. (2018). Phonation and laryngeal specification in American English voiceless obstruents. *J of the International Phonetic Association*, *48*(3). 331-356. doi:10.1017/S0025100317000330

Flege, J.E., & Bohn, O.S. (2021). The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, *10*.

Kager, R., Van der Feest, S.V.H., Fikkert, P., Kerkhoff, A.O. & Zamuner, T.S. (2007). 'Representations of [voice]: evidence from acquisition.' In: J. van der Weijer & E.J. van der Torre (Eds.) *Voicing in Dutch*: *(De)voicing-Phonology, phonetics and psycholinguistics.* Current Issues in Linguistic Theory 286. John Benjamins. 41-80.

Mikuteit, S., & Reetz, H. (2007). Caught in the ACT: The Timing of Aspiration and Voicing in East Bengali. *Language and Speech, 50*(2), 247-277. doi:10.1177/00238309070500020401

13. Cross-linguistics differences in word stress production in varieties of North American Heritage Icelandic

Christiane Ulbrich[1], Nicole Dehé[2], Jörn Krantz[1], Meike Rommel[2]
[1]University of Cologne, [2]University of Konstanz

Our study investigates phonological phenomena in North American Heritage Icelandic (NAmIce) spoken in Manitoba. The study of Heritage Languages (HL) provides a valuable contribution to our understanding of how grammar is acquired and maintained. NAmIce is especially suitable for studying heritage language phonology because (i) Modern Icelandic (ModIce) is more conservative compared to its Scandinavian relatives, (ii) many speakers of moribund NAmIce were raised by 1st generation immigrants who actively preserved the language, (iii) Icelandic exhibit only minor dialectal variation, and (iv) there has not been a continuous influx since 1914, allowing researchers to trace moribund NAmIce back to varieties brought by immigrants from 1870 to 1914 (Arnbjörnsdóttir 2006, 2015; et al. 2018). The current study aims to explore cross-linguistic influence (CLI) in phonology, focusing on word-level primary stress. It is widely accepted that CLI is bi-directional and depends on several factors, which help to evaluate the entrenchment of a specific linguistic structure in the individuals' mental lexicon (Ulbrich & Ordin 2014; Ulbrich & Wiese 2018).

Word-level stress is typically acquired early. There is, however, disagreement among researchers regarding strategies for acquiring word stress, the age at which it becomes part of the phonological system, and how it interacts with other linguistic elements like word segmentation. One important shortcoming of studies on HLs phonology (e.g. Polinsky 2018) is that they are based on large group data with multilingual and multiethnic homeland situations and often do not include comparisons of the HL data with the two other languages in the heritage dyad (i.e., the dominant ambient language and the homeland language). We test the full heritage dyad for Icelandic (see speaker groups and respective numbers in table 1) in a picture naming task with a total of 2460 datapoints across all groups.

Unlike English, ModIce has fixed primary word stress on the initial syllable of a word, with very few exceptions (Árnason 2011). Our data set included English/Icelandic word pairs with same (e.g., Engl. *ˈhamburger*, Icel. *ˈhamborgari*) or different (e.g., Engl. *proˈfessor,* Icel. *ˈprófessor*) primary stress position (see table 2). Given that fixed word stress in Icelandic may not be phonologically represented, we hypothesized word stress in NAmIce (Group B) to be deviant from word stress in ModIce spoken in Iceland (group A) in both position and phonetic realization. We did not expect to find differences between moribund NAmIce speakers and 1st generation descendants of more recent immigrants (C) in either their heritage Icelandic or their dominant English productions. However, we expected differences between those two NAmIce groups and the 2nd generation descendants (D) due to the timing of the onset of acquisition of English: only the 2nd generation speakers of heritage Icelandic are confronted with competing stress systems simultaneously. This was also expected to lead to differences between second language learners of Icelandic (E) on the one hand, and Groups B and C on the other.

Statistical analysis of acoustically using PRAAT (Boersma & Weenink, 2022) and auditorily identified position of word stress, carried out in R (R core Team, 2022), shows differences between speakers of modern Icelandic and heritage speakers of Icelandic in NAmIce. Unexpectedly, we did not find the divide between 2nd generation speakers and the moribund and 1st generation speakers. Rather, 1st and 2nd generation speakers produced primary word stress more like modern Icelandic than moribund speakers, who differ most from the Icelandic controls and showed transfer effects from their dominant North American English. We also find significant differences between heritage speakers in 1st and 2nd generation compared to L2 learners of Icelandic. Considering English, there were no significant differences between the three NAmIce groups compared to the English controls.

Group	Explanation	No. of speakers
A	Speakers of Modern Icelandic A1 aged 70+; A2 aged 20-50	21 / 20
B	HL Speakers of moribund North American Icelandic aged 70+: sequential bilinguals (pre-puberty)	27
C	Speakers of NAmIce-gen1: 1st generation immigrants to North America: sequential bilinguals (post-puberty)	10
D	Speakers of NAmIce-gen2: 2nd generation immigrants born in North America to Icelandic parents, 2L1 Icelandic and English simultaneous bilinguals (pre-puberty)	26
E	Speakers of Icelandic as L2 with L1 NAmEngl, sequential bilinguals (post-puberty)	12
F	Speakers of NAmEngl F1 aged 70+; F2 aged 20-50	7 / 10

Table 1: *Test groups for full heritage dyad (heritage) Icelandic*

	Question	Target response Icelandic	Word stress in Icelandic	Target response English	Word stress in English	number of items
a.	 Hvað sérðu á myndinni? What do you see in the picture?	(Þetta er) **hamborgari**.	ˈσ σ ˌσ σ	(It's a) **hamburger**.	ˈσ σ σ	10
b.	Hvað sérðu á myndinni? What do you see in the picture?	(Þetta er) **banani.**	ˈσ σ ˌσ	(It's a) **banana**.	σ ˈσ σ	10

Table 2: *Word stress, testing primary stress; the full data set included 2460 items, tested for all 6 test groups.*

Árnason, K. 2011. *The phonology of Icelandic and Faroese*. Oxford: Oxford University Press.

Arnbjörnsdóttir, B. 2006. *North American Icelandic: The Life of a Language.* University of Manitoba Press.

Arnbjörnsdóttir, B. 2015. Reexamining Icelandic as a heritage language in North America. In: *Germanic heritage languages in North America: Acquisition, attrition and change*, ed. by Janne Bondi Johannessen & Joseph C. Salmons, 72–93. Amsterdam: John Benjamins.

Arnbjörnsdóttir, B., H. Þráinsson & Ú. Bragason (eds.). 2018. *Sígurtunga: Vesturíslenskt mál og mennig*. Reykjavík: Háskólaútgáfan.

Paul Boersma & David Weenink (1992–2022): Praat: doing phonetics by computer [Computer program].

Version 6.2.06, retrieved 23 January 2022 from https://www.praat.org.

Polinsky, M. 2018. *Heritage Languages and Their Speakers*. Cambridge University Press, Cambridge.

R Core Team, (2022). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. Thráinsson, H. 1994. Icelandic, in: König, E., van der Auwera, J. (Eds.), The Germanic Languages. Routledge, London & New York. 142–189.

Ulbrich, C. & M. Ordin. 2014. Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics* 45C: 26-42.

Ulbrich, C. & R. Wiese. 2018. Phonotactic principles and exposure in second language processing. In: C. Ulbrich, A. Werth & R. Wiese (eds.), *Empirical approaches to the phonological structure of words.* Linguistische Arbeiten 567. Berlin: De Gruyter, 153-182.

14. L2 and heritage Spanish listeners' perceptual acquisition of a new dialect-specific sound during a short-term study abroad in Sevilla

Brendan Regan
Texas Tech University

There is an increased interest in the acquisition of dialect-specific sounds by L2 and heritage Spanish speakers during study abroad (henceforth SA), particularly phonetic production (George 2014; Knouse 2012; Pozzi & Bayley 2020; Ringer-Hilfinger 2012; Schmidt 2020). Less is known regarding the perception of dialect-specific phones. A few studies have examined coda /s/ aspiration (Escalante 2018; Del Saz 2019; Schmidt 2018) and Andalusian post-consonant aspiration [t^h] for /s.t/ (Bedinghaus 2015), finding that SA listeners increase their ability to perceive/identify regional allophones over time. Most studies, however, have not included gradient auditory stimuli, heritage speakers, or proficiency level comparisons. This study aims to address these gaps by examining the perception of Andalusian post-consonant aspiration of word-internal /sp, st, sk/ (*pasta* as [ˈpa.t^ha]) (Ruch & Harrington 2014; Ruch & Peters 2016).

The aims were three-fold: (i) to examine the effect of a short-term SA on the perception of post-consonant aspiration (/s.t/ as [t^h]) among L2 and heritage Spanish listeners; (ii) how perception varies by proficiency level; (iii) how perception varies by bilingualism type. 57 U.S. undergraduate students (43 women, 14 men; ages 18-23) (25 intermediate [19 L2, 6 heritage], 32 advanced [24 L2, 8 heritage]) participated in a 5.5-week SA program in Sevilla. 17 *sevillano/as* (11 women, 6 men; ages 19-38) also participated. A realization of *pasta* (/s.t/), *tasca* (/s.k/), and *caspa* (/s.p/) produced by a 42-year-old *sevillana* was phonetically manipulated in Praat (Boersma & Weenink 2023) to create a 7-step VOT continuum (65, 55, 45, 35, 25, 15, 5ms), following Ruch and Harrington (2014). These stimuli were used to construct a binary forced-choice identification task consisting of 168 audios (3 places of articulation x 7-steps x 8 repetitions) in which listeners selected between *pata-pasta*, *capa-caspa*, and *taca-tasca* in PsychoPy3 (Peirce et al. 2019) at the beginning and end of SA. The independent variables consisted of VOT (7-steps), time (T1, T2), proficiency level (intermediate, advanced), and bilingualism type (L2, heritage) with the random effect of participant. The dependent variable was the binary selection of minimal pairs.

Mixed effects logistic regressions in R (R Team 2024) using the *lmer* function (Bates et al. 2015) found that *sevillano/as* were sensitive to changes in VOT for all places of articulation (Figure 1A), while L2 and heritage Spanish speakers only showed a change in perception over time with /s.t/ realizations in which they perceived less VOT as /s.t/ at a cross-over point nearly identical to *sevillano/as* (Figures 1B-D). There were no significant differences for proficiency level nor bilingualism type. The implications are two-fold: (1) as [t^h] is more phonetically salient than [p^h] and [k^h] for *sevillano/as* (Ruch & Peters 2016), L2 and heritage listeners show more sensitivity to this salience revealing its role in acquisition; (2) gradient stimuli allow L2 and heritage speakers to show gradient gains in the perception of dialect-specific phones during a short-term SA (increased perception of [t^h] as /s.t/). Future studies should examine if a longer SA shows more gains in for /sp, sk/ and differences for bilingualism type or proficiency levels. In conclusion, examining the perception of dialect-specific sounds during SA, provides insights into L2/D2 acquisition.

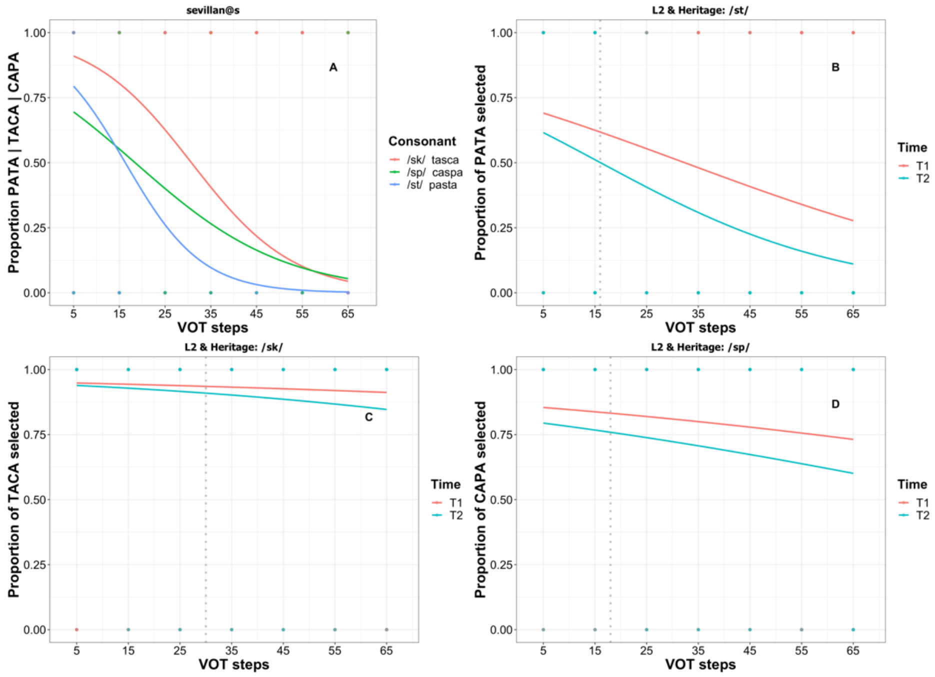


Figure 1. *All three places of articulation for 'sevillan@s' (A); Time 1 versus Time 2 for L2 and heritage listeners for /st/ (B), /sk/ (C), and /sp/ (D). Note: gray dotted line indicates cross-over point for 'sevillan@s', 1 = perceived 'pata/taca/capa', 0 = perceived 'pasta/tasca/caspa'. DISCLAIMER: 'Sevillan@s' were not included as a "monolingual control group", but rather to examine how speakers of this particular dialect perceptually utilize the VOT cues as Andalusian Spanish is the only variety that demonstrates this dialect-specific feature.*

[1] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

[2] Bedinghaus, R. W. (2015). The effect of exposure to phonological variation on perceptual categorization and lexical access in second language Spanish: The case of /s/-aspiration in Western Andalusian Spanish [Unpublished doctoral dissertation]. Indiana University.

[3] Boersma, P., & Weenink, D. (2023). *Praat*: A system for doing phonetics by computer.

[4] George, A. (2014). Study abroad in central Spain: The development of regional phonological features. *Foreign Language Annals*, 47, 97–114.

[5] Knouse, S. M. (2012). The acquisition of dialectal phonemes in a study abroad context: The case of the Castilian theta. *Foreign Language Annals*, 45, 512–542.

[6] Pozzi, R., & Bayley, R. (2020). The development of a regional phonological feature during a semester abroad in Argentina. *Studies in Second Language Acquisition*, 43, 109–132.

[7] R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. www.R-project.org.

[8] Ruch, H., & Harrington, J. (2014). Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. *Journal of Phonetics*, 45, 12-25.

[9] Ruch, H., & Peters, S. (2016). On the origin of post-aspirated stops: production and perception of /s/ + voiceless stop sequences in Andalusian Spanish. *Laboratory Phonology*, 7(1), 1-36.

[10] Ringer-Hilfinger, K. (2012). Learner acquisition of dialect variation in a study abroad context: The case of Spanish [θ]. *Foreign Language Annals*, 45, 430–446.

[11] Schmidt, L. B. (2020). Role of developing language attitudes in study abroad context on adoption of dialectal pronunciations. *Foreign Language Annals*, 53, 785–806.

15. Perception and production of English silent letters by L1 Spanish CLIL and EFL primary students

Evelyn Gandón Chapela[1], Marta Gómez Martínez[1], Esther Gómez Lacabex[2]
[1]Universidad de Cantabria, [2]Universidad del País Vasco

There is ample evidence that graphemic information is activated during speaking activity [1, 2]. Specifically, silent letters (e.g.: *listen*, *scene*) have been described as a frequent cause of mispronunciation in L2 speech [3, 4], also in the case of Spanish L2 learners [5, 6], whose L1 is highly transparent, exhibiting strong grapheme-to-phoneme associations. Interestingly, teaching programmes which bring along an increase of exposure to English such as Content and Language Integrated Learning (CLIL) could have an impact on the development of phonological acquisition. In fact, recent work with young L2 learners in bilingual programmes has suggested that they can successfully activate graphemic information from both L1 and L2 orthographic systems during reading [7]. The present study investigates the identification and production skills of English silent consonants by young Spanish schoolchildren in two differentiated English learning programmes. These differed in increase of exposure, namely a traditional EFL programme and a CLIL programme where learners received such an input in addition to approximately 200 hours of English medium instructed sessions of subjects such as Mathematics, Science or Social Science.

235 Spanish primary students aged between 8 and 10 performed a auditory decision task and a read aloud task of 12 English words containing 6 exocentric silent consonants (*comb*/*lamb*, *knife*/*knee*, *listen*/*fasten*, *walk*/*half*, *scissors/scent*). As for the auditory decision task, students chose whether the word they heard and saw was a correct pronunciation of an English word or not in a two-alternative forced-choice format. The words were both presented as correct pronunciations in English (*comb* /kəʊm/; *walk* /wɔ:k/; *knife* /naɪf/) and as incorrect pronunciations based on a transparent sound-grapheme congruence, representative of their L1, Spanish: (*comb* /kəʊmb/; *walk* /wɔ:lk/; *knife* /knaɪf/). In the read aloud task, learners were asked to pronounce the word orthographically presented on the screen. Accurate identifications of correct and incorrect identifications and pronunciations were calculated in CLIL and EFL settings.

As for the auditory task, average accuracy of acceptance of pronunciation of English silent letters was at 71%, while average accuracy of rejection of pronunciations of English silent letters actually pronounced (on account of Spanish grapheme-to-phoneme transparent correspondence) was at 37%, evincing that these young learners frequently accept mispronunciations of English silent letters in their interlanguage. As for production, results indicated that the average accuracy of correct pronunciations of the silent consonants was at 35.6%. We found no differences between CLIL and EFL groups in their overall identification and production performances, while few differences were found when silent letters were computed individually in favour of the CLIL group. Results also indicated differences in identification scores on account of the silent consonant inspected (e.g: a velar plosive in *knife/knee* better identified as incorrect than a bilabial plosive in *lamb/comb*). Production results revealed significant differences in some of the word pairs (e.g: *listen* > *fasten*; *scissors* > *scent*). Results will be discussed in the light of the moderate advantages of CLIL as for pronunciation and the impact of word frequency on production and perception.

[1] Saletta, M., Goffman, L., & Brentari, D. (2016). Reading skill and exposure to orthography influence speech production. *Applied psycholinguistics*, *37*(2), 411-434.
[2] Ranbom, L. J., & Connine, C. M. (2011). Silent letters are activated in spoken word recognition. *Language and Cognitive Processes*, *26*(2), 236-261.

[3] Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics*, *36*(1), 67-91.
[4] Mouquet, M., & Mairano, P. (2023). Effects of silent letters on the L2 English pronunciation of L1 French learners. In *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 188-198).
[5] Charpentier-Jiménez, W. (2022). EFL University student's production of English words containing silent letters. *Lenguas Modernas*, (60), 39-56.
[6] Mompean, J. A., & Fouz-González, J. (2016). Twitter-based EFL pronunciation instruction. *Language Learning and Technology* (20)1, 166-190.
[7] Hevia-Tuero, C., Incera, S., & Suárez-Coalla, P. (2022). Influences of first and second language phonology on Spanish children learning to read in English. *Frontiers in Psychology*, *13*, 803518. https://doi.org/10.3389/fpsyg.2022.803518

16. Effects of language background on sentence type identification in Brazilian Portuguese

Jaqueline Cerqueira Ionson[1], Meghan Armstrong[1], Luciana Lucente[2]
[1]University of Massachusetts Amherst, [2]Universidade Federal de Minas Gerais

To what extent are intonational question cues accessible to speakers of differing language profiles? Polar question (PQ) marking by intonation only is a less common cross-linguistically. [1] report that just 18% of the languages surveyed for the World Atlas of Language Structures Online use this distinction. Brazilian Portuguese is one such case. [2]'s inventory of tune types for BP annotated with the P_ToBI system show falls for both declaratives (H+L* L%) and PQs (L+H* L%). [3]'s recent study using the DaTo system also describes a late rising contour >LH with a L boundary tone for PQs in BP (see also [4]). This differs from American English (AE), for example, where unbiased PQs typically have a rise (a low rise or a high rise, see [5]), *do* support, and inversion. Despite the lack of a rise at the end of a question, some languages still show a high tonal target somewhere in the question. Puerto Rican Spanish (PRS) presents an extra high tone ¡H* in the nuclear syllable followed by a fall to a low boundary tone L%, though this may be truncated for oxytone words [6].

We hypothesized that AE speakers would tend to hear both BP fall types as declaratives, since we expect this group to expect rises for PQs. For PRS speakers, we asked to what extent having a falling tune for questions might provide an advantage, even though the PQ fall in PRS is quite different from the PQ fall in BP. We also asked to what extent having some instruction in Portuguese helped AE speakers. Below we describe an identification task with varying language profiles: 15 L1 American English (AE) speakers who had never studied BP, 15 AE speakers that had been taking Portuguese classes for a few weeks, 15 AE speakers that had studied Portuguese in the past, 12 Puerto Rican Spanish (PRS) speakers (no Portuguese), and 13 L1 BP speakers (total 70 participants). The task was designed to assess participants' abilities in discriminating declaratives from PQs. Targets utterances were identical segmental strings that differed only in intonation – (e.g. *A festa foi no clube* "The party was at the club" with either the declarative or PQ fall (see annotation above). Stimuli were recorded by two BP speakers from Southeastern Brazil (1 male, 1 female). Participants listened to an utterance and decided whether it was a declarative, PQ, exclamative or wh-question. Eight distractors (exclamatives and wh-questions) were included, with a total of 40 items. The experiment was administered online.

There was an overall effect of utterance type, such that all groups (except the BP group) performed significantly worse in identifying questions when compared to declaratives. However, even though all groups performed significantly better for declaratives, and likely because of the tendency to associate falls with declaratives more generally, none of the non-BP groups were at ceiling for that condition, though the AE speakers who had learned Portuguese at some point in the past were approaching ceiling. This group was the only group that did *not* perform significantly worse than BP speakers for declarative identification. All non-BP groups performed significantly worse than BP participants for the PQ condition, though AE speakers with past learning experience with Portuguese performed significantly better than those who had never studied it. Interestingly, the fact that PRS has falling polar questions did not seem to provide any advantage for this group, evidence the specific tonal targets are conventionally associated with linguistic meaning. Data collection is continuing for the new Portuguese learners, who will be tested again at the end of academic year. Our results show that speakers rely on their L1 associations between intonational forms and their meanings, in line with [7]'s "semantic dimension" in the LiLt model of intonational acquisition. While our participants are not all L2 learners, the results provide us with insight about what tendencies we might expect learners to come to the L2 classroom with, and suggests that explicit intonational instruction could be beneficial.

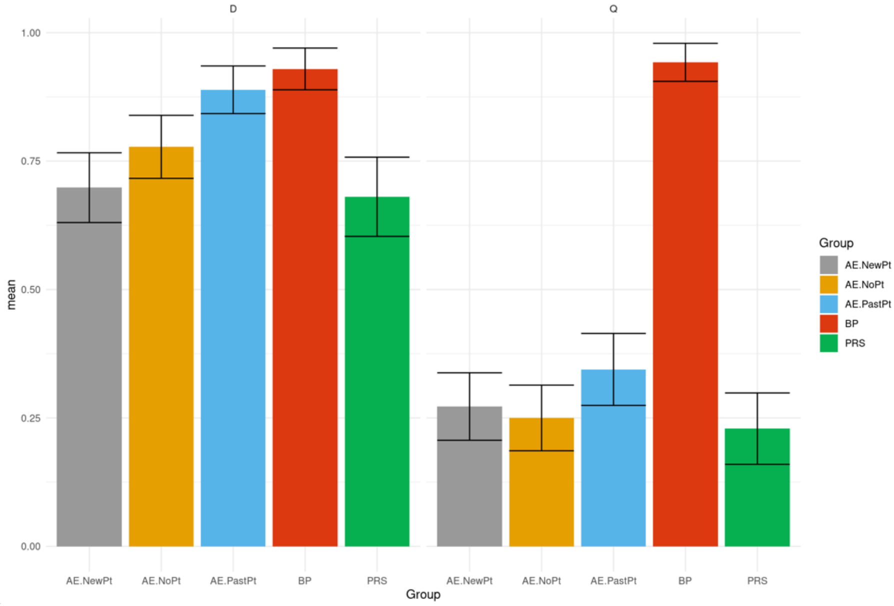


Figure 1. *Mean correct identification for Declaratives (D) and Polar Questions (Q), by Group.*

[1] Dryer, M. S., & Haspelmath, M. (2013). Polar questions. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. Retrieved from https://wals.info/chapter/116.

[2] Frota, S., Cruz, M., Svartman, F., Collischonn, G., Fonseca, A., Serra, C., Oliveira, P., & Vigário, M. (2015). Intonational variation in Portuguese: European and Brazilian varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 235–283).

[3] Kachkovskaia, T., Lucente, L., Smirnova Henrqiues, A., de Souza Fontes, M.A., Skrelin, P., & Madureira, S. (2024). L1-L2 influence in intonation: A case of Russophone immigrants in Brazil. *Languages, 9*(6), 212; https://doi.org/10.3390/languages9060212

[4] Lucente, Luciana. 2015. Funções comunicativas e restrições articulatórias na determinação do alinhamento de contornos dinâmicos. *Revista Diadorim* 17: 52–66

[5] Hirschberg, J. & Ward, G. (1995). The interpretation of the high-rise question contour in English. *Journal of Pragmatics, 24*(4), 407-412.

[6] Armstrong, M.E. (2015). Accounting for intonational form and function in Puerto Rican Spanish polar questions. *Probus, 29*(1), 1-40.

[7] Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations* (pp. 171–188). Berlin: Springer.

17. Through *dick* & *dünn*: Cognates and the legacy of diachronic change in synchronic processing

Joshua Booth, Felicia Stich, Isabella Fritz & Aditi Lahiri
University of Oxford

Linguists interested in the historical development of words have always differentiated 'cognates' and 'loans': English (E) *reptile* and German (G) *Reptil* are loans, as they were both independently borrowed from French, whilst E *pound*~G *Pfund* are cognates, as they were both inherited from the same West Germanic source (developing differently over time). However, most modern experimental research disregards this distinction (particularly in L2 literature). Instead, any words with form–meaning overlap across two languages are (mis)termed 'cognates': e.g. E *ambulance* ~ Welsh *ambiwlans* (or even E *chocolate* ~ Mandarin *qiǎokèlì!*). This definition furthermore excludes many true (inherited) cognates, e.g. E *starve* ~ G *sterben*.

German and English share many genuine cognates, but the question we ask is whether they *hinder* or *facilitate* L2 word recognition. How do systematic sound relations affect processing, and do the phonological feature specifications of cognates matter? Does the phonological representation of *Pfund* affect the access of *pound* for fluent L2 E speakers? We present the results of a recent EEG study, arguing that the distinction between cognates and (shared) loans is as relevant for work on synchronic language processing as historical linguistics.

Although modern speakers have no knowledge of their language's history, sound change is *systematic*, resulting in *regular correspondences* between related languages. This necessarily plays a role in the phonology (even where there is no segmental or semantic identity). Much L2 literature relies on algorithms of orthographic or phonemic similarity, but not all change is equal: *the direction matters*. Due to a phonological change known as the 'Second Sound Shift' and subsequent loss of /θ/ in G (see Table 1), certain classes of E–G cognates systematically differ in their **initial** consonant, which might be a stop, fricative or affricate.

In our experiment, native G speakers (with high E proficiency) completed a cross-modal lexical decision task in E, hearing a CV fragment prime (e.g. [θɔː]) before seeing a visual target. Their task was to identify whether or not the target was an E word. Reaction times and brain activity were recorded. We entertained the hypothesis that lexical access is not language-selective and fragments may thus *not only* activate the relevant E word, *but also* its G cognate (recall that word-initial sounds are critical in word recognition). Our predictions are given in Table 2. We expected near-identical G cognates to compete and thus *hinder* the activation of the E word (cf. [1]); furthermore, depending on whether the feature specification of the initial consonant of the corresponding cognate **matches**, ***mis*matches** or is tolerated (**no-mismatch**), lexical activation of the target might be inhibited or facilitated.

The results support our predictions. First, the priming effect was affected by condition, with 'identical' cognates having the **slowest** reaction time (and thus smallest priming effect): competing cognate forms were activated and hindered access. However, the features of the initial consonant were crucial. Although the identity condition was significantly slower than all others, the non-mismatching cognates (e.g. *thorn*) also had significantly slower reaction times than the non-existent condition (where there was no competition and access was facilitated). This was not true of the mismatching condition (which showed no activation of an L1 competitor, behaving as though no cognate existed). Results are summarised in Table 3.

Of particular interest was the 'N400' ERP component, which reflects ease of lexical access (more attenuated with greater facilitation). We expected fragments to activate and prime targets, but that interference from competing L1 cognates would lead to an increased reaction time and less-attenuated N400. **Most strikingly, the N400 was only significantly attenuated for the mismatching and non-existent conditions.** These words were therefore easier to access than the other conditions (where the competing L1 cognate *was* activated, leading to competition and inhibition).

	Proto-Germanic	English		German	
/p/ → /p͡f/	p	p	[pɑːθ] *path*	p͡f	[p͡faːt] *Pfad*
/t/ → /t͡s/	t	t	[tɪn] *tin*	t͡s	[t͡sɪn] *Zinn*
/θ/ → /d/	θ	θ	[θɪk] thick	d	[dɪk] *dick*
—	f	f	[fɪʃ] *fish*	f	[fɪʃ] *Fisch*
—	m	m	[maʊ̯s] *mouse*	m	[maʊ̯s] *Maus*

Table 1: *Summary of the effects of the Second Sound Shift and subsequent loss of /θ/ on initial consonants in German*

	English	Extracted features	German cognate feature representation	Feature matching	Activation effect
(i) identical	**m**ild	[nasal]	[nasal]	match	Activated: competition (+)
(ii) tolerated	**th**ick	[continuant]	underspecified [plosive]	no-mismatch	Activated: competition
(iii) mismatching	**t**in	[plosive]	[strident]	mismatch	Not activated: no competition
(iv) non-existent	**p**ig	[plosive]	—	—	Not activated: no competition

Table 2: *Predicted activation effect of the G cognate, depending on whether the initial consonant's feature specification is a match, mismatch or no-mismatch for those accessed from the E; [PLOSIVE] is underspecified and thus not stored in the lexical representation (but is present in the signal, cf. [2]).*

	English	German	Priming Effect	N400
(i) identical	fish, mild	Fisch, mild	21.3ms	—
(ii) tolerated	thorn, thick	Dorn, dick	42.3ms	—
(iii) mismatching	path, tin	Pfad, Zinn	50.7ms	attenuated
(iv) non-existent	pig, thug	N/A	57.1ms	attenuated

Table 3: *Summary of experimental results*

[1] Fritz, I., Lahiri, A. & Kotzor, S. 2023. Shared Loanword Recognition in German–English Bilinguals: The Role of Metrical Phonology. *Journal of Cognitive Neuroscience* 35(12), 2028–2048. doi: https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1162/jocn_a_02063

[2] Lahiri, A. & Reetz, H. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38: 44–59.

Poster session 2a

1. Pupil dilation patterns in response to accented speech during simultaneous interpreting

Karolina Broś, Katarzyna Czarnocka, Janina Mołczanow, Małgorzata Szupica-Pyrzanowska
University of Warsaw

Simultaneous interpreting is a highly complex process whose success rests on many factors, including speech rate, information density, syntactic complexity, and sound quality ([1]). One of the largely underresearched factors potentially affecting this process is foreign accent which is known to hamper speech comprehension ([2]). However, there is scarce evidence of the impact of accented speech on cognitive effort during interpreting. The present study aims to fill this void by investigating the effect of accented speech on the interpreters' cognitive effort.

We tested 23 interpreting trainees and 27 professional interpreters. The procedure consisted in simultaneous interpretation of five short speeches delivered in English by different speakers. This resulted in a mixed, 5x5x2 design with 5 trials of 5 accents interpreted by two independent participant groups. The speeches were simple and did not contain specialized vocabulary to ensure comparability. Each speaker had a different accent in English, including one native (American English), treated as a baseline condition and four foreign accents: Polish, Italian, Swedish, Mandarin that differed phonetically and typologically from one another. Overall, the stimuli exhibited comparable non-native features, as confirmed by native speaker ratings of foreign accentedness, which were consistent across accents. We used an EyeLink Portable Duo eye-tracker (1000 Hz sampling frequency) to measure pupil size. The data were extracted and analysed in R. We checked language proficiency and information on the age and years of experience on the interpreting market of each participant. We expected pupil size to vary according to accent, reflecting the differences in cognitive and perceptual processing demands imposed by the properties of accented speech. Based on our expectations concerning allophonic feature transfer from L1, we predicted that cognitive effort would be lower for Polish-accented speech, possibly similar to the baseline American accent. The more the accent diverges phonetically from the baseline and the native accent of the participants (i.e. Polish), the higher the expected cognitive effort, potentially peaking with the typologically distant Chinese accent.

The results of the study suggest that trainees were significantly affected by different accents, while professional interpreters exhibited similar reactions to the various accented speeches demonstrating lower pupil activation specifically for the baseline American English accent (Fig.1). Though overall both professionals and trainees show similar responses to different accents, statistical significance was not reached in professionals beyond the first 60 seconds of each trial (cf. Fig.2). In the case of trainees, the cognitive effort associated with accented speech processing is greater throughout the trial. While the Polish accent elicited a moderate response in both professionals and trainees, suggesting it poses the least challenge for native Polish interpreters, only trainees seem to be sensitive to the type of accented speech applied, with a particular indication of Italian and, paradoxically, American accents as the most difficult. This may be due to the differences in exposure and familiarity with American English between the groups, as student training is mainly based on British pronunciation. As for the Chinese and Swedish accents, an initial increase in pupil size was followed by a steep decline in many participants, mostly in the trainee group. It is possible that motivation played a role, i.e. listeners may have found the acoustic challenge too demanding to engage additional resources [3], hence the decrease in pupil size to or below baseline shortly after the beginning of the trial. Furthermore, there were differences in pupil size depending on whether English is a primary or secondary working language for a given interpreter but no proficiency or age results were found. All in all, our findings suggest that accented speech affects speech processing in interpreters in different ways depending on experience and active language use.

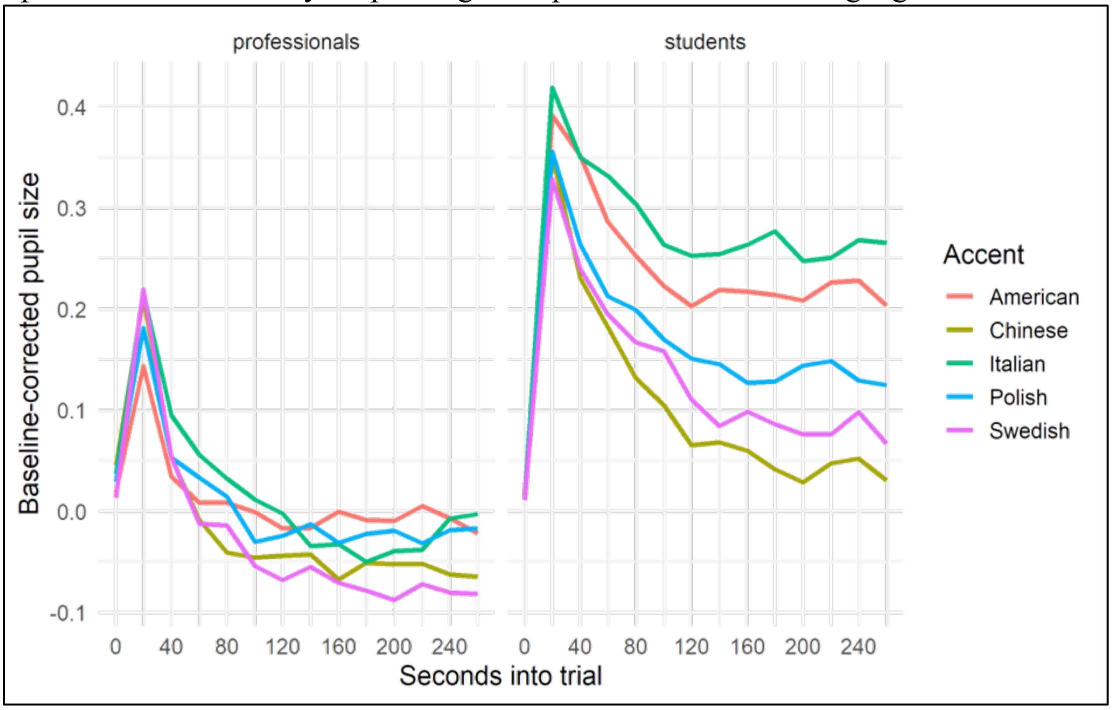


Figure 1. *Time course of pupil size variations: A comparison across groups and accents.*

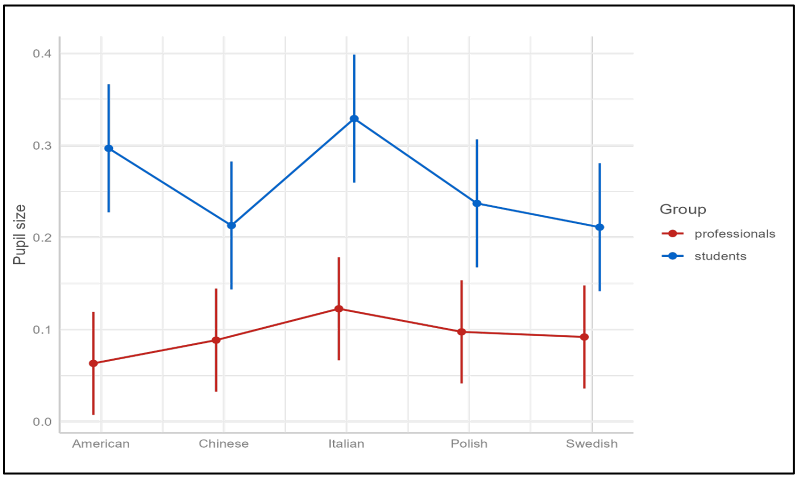


Figure 2. *Effects plot of the interaction between group and accent based on the GCA model fit corresponding to the first 60 seconds of each trial.*

[1] Gerver, D. 1976. Empirical studies of simultaneous interpretation: A review and a model. In R. W. Briskin (Ed.), *Translation: Applications and Research*. New York: Gardner Press, 165-207.

[2] van Engen, K. J. & Peelle, J. E. 2014. Listening effort and accented speech. *Frontiers in Human Neuroscience* 8, 577.

[3] Eckert, M. A., Teubner-Rhodes, S. & Vaden, K. I. 2016. Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. *Ear and Hearing* 37(Suppl 1), 101S-110S.

2. Prosody meets phonation: a dialectal investigation in Peninsular Spanish

Susan Cox
Florida State University

The purpose of the current study is to investigate dialectal differences in intonation and non-modal phonation in Northern and Southern dialects of Peninsular Spanish to better understand how they manifest and interact within different prosodic contexts. This study focuses specifically on utterance-final vowels in two prosodic contexts (Garellek & Keating, 2015; González et al., 2022; Gordon & Ladefoged, 2001). Based on previous studies, dialectal differences are expected, with breathy voice anticipated in the South (Mendoza et al., 1996; O'Neill, 2005; Trittin & De Santos Y Lleó, 1995), and creaky voice in the North (Armstrong et al., 2015; González et al., 2022). Non-modal phonation is expected to be more prevalent in final intonational phrases than in intermediate phrases (Garellek & Keating, 2015; González et al., 2022; Gordon & Ladefoged, 2001; Morrison & Escudero, 2007). It is anticipated that breathier phonation will be found in females, while males will have more creak (Armstrong et al., 2015; Garellek & Keating, 2015; González et al., 2022; Mendoza et al., 1996; Morrison & Escudero, 2007; Trittin & De Santos Y Lleó, 1995; Yuasa, 2010).

The term *voice quality* refers to phonation, or how your voice sounds, resulting from the vibration of the vocal folds in the larynx along with sustained airflow from the breath. Modal voice quality is considered the neutral setting of the larynx, or the default way we produce sound. During modal voice, our breath causes the vocal folds to vibrate regularly, creating a smooth sound. In contrast, creaky voice and breathy voice are two types of non-modal phonation that diverge from modal voice to create different qualities of sound and can be realized by manipulating the degree of constriction in the glottis (Esling et al., 2019). Creaky voice occurs when the vocal folds vibrate slowly and with more constriction, causing the voice to lower in pitch and become creaky. Breathy voice occurs when air escapes through the less constricted vocal folds while producing sound, giving it an airy quality like a whisper, only louder (Esling et al., 2019).

Data for the current study were collected from 29 monolingual native speakers of Peninsular Spanish from Northern and Southern dialects using a reading task and an elicitation task. Data from the elicitation task is the focus of the current study. Participants saw images of twelve target items either at the end of an intonational phrase (IP), or mid-phrase at the end of an intermediate phrase (ip). For example, in the intonational phrase (IP) context, *Veo una **boca***]$_{IP}$ ('I see a mouth'); and in the intermediate phrase (ip) context, *Veo una **boca***]$_{IP}$ *y una nariz* ('I see a mouth and a nose'). The non-target words that followed the target word in the intermediate phrase context were semantically related to the target. Participants performed two repetitions, totaling 1080 tokens.

Data is being analyzed acoustically in Praat following Spanish ToBi (Aguilar et al., 2009; Boersma & Weenik, 2022). Preliminary analyses of 9 participants indicate variation in intonation and phonation. For both dialects, in the (IP) context, final boundaries vary between low (L%) and high (H%). In the Southern dialect, devoicing, breathiness, and creak occur frequently at final boundaries. While there is less breathiness in the Northern dialect, devoicing and creak are present. For both dialects, in the (ip) context, a high boundary tone (H-) occurs frequently followed by a long pause with little non-modal phonation. Ongoing analyses will refine these findings and results will be shared during the conference presentation.

Aguilar, L., De-la-Mota, C., & Prieto, P. 2009. *Sp_ToBI Training Materials* [Computer software]. http://prosodia.upf.edu/sp_tobi/

Armstrong, M. E., Henriksen, N. C., & DiCanio, C. 2015. Sociophonetic analysis of young Peninsular Spanish women's voice quality. In R. Klassen, J. M. Liceras, & E. Valenzuela (Eds.), *Issues in Hispanic and Lusophone Linguistics* John Benjamins Publishing Company: Vol. 4, pp. 293–312

Boersma, P., & Weenik, D. 2022. *Praat: Doing phonetics by computer* (6.2.22) [Computer software]. praat.org

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*(4), 423–444.

Esling, J. H., Moisik, S. R., Benner, A., & Crevier-Buchman, L. 2019. *Voice quality: The laryngeal articulator model*. Cambridge University Press.

Garellek, M., & Keating, P. 2015. *Phrase-final creak: Articulation, acoustics, and distribution*. The 89th Meeting of the Linguistic Society of America, Portland, Oregon.

González, C., Weissglass, C., & Bates, D. 2022. Creaky Voice and Prosodic Boundaries in Spanish: An Acoustic Study. *Studies in Hispanic and Lusophone Linguistics*, *15*(1), 33–65.

Gordon, M., & Ladefoged, P. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383–406.

Mendoza, E., Valencia, N., Muñoz, J., & Trujillo, H. 1996. Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of Voice*, *10*(1), 59–66.

O'Neill, P. 2005. Utterance final /s/ in Andalusian Spanish. The phonetic neutralization of a phonological contrast. Language Design (7), 151–166.

Trittin, P. J., & De Santos Y Lleó, A. 1995. Voice quality analysis of male and female Spanish speakers. *Speech Communication*, *16*(4), 359–368.

3. Can L2 learners be trained to distinguish L2 and L1 categories?

Celia Gorba[1], Angélica Carlet[2], Juli Cebrian[1], Núria Gavaldà[3], Wolf De Witte[1], Susana Cortés[4]

[1]Universitat Autònoma de Barcelona, [2]Charles Darwin University, [3]Universitat de Barcelona, [4]Universitat de les Illes Balears

High Variability Phonetic Training (HVPT) has been shown to enhance second language (L2) perception and production by exposing learners to highly variable input, involving stimuli from various talkers and different contexts [1, 2]. According to the most influential models of L2 speech, L2 learners need to distinguish between native and non-native sounds in order to establish target-like L2 categories. The ability to discern differences between L1 and L2 sounds may decrease with age but it may improve with L2 exposure and experience [3]. To date, few studies have examined the effect of HVPT on the ability to distinguish L2 from L1 sounds, and these have reported little or no effect of perceptual training on cross-linguistic similarity of vowels [4] (cf. [5] on consonants). Further, HVPT studies typically involve perception and/or production tasks that focus on target language sounds and sound contrasts. To our knowledge, no prior training studies have included L1-L2 sound contrasts in their training regimes. Finally, while feedback is a crucial component of phonetic training, few studies have investigated whether the presence and the type of feedback provided (i.e., from more to less explicit feedback) has an effect on training outcomes [2].

To address these issues, the current study aims to compare two HVPT regimes that differ in the type of stimuli used and the type of feedback provided. Thus, a training task that involves distinguishing between L2 sounds is contrasted with a training task that contains both L1 and L2 stimuli. Regarding feedback, two types of feedback are implemented differing in whether or not the identity of the stimuli is revealed after each trial. The research questions thus address (a) whether training to distinguish L1 and L2 vowels improves L2 perception and production, and (b) how specific HVPT feedback needs to be.

The participants are Spanish/Catalan bilinguals who are L2 English learners, mostly undergraduate students in an English Studies degree. Half the participants are trained with L2 stimuli only and the other half with both L1 and L2 stimuli. Each group is further subdivided on the basis of feedback type: one subgroup receives feedback directing attention to the correct answer, while the other receives right/wrong feedback without stimulus identification. L2 target vowel and L1-L2 contrasts were selected based on previous findings (e.g., [6], [7]). Training is implemented by means of an identification task (involving either L2 stimuli only or L1-L2 stimuli). The training stimuli consist of non-sense words that are phonotactically possible in both the L1 and the L2 (e.g., /fVpi/, /nVfi/, /sVfi/). Before and after training, participants complete a perceptual assimilation task, and L2 vowel identification and production tasks involving non-word and real word stimuli to assess generalization of learning to untrained stimuli. Identification and production tasks include Standard Southern British English vowels present in the training regimes: /iː/, /ɪ/, /æ/, /ʌ/.

The study is currently in progress and the results will be discussed in light of the greater efficacy predicted by current L2 speech models for a novel training method focusing directly on the ability to discern L2 from L1 sounds, as well as the applied and theoretical implications of this approach. The methodological challenges posed by a mixed-languages training regime as well as the implementation of the different feedback types and their consequences will also be considered. These issues are crucial for understanding HVPT's efficacy and limitations, and for guiding future research.

[1] Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. The Journal of the Acoustical Society of America, 89(2), 874-886. doi:10.1121/1.400497.
[2] Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, 4(2), 208-231.
[3] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research (pp. 233-277). Timonium, MD: York Press.
[4] Cebrian, J., Carlet, A., Gorba, A., & Gavaldà, N. (2019). Perceptual training affects L2 perception but not cross-linguistic similarity. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 929-933). Australasian Speech Science and Technology Association Inc.
[5] Gong, J. García-Lecumberri, M. L., & Cooke, M. (2017). Ab initio perceptual learning of foreign language sounds: Spanish consonant acquisition by Chinese learners. *System* 66, 142-155.
[6] Rallo Fabra, L. (2005). Predicting ease of acquisition of L2 speech sounds. A perceived dissimilarity test. *Vigo International Journal of Applied Linguistics*, (2), 75-92.
[7] Cebrian, J. (2021) Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *Journal of the Acoustical Society of America*, 149(4), 2671-2685.

4. Mouse-tracking reveals L2 listeners' online processing and changes of mind during an identification task

Yizhou Wang[1], Rikke Bundgaard-Nielsen[12]
[1]The University of Melbourne, [2]MARCS institute

We report on a study investigating whether L2 English learners (L1 Mandarin, *n* = 20) can identify English affricate onsets, /tɹ/-/tʃ/ and /dɹ/-/dʒ/ in a two-alternative forced-choice (2AFC) task using mouse-tracking (MT) technology. MT trajectories are particularly informative in psycholinguistic research, as they provide both outcome-based measures (e.g., error rates) and easily accessible measures of fine-grained online processing and decision-making because goal-directed behaviour results from the tightly coupled relationship between vision, motion, and language cognition [1], [2]. This is the case even when the decision-making process involves complex steps, e.g., the participant may change their mind multiple times. In contrast, eye-tracking paradigms often employed in psycholinguistic studies, are less transparent regarding the underlying cognitive processes involved, and it is difficult to differentiate exploratory, verification, and decisive gazes.

The stimuli in the present study comprised a set of disyllabic pseudo-words (/CV.ti/), beginning with either /tɹ/ or /tʃ/ and /dɹ/ or /dʒ/. These consonants were followed by one of two nuclear vowels, /i/ and /u/, as research suggests that vowel context sometimes affects L2 affricate perception [3]. The second syllable provided a controlled context. All stimuli were produced by a male Australian English speaker. During the experiment, the participants heard a target pseudoword over headphones and then indicated which onset they heard by mouse-clicking one of two phonological labels, /tɹ/ versus /dɹ/ or /tʃ/ verus /dʒ/, presented in the top left and top right corners of their computer screen. For all trials, cursor position was returned to a central point at the bottom of the screen, and the trajectory of each cursor movement was recorded within that trial, together with the temporal information, i.e., cursor location at any given point in time. Unlike previous research that has relied on aggregated measures [4], we categorised all responses into five MT trajectory types: a straight line, a curved line, a continuous change of mind (cCoM), a discrete change of mind (dCoM), and discrete double changes of mind (dCoM2), using a recently developed classification system [5], [6], see Figure 1. We predict **(1)** that different types of MT trajectories have different spatiotemporal qualities, and **(2)** that the vowel context could affect L2 processing of onset consonants.

In the analysis, we measured and compared total trajectory distance and the response time (RT) for the five trajectory types. Consistent with **(1)**, the comparisons showed that straight < curved < cCoM < dCoM < dCoM2, in terms of distance covered by the cursor movement, and that {straight = curved} < cCoM < dCoM < dCoM2, in terms of movement duration (RT). Consistent with **(2)**, an analysis of the distributions of each trajectory type in two vowel contexts indicated that L2 listeners' affricate perception was indeed affected by the nucleus vowel context: Mandarin-speaking learners experience more perceptual confusion when the onset affricates are followed by /u/ than /i/. This effect was consistent between the accuracy data, RT data, and the distribution of trajectory types. The exact mechanism for this contextual effect requires further research, but it is possible that such an effect roots in the phonotactic differences between L1 and L2, and language-specific distributional properties, which may affect how affricate-vowel sequences are perceptually adapted in the listener's interlanguage system. For instance, Mandarin phonology permits affricate+/w/ but not affricate+/ɹ/ sequences, meaning that when Mandarin listeners perceive English /tɹ, dɹ/, they may rely less on the lingual features than the labial gestures of the rhotic segment, which become masked in the /u/ context (but not in the /i/ context) due to anticipatory coarticulation.

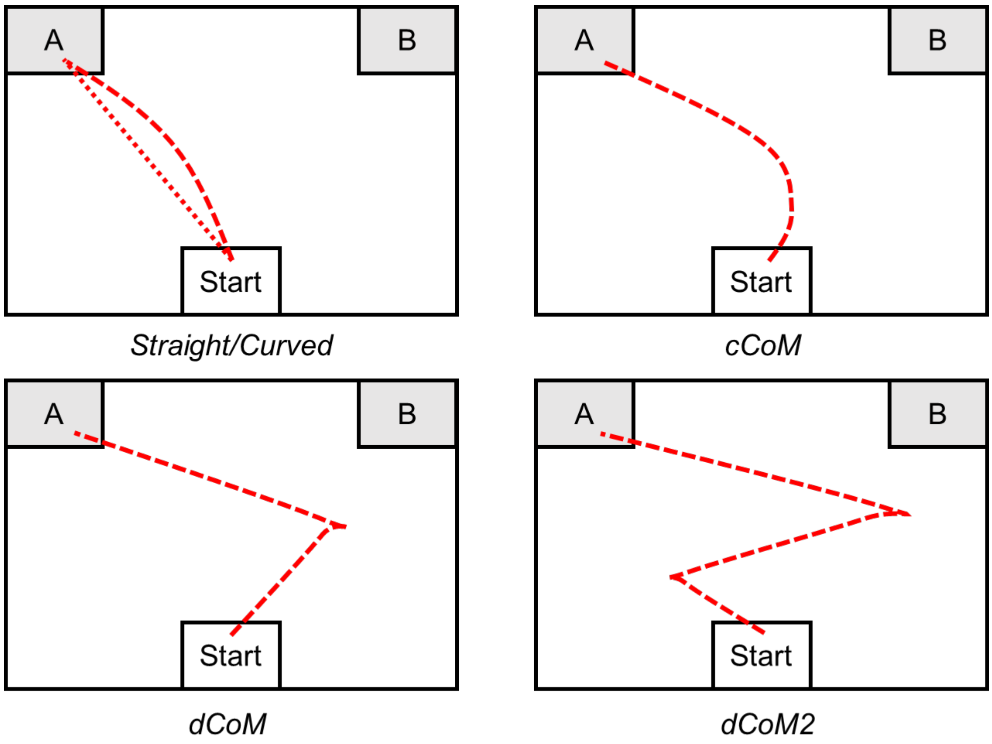


Figure 1. *Mouse-tracking trajectories with or without changes of mind.*

[1] M. J. Spivey, M. Grosjean, and G. Knoblich, “Continuous attraction toward phonological competitors,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 29, pp. 10393–10398, 2005, doi: 10.1073/pnas.0503903102.

[2] M. J. Spivey, D. Richardson, and R. Dale, “The movement of eye and hand as a window into language and cognition,” in *Oxford handbook of human action*, E. Morsella, J. A. Bargh, and P. M. Gollwitzer, Eds. Oxford University Press, 2009, pp. 225–249.

[3] Y. Lan, “Vowel effects on L2 perception of English consonants by advanced learners of English,” in *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, M. Le Nguyen, M. C. Luong, and S. Song, Eds. Hanoi, Vietnam: University of Science, Vietnam National University, 2020, pp. 149–157.

[4] Y. Wang, R. L. Bundgaard-Nielsen, B. J. Baker, and O. Maxwell, “Native phonotactic interference in L2 vowel processing: Mouse-tracking reveals cognitive conflicts during identification,” in *Proceedings of INTERSPEECH 2022*, H. Ko and J. H. L. Hansen, Eds. International Speech Communication Association, 2022, pp. 5223–5227. doi: 10.21437/Interspeech.2022-12.

[5] D. U. Wulff, P. J. Kieslich, F. Henninger, J. Hasbeck, and M. Schulte-Mecklenbeck, “Movement tracking of cognitive processes: A tutorial using mousetrap,” *PsyArXiv*, 2023, doi: 10.31234/osf.io/v685r.

[6] P. J. Kieslich, “Package ‘mousetrap.’” CRAN, 2021. [Online]. Available: https://github.com/pascalkieslich/mousetrap

5. Dynamics of the vowel sequence /ia/ in naturalistic Portuguese and Spanish

Johanna Cronenberg, Ioana Chitoran
Université Paris Cité

This study is concerned with the acoustics of the vowel sequence /ia/ in European Portuguese (PT) and Spanish (SP). While /ia/ is always phonologically classified as a hiatus in PT, it can be either a diphthong /ja/ or a hiatus in SP [1]. However, these canonical forms can be subject to adaptations in the production process due to the influence of phonetic factors. In small samples of laboratory speech, it has been found that SP prefers diphthongs, esp. with increased distance from the stressed syllable, e.g. *diálogo* /di'alogo/ (engl. dialogue) vs. *dialogó* /djalo'go/ (engl. s/he conversed) [2]. For PT, it has been claimed that diphthongization might exceptionally occur in informal speech or in post-tonic position [1]. If SP and PT differ in the phonological status of /ia/, we hypothesize that this difference should also be maintained in spontaneous productions despite the reduction processes associated with continuous, naturalistic speech and lexical stress. In line with previous research [3, 4], we expect hiatuses to have longer, steeper, and curvier formants than diphthongs.

The data for this study came from corpora of European radio and TV shows in PT (114 hours) and SP (223 hours; [5]) which included automatic segmentations on the word and phoneme levels [6]. For the available 8,107 tokens of /ia/ for PT and 31,192 tokens for SP, we measured F1 and F2 using the *forest* algorithm [7] and then z-scored and time-normalized the formants. The duration of the sequences was normalized for speech rate differences, and then time-warped into a horizontal, time-normalized signal with the rate-normalized duration as the y-intercept. The three signals per token were submitted to Functional PCA [8] in order to identify systematic patterns of variation in the formants' shapes and their duration. PC1, PC2, PC3, and PC5 were found to contribute information that was meaningful with respect to the hiatus-diphthong contrast. The corresponding PC scores were used as dependent variables in LMERs, with language, gender (male/female), and proximity-to-stress (tonic/pre- tonic/post-tonic) as fixed factors as well as a random intercept for word and a random slope for gender by word. Post-hoc comparisons were computed using estimated marginal means.

Figure 1 shows the result of reconstructing the formants and their duration using the estimated marginal means of the four PC scores, separately for the languages, genders, and proximity-to-stress conditions. The vowel sequence /ia/ is consistently shorter in SP than in PT. For both languages, the tonic sequences are longer compared to the pre- and post-tonic ones. However, the effect of stress on the duration of /ia/ is more pronounced in PT than in SP. In addition to being shorter, pre- and post-tonic sequences in both languages present a reduced movement in F1, especially in PT. The dynamics of F2 for pre- and post-tonic SP /ia/ as well as pre-tonic PT /ia/ are similar to those of their tonic counterparts. The post-tonic sequences in PT, however, additionally show a reduction of the F2 movement.

In summary, tonic /ia/ appears to be the most hiatus-like for both languages because of its long formants with steep transitions, while the same sequence in pre- or post-tonic position is shorter and shows significant reductions of the formant dynamics. That is, unstressed /ia/ aligns more with the phonological characteristics of a diphthong than a hiatus. In addition, the duration of SP /ia/, irrespective of lexical stress, is closer to the duration of diphthongs than that of hiatuses as described by [3] (133 vs. 181 ms). Thus, and in line with [9], SP shows a preference for diphthongal sequences. PT, on the other hand, maintains the hiatus-like acoustic configuration of /ia/ in tonic position, but is prone to diphthongization and even monophthongal tendencies in pre- and post-tonic position. Beyond contributing to our understanding of the phonetic gradience of the phonological diphthong-hiatus contrast in Romance languages, this study also highlights the benefits and challenges of working with large corpora of naturalistic speech.

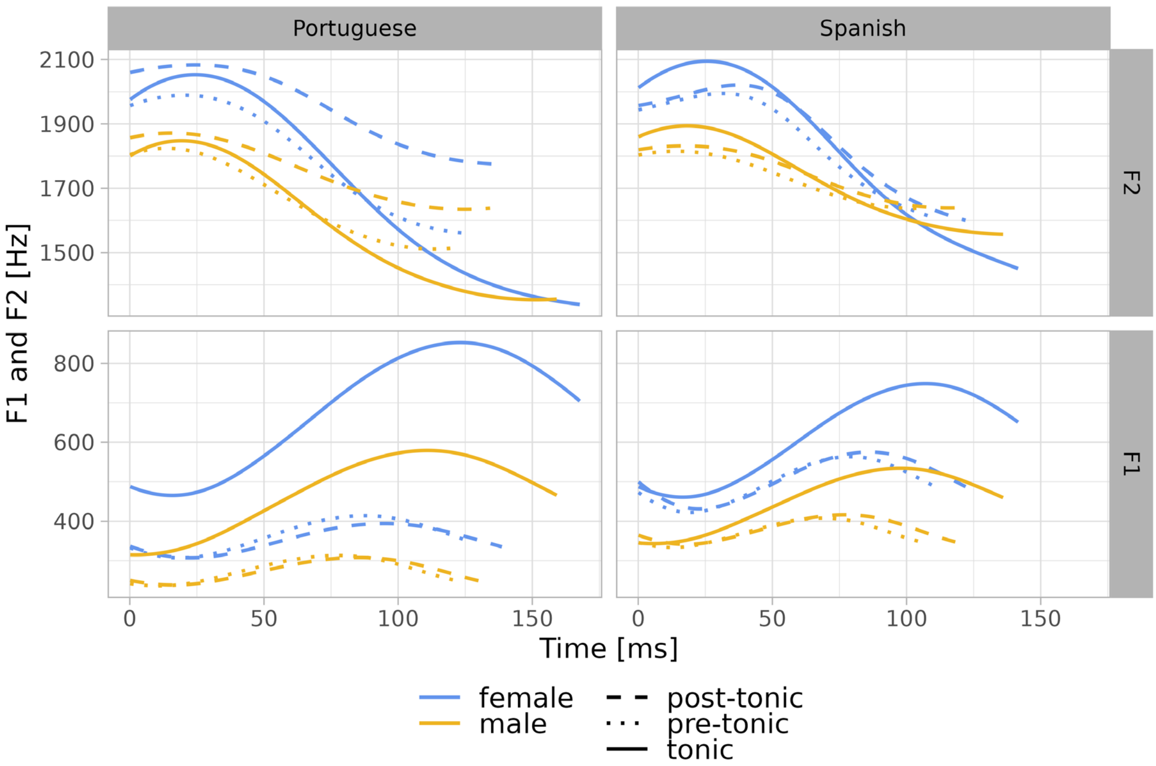


Figure 1. *F1 and F2, reconstructed using estimated marginal means of four PC scores, separately for Portuguese and Spanish, female and male speakers, as well as the three proximity-to-stress conditions.*

[1] Chitoran, J. & Hualde, J. (2007): From hiatus to diphthong: the evolution of vowel sequences in Romance. *Phonology* 24, 37-75.

[2] Cabré, T. & Prieto, P. (2007): Exceptional hiatuses in Spanish. In Colina, S. & Martínez-Gil, F: (Eds.), *Optimality-Theoretic Advances in Spanish Phonology*. Amsterdam: John Benjamins, 205-238.

[3] Gubian, M., Torreira, F. & Boves, L. (2015): Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* 49, 16-40.

[4] Aguilar, L. (1999): Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech Communication* 28, 57-74.

[5] Vasilescu, I., Yaru, W., Jatteau, A., Adda-Decker, M. & Lamel, L. (2020): Alternances de voisement et processus de lénition et de fortition: une étude automatisée de grands corpus en cinq langues romanes. *Traitement Automatique des Langues* 61, 11-36.

[6] Lamel, L. & Gauvain, J.-L. (1992): Continuous Speech Recognition at LIMSI. *Proceedings of the Final Review of the DARPA ANNT Speech Program*, 1-7.

[7] Winkelmann, R., Bombien, L., Scheffers, M. & Jochim, M. (2024): *wrassp: Interface to the 'ASSP' Library*. R package version 1.0.5.

[8] Ramsay, J. & Silverman, B. (2010): *Functional Data Analysis*. New York: Springer.

[9] Herrero de Haro, A. & Alcoholado Feltstrom, A. (2024): Anti-hiatus tendencies in Spanish: rate of occurrence and phonetic identification. *Linguistics* 62, 203-228.

6. Turn-taking in online conversations with people who do and do not stutter

Lotte Eijk, Sophie Meekings
Department of Psychology, University of York

People most commonly use speech in interactions with other people, where taking turns seems to flow effortlessly. Interlocutors time their utterances based on predictions about the other speaker's timings [1, 2]. Gaps between speakers' turns are only 300ms or less in many languages [3]. We have a lot of information on turn-taking in speakers without speech disorders (e.g., [4, 5, 6]). However, there is not as much focus on turn-taking in conversations including people with atypical speech, like people who stutter (PWS). PWS frequently experience involuntary syllable repetitions, prolongations, and 'blocks' where they cannot produce a sound. These factors make the speech of a PWS less predictable, which could influence turn-taking. A previous study [7] has found that PWS are more often interrupted or have their utterances completed in a conversation than a speaker who does not stutter. We aim to explore this further by investigating turn-taking in conversations with PWS in more detail. In this study, we examine whether there are differences in turn-taking speed, whether PWS get a similar amount of speaking time as typical speakers, and whether PWS are more likely to be interrupted than typical speakers or differ in the amount of backchanneling.

We analysed twenty conversations - half of the conversations were between two typical speakers (age: $M = 29.7$, $SD = 10.5$; gender: 6 F-F, 3 F-M, 1 M-M), and the other half consisted of typical-PWS pairs (age: $M = 32.8$, $SD = 12.2$; gender: 2 F-F, 7 F-M, 1 M-M). PWS were self-identified people who stutter.

Speakers participated in a Diapix spot-the-differences task [8] conducted over Zoom. Each pair participated in two rounds with two different pictures. Each picture contained 12 differences to be found in 10 minutes. In each round, one participant was appointed the leader who started the description, while the other participant was the follower. The participant who stutters always took the role of leader in the first round, after which roles were switched.

Backchannels and interruptions were automatically coded. Fully overlapping turns were considered backchannels, and turns that start before, but finish after the end of the previous speaker's turn were categorised as interruptions. We performed three mixed effects models with random effects for speaker and transcriber. The first model examined gap duration based on turn transition type (PWS to typical, typical to PWS, or typical to typical). The second model assessed turn duration based on speaker group (PWS, typical interacting with PWS, or with typical) and role (leader or follower), with an interaction between the two. The third model predicted the number of backchannels versus interruptions by speaker group and role.

Preliminary results indicated that leaders had longer turns, and role influenced the type of overlapping speech, such that the follower used more backchannels than the leader. No significant differences were found between the speaker groups in gap duration across the different turn transitions, nor in turn duration or type of overlap. These findings may suggest assigning clear roles in conversation could be helpful to PWS. The final results will be presented at the conference. Future research could benefit from manual coding of backchannels and interruptions, and could focus on exploring the relationship between stuttering severity and turn-taking behaviours.

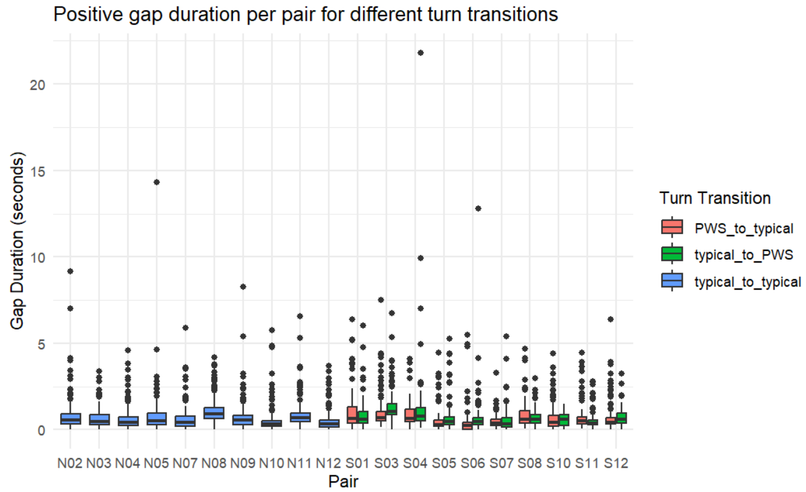


Figure 1. *Positive gap duration per pair (S = PWS-typical pair, N = typical-typical pair).*

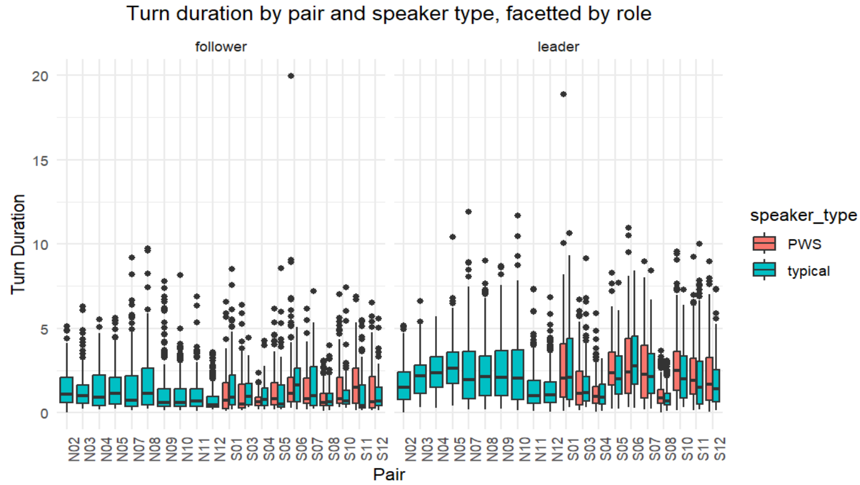


Figure 2. *Turn duration by pair (S = PWS-typical pair, N = typical-typical pair), and speaker type (PWS and typical), facetted by speaker role (leader vs follower).*

[1] Roelofs, A., & Ferreira, V. S. (2019). The architecture of speaking. Human language: From genes and brains to behavior, 35-50.

[2] Meyer, A. S. (2023). Timing in conversation. *Journal of Cognition, 6*(1).

[3] Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J.P., Yoon, K-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587-10592.

[4] Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics, 38*(4), 555-568.

[5] Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences, 119*(4), e2116915119.

[6] Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no… How the brain interprets the pregnant pause in conversation. *PloS one, 10*(12), e014547.

[7] Freud, D., Moria, L., Ezrati-Vinacour, R., & Amir, O. (2016). Turn-taking behaviors during interaction with adults-who-stutter. *Journal of Developmental and Physical Disabilities, 28*, 509-522.

[8] Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods, 43*, 761-770.

7. Acoustical correlates of femininity and masculinity in Spanish voices

Nuria Polo[1,4], Filipa M. B. Lã[2, 4], Paula Barriendo-Cebrián[1, 4], Mario Casado-Mancebo[3, 4]
[1]Faculty of Philology, UNED, [2]Faculty of Education, UNED, [3]Faculty of Philology, UCM, [4]Laboratory of Voice, Music and Language, UNED

Morphological, histological, acoustical and aerodynamical sex-related voice differences have been described in several investigations. However, the understanding of how these differences can be connected to the gender of a voice is still scarce. About 25 % of Spanish females believe that their voices do not sound sufficiently feminine [1]. Also, to sound feminine one needs to sound young. On the contrary, to sound masculine one must not sound young.

The question now is whether there are acoustical vocal features that correlate with these believes. The present investigation seeks to define the acoustical metrics that correlate with cross-gender laryngeal and vocal tract adjustments performed by Spanish speakers to sound feminine or masculine.

A total of 22 female and 20 male healthy young Spanish native speakers were recorded while reading a phrase containing all Spanish vowels: "he cumplimentado el formulario amarillo" (*I had filled in the yellow form*). All participants were instructed to read this phrase in five conditions: (1) habitual voice; (2) sounding feminine; (3) sounding masculine; (4) sounding feminine, while visualising a picture of Betty Boop (feminine stereotype); and (5) sounding masculine, while visualising a picture of Brutus, from Popeye cartoon series (masculine stereotype). All recordings were made in a sound treated booth using the free FonaDyn software (Version 2.4 by Sten Ternström, Sweden). The audio signal was collected with a headset omnidirectional microphone (DPA microphone, Denmark), attached to an external sound card (Fireface UCX, RME Audio, Germany), and calibrated using a sound level calibrator (Extech Instruments, Germany). Mean fundamental frequency (f_o) and vowels' formant frequencies and speech fluency-related metrics were extracted using a Praat script.

Preliminary results suggest that traits of femininity are related to elevated mean f_o and extreme values for formant frequencies. In the stereotyped condition of femininity, lower speech rate were found in both male and female voices. Masculine voices were created by decreasing f_o and increasing speech rate. These characteristics were exacerbated for the masculine stereotype condition. In future investigations, it seems worthwhile to investigate the extent to which these characteristics may correlate with self-perceptions of gendered personality traits.

[1] Polo, N. & González-Sanz, M. 2025. Creencias en torno a la manifestación del género en las voces españolas, *ELUA: Estudios de Lingüística de la Universidad de Alicante*, 43.

8. Intonational dynamics of German backchannels across conversational contexts

Eduardo Möking, Simon Wehrle, Constantijn Kaland & Martine Grice
IfL-Phonetik, University of Cologne, Germany

This study investigates how conversational context influences the intonation of backchannels (BCs) such as 'ja', 'mmhm' and 'mm' in task-oriented dialogue and spontaneous speech. We analysed speech data from a multimodal corpus involving 14 pairs of participants (who had not previously met), interacting in three communicative contexts sequentially: introductory small talk, a collaborative Tangram game, and a discussion of the Tangram game just played, introducing a fixed topic and a degree of common ground. The conversational contexts will henceforth be referred to Intro, Tangram and Discussion, respectively; more details can be found in [1]. We employ two different techniques for analysing pitch— global f0 direction and contour clustering—each of which captures distinct aspects of BC intonation. Global f0 direction is calculated as the semitone difference between the beginning and the end of a token, enabling a broad categorization of simple rises, falls, or level contours. While this relatively straightforward method has successfully been applied in previous work (e.g. [2], [3]), it does not account for the shape of pitch movements, or for token duration. To capture the shape of pitch movements, we hence adopted Kaland's [4] contour clustering method (v. 2024-08), which we used to group similar productions based on their f0 contour and duration. F0 was tracked as a time series using the default settings and 10 measurement points (f0 fit set to 0.6). The time series were speaker-standardized (z-scored). Hierarchical agglomerative clustering was performed with complete linkage as well as Euclidean distances between the time-series f0 and duration measures.

The two analysis techniques together reveal context-specific effects on BC intonation. In terms of global f0 direction, there was a greater amount of intonational rises in task-oriented speech (Tangram) than in either of the two spontaneous contexts (Intro and Discussion), suggesting a shift towards more prominent BCs in contexts that necessitate efficient, goal-directed feedback; see Fig. 1. Considering different lexical realisations of BCs, it can be seen that although all three BC types are more often rising in the Tangram condition, the baseline for each BC is different: 'mmhm' is predominantly rising throughout, reflecting its inherent role as a continuer, whereas 'ja' is predominantly falling or level, and 'mm' has a more balanced distribution of contours. While all BC types showed an increase in rising contours in task-oriented interactions, contour clustering allowed us to identify distinct shapes within these globally rising contours. Focusing on the BC types with the highest overall proportion of rises ('mm' and 'mmhm'), a predominance of late rising contours, including a fall-rise, was found for 'mm', while a preference for early rises was found for 'mmhm'. There was a general increase of late rises in task-oriented speech, which was more subtle for 'mmhm'; see Fig. 2.

Our results show that task-oriented contexts, which demand clear communication and attentiveness, evoke more dynamic and rising BC intonation compared to more spontaneous conversation, regardless of whether it involved small talk or a more targeted discussion (cf. [1], [5]). Our findings have broader implications for understanding the role of BCs in interactive speech, where the intonational qualities of these minimal responses are crucial for facilitating mutual understanding and conversational flow. Furthermore, we were able to identify both broad and more nuanced intonational features by using complementary techniques, thereby advancing previous approaches. The analysis of very short utterances, exemplified by BCs, can be seen as a test case for the challenges and possibilities inherent in the measurement and evaluation of f0 contours. Our findings also underline the importance of dyad-specific behaviour, reflected in the finding that while most dyads followed the group-level pattern, some deviated notably, emphasizing the role of individual variability in conversational behaviour.

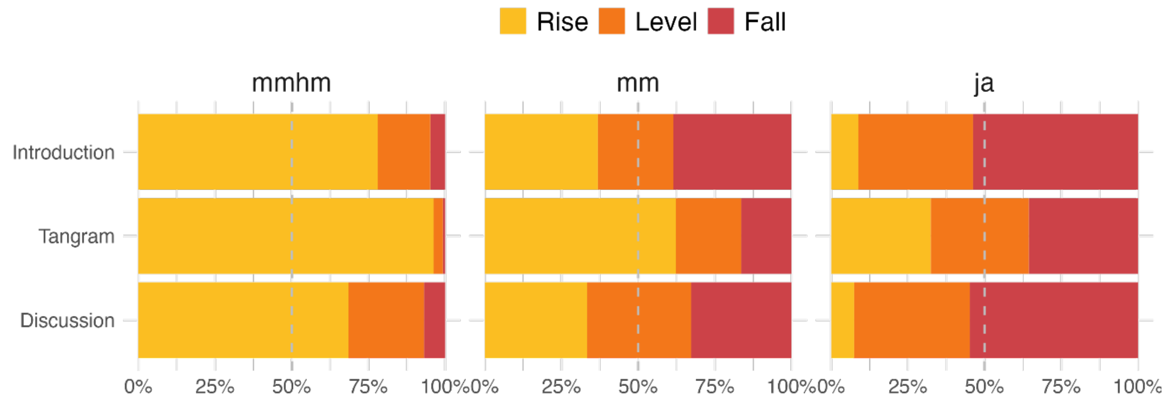


Figure 1: *Contour categories by BC type across conversational contexts. Rises (in yellow) are more frequent in the Tangram condition than in the other two spontaneous conditions (Introduction and Discussion). Note also the predominance of rises in 'mmhm' and falls (and level contours) in 'ja'.*

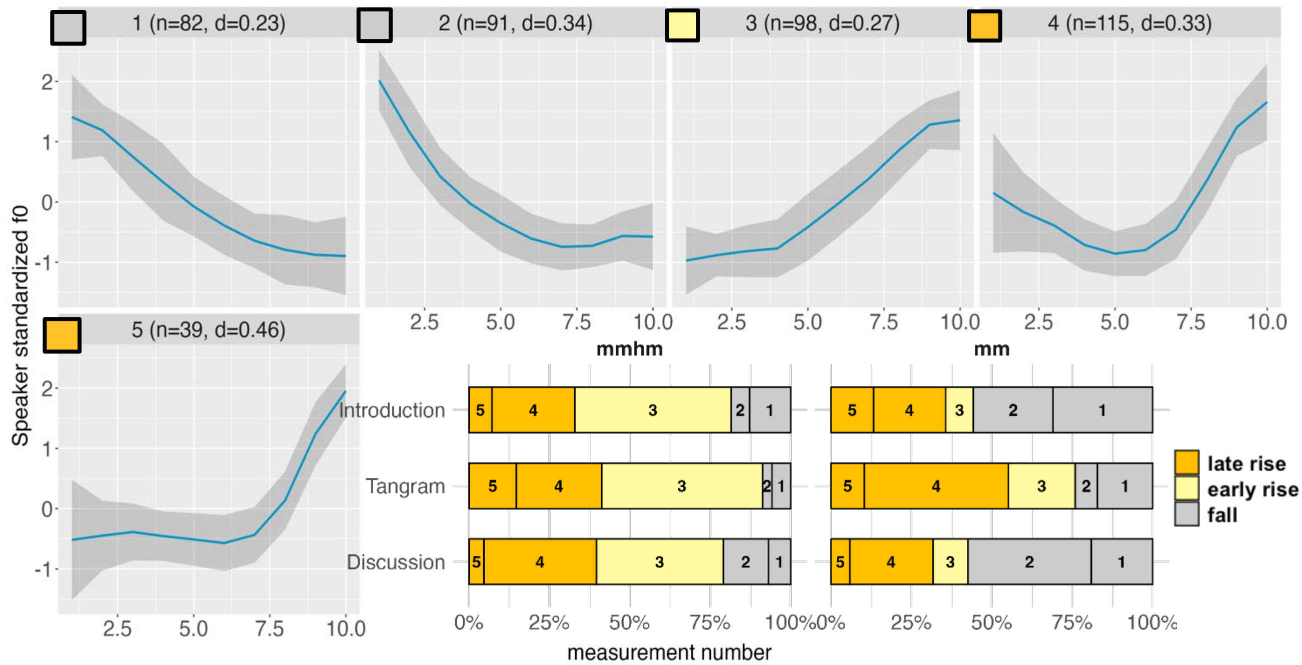


Figure 2: *Speaker-standardized f0 contours for 'mmhm' and 'mm' in five clusters. Measurement number (10 per BC) represents a time-normalized scale; n = number of observations in cluster, d = average duration in seconds.*

[1] Spaniol, M., Wehrle, S., Janz, A., Vogeley, K., & Grice, M. 2024. The influence of conversational context on lexical and prosodic aspects of backchannels and gaze behaviour. Proc. SpeechProsody 2024, 607–11. doi:10.21437/SpeechProsody.2024-123.

[2] Sbranna, S., Möking, E., Wehrle, S., & Grice, M. 2022. Backchannelling across Languages: Rate, Lexical Choice and Intonation in L1 Italian, L1 German and L2 German. Proc. Speech Prosody 2022, 734–38.

[3] Wehrle, S., Vogeley, K., & Grice, M. 2023. Backchannels in conversations between autistic adults are less frequent and less diverse prosodically and lexically. Language and Cognition, 1–26. doi:10.1017/langcog.2023.21.

[4] Kaland, C. 2023. Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. Journal of the International Phonetic Association, 53(1), 159–88. doi:10.1017/S0025100321000049.
[5] Janz, A. 2022. Navigating Common Ground Using Feedback in Conversation: A Phonetic Analysis. MA thesis, University of Cologne, Germany; https://kups.ub.uni-koeln.de/72407/.

9. Pitch patterns among female speakers in three dialects of English: Newcastle, London and Dublin

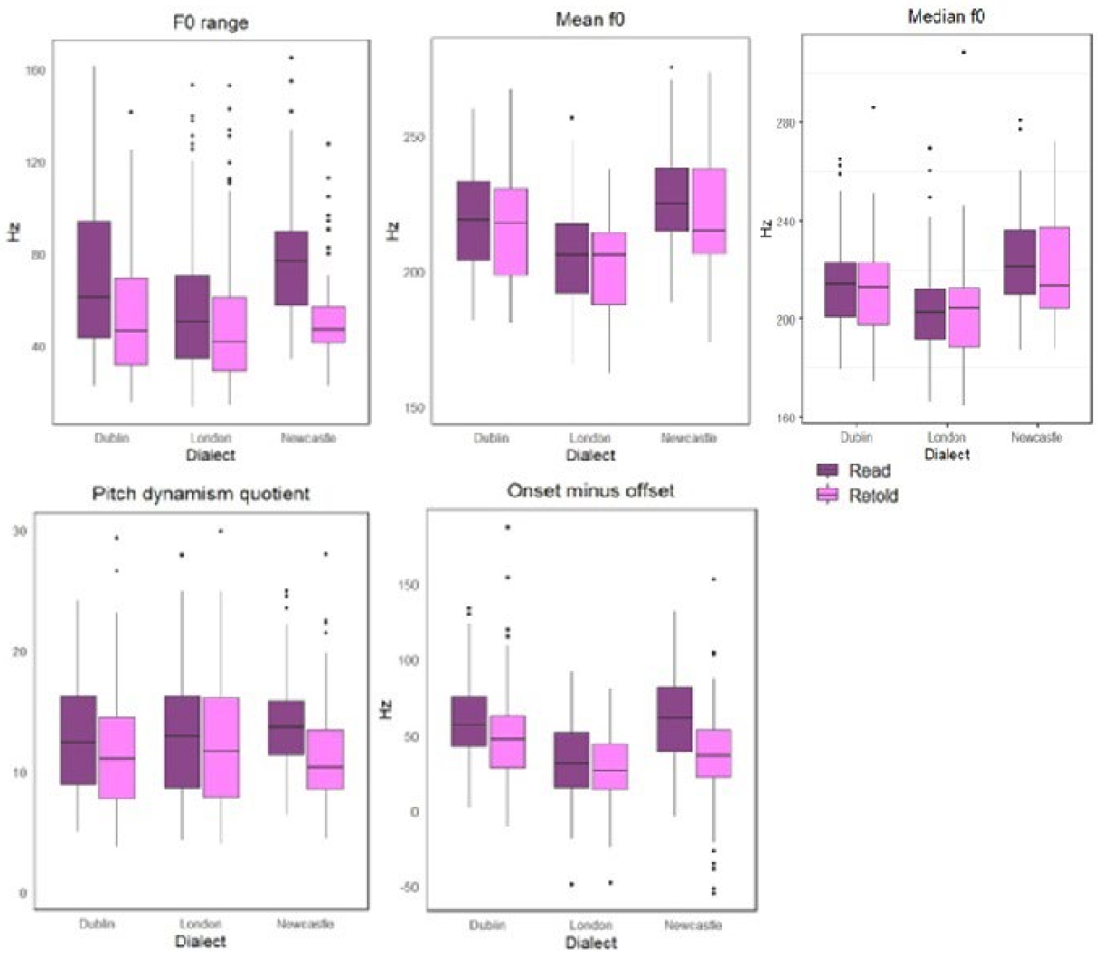
Niamh Kelly
Newcastle University

Pitch (acoustically, fundamental frequency or f0) range and pitch dynamism are suprasegmental factors which have been shown to differ across dialects (Meer & Fuchs 2022) and languages (Andreeva et al. 2014). There are also interactions with gender and language background. Generally, female speakers tend to have higher pitch than male speakers due to having shorter vocal folds (e.g., Hollien 1960) which vibrate at a higher frequency. Importantly, though, there is a large amount of variability within these groups, and sociocultural factors also play an important role in how pitch is used by speakers to conform - or not - to gender role expectations (Van Bezooijen 1995, Li et al. 2022). Some bilingual speakers have been found to use different pitch ranges in their languages, for example, female bilingual speakers of Welsh and English had a wider pitch range in Welsh than in English (Ordin & Mennen 2017). Languages/dialects can also differ in their intonational patterns, for example, German and English had a narrower pitch range and less variable pitch than Bulgarian and Polish (Andreeva et al. 2014). Across varieties of English, research has found differences in phonological pitch accent categories as well as the frequency with which particular pitch accents are used (Grabe & Post 2002, Grabe et al. 2005). However, less work has examined to what extent there are differences in overall phonetic pitch patterns across varieties of English.

The current study examines (in hertz) f0 range, mean f0, median f0, onset minus offset f0, and pitch dynamism quotient (PDQ, a measure of overall variability of pitch (Hincks 2004, Meer & Fuchs 2021)) of female speakers of three varieties of English: Newcastle, London (where the speakers had a Jamaican background) and Ireland (Dublin) from a reading task and a story retelling task. Recordings from the IViE corpus (Grabe et al. 2002) were used. These were annotated into sentences, and Praat scripts were run to take the above measures for each sentence. Four female speakers of each variety, aged around 16 years old, were examined. 30-40 sentences were examined per speaker per task. There was a total of 766 sentences.

Linear mixed effects regression models were run for the four acoustic measures separately. For each measure, models were built up term by term to determine the best model, using model comparison (the *anova* function in R). The fixed factors that were explored were Dialect and Task (read or retold). Speaker was included as a random factor. For four of the five measures - range, mean, median and PDQ - the best model was one with just Task as a fixed factor, where the Retold task had a significantly lower mean and median f0, narrower range and smaller PDQ than the Read task for all dialects. For onset minus offset, the best model was one with an interaction of Task and Dialect. Pairwise tests showed that only for Newcastle and Dublin, the Retold task had a smaller value than the Read task, while for London there was no difference. For the Read task, London had a smaller value than the other two dialects.

The results indicate that the dialects do not differ in most of the measures, at least for the speakers examined. Examining the data qualitatively shows that Newcastle and London speakers showed instances of a negative number for onset minus offset, due to final rises. Read speech had more pitch movement and a wider range and higher mean than a retold story. This may be to do with the presence of dialogue in the read task. These findings demonstrate that the type of task used in intonation research has significant effects. Further dialects and male speakers are currently being examined.



Andreeva, B., Demenko, G. Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Oleskowicz-Popiel, M. & Trouvain, J. (2014). Comparison of pitch range and pitch variation in Slavic and Germanic languages. *Speech Prosody*, 776-780.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech, 38*(3), 253-265.

Grabe, E. & Post, B. (2002). Intonational variation in the British Isles. *Speech Prosody*.

Grabe, E., Kochanski, G., & Coleman, J. (2005). The intonation of native accent varieties in the British Isles: potential for miscommunication? *English Pronunciation Models: A Changing Scene* (pp. 311-337).

Grabe, E., Nolan, F. & Post, B. (2002). *English Intonation in the British Isles: The IViE Corpus*. Phonetics Laboratory, University of Oxford, Department of Linguistics, University of Cambridge, ESRC Grant R000237145.

Hincks, R. (2004). Processing the prosody of oral presentations. *Proceedings of InSTIL/ICALL*, Venice, Italy.

Hollien, H. (1960). Some laryngeal correlates of vocal pitch. *J. Speech & Hearing Research, 3*, 52-58.

Li, A., Purse, R., & Holliday, N. (2022). Variation in global and intonational pitch settings among black and white speakers of Southern American English. *JASA*, 152, 2617–2628.

Meer, P. & Fuchs, R. (2022). The Trini sing-song: sociophonetic variation in Trinidadian English prosody and differences to other varieties. *Lang. & Speech, 65*(4), 923-957.

Ordin, M. & Mennen, I. (2017). Cross-linguistic differences in bilinguals' fundamental frequency ranges. *J. Speech & Hearing Research, 60*, 1493-1506.

10. Italians in contact: obstruents' accommodation among internal migrants in Florence.

Giuditta Avano
University of Pisa

Research on Second Dialect Acquisition (SDA) [1][2] investigates the linguistic effects of long-term accommodation [3][4] on speakers of a dialect (D1) who migrate to an area where a different variety of their language (D2) in spoken. The peculiarity of this migration situation is that the two languages in contact are mutually intelligible, so migrating would not necessarily require the acquisition of a new code to communicate, as it would happen for contexts of Second Language Acquisition [1], thus language change might occur for the will to get approval from the arrival community [3][4]. Within the Italian context, studies on the linguistic effect of varieties' contact due to migration have been quite rare so far [5], [6], [7], [8]. We believe that introducing SDA studies within the Italian context would be of particular interest to understand the linguistic and sociolinguistic effect on accommodation due migration contact [9], as Italian internal migration is a relevant phenomenon [10] and the country presents a very complex linguistic situation in which there is a large diatopic variation, especially concerning phonetics [11]. In fact, as "pronunciation is less subject to the pressure of the standard norm than other levels of the language system" , spoken Italian without regional phonetic features is extremely rare, even in formal situation and among speakers with a high level of education [12:20]. To investigate the effects of migration linguistic contact on Italian varieties' ongoing changes, we chose to explore the speech of internal Italian migrants living in Florence. In fact, among Italian varieties, Florentine has a special position, as it was used as a model during the past centuries for the definition of a Standard Italian (SI) [13]; however it does not corresponds to the standard norm, as it shows some local phonetic features, particularly regarding obstruents [14]. Furthermore, recently, it might have lost prestige at the national level [15], [16].Will internal migrants, speakers of other varieties, adopt the Florentine pronunciation, even when it differs from the national norm, or will they preserve their own phonetic features, given the high tolerance on pronunciation? Which linguistic and sociolinguistic factors would influence these dynamics?

In order to answer the research questions above, we run an analysis of intervocalic obstruents' realizations among 29 speakers of different regional Italian - 10 Campanian (CamI), 7 Lombard (LI) and 12 Calabrese (CalI) speakers, aged between 30 and 66 y.o. - who arrived in Florence as adults (age of arrival > than 19 y.o) and had been living in town for more than 10 years. We analysed the phonetic realizations of:

1. intervocalic stops (realized as such [p t k] in SI, CalI and LI and as lenited outputs [p̬ t̬ k̬] in CamI) to observe for an adoption of the Florentine Italian (FI) spirantized pronunciation [ɸ θ h]. e.g. SI [buːka], FI [buːha] "hole".
2. intervocalic affricates (realized as such [ʧ ʤ] in SI and LI and [ʧ ʤː] in CalI, [ʃ ʤː] in CamI) to observe for an adoption of FI de-affricated pronunciation [ʃ ʒ]. e,g. SI [baːʧo], FI [baːʃo] "kiss".

Obstruents (N= 2278) were elicited though a sentence reading task and classified with *praat* [17]. Acoustic cues as CoG [18], [19], duration and intensity [20] were extracted to compare migrants' realizations to those of 20 Florentine speakers of about the same age.

SDA studies [21], [22], have showed that each feature social prestige and salience have a role in determining its accommodation or maintenance. Given that deaffrication is considered as less salient and sociolinguistically marked than intervocalic spirantization [23], we expect to find a greater amount of Florentine-like realization for the former phenomenon than for the latter one. Given that speakers are aware of Florentine spirantization for the /k/ segment only [24, p. 20], we will try to understand if its higher salience shows an effect in the direction of a higher or a lower degree of accommodation toward Florentine pronunciation. Moreover, we will investigate if different speakers' D1 has an effect on accommodation both for linguistic (major or minor linguistic distance between D1 and D2 pronunciation) and sociolinguistic reasons [15], [16]. To conclude, we will investigate if, together with an effect of sex [8], [25], [26] and education, an effect of a positive explicit attitude [27] toward the Florentine accent is found on Florentine features' adoption.

[1] Siegel, J., 2010. *Second Dialect Acquisition*, 1st ed. Cambridge University Press.

[2] Nycz, J., 2015. Second Dialect Acquisition: A Sociophonetic Perspective. *Language and Linguistics Compass*, 9, 11, 469–482.

[3] Giles, H., Taylor, D. M., & Bourhis, R., 1973. Towards a theory of interpersonal accommodation through language: some Canadian data. *Lang. Soc.*, 2, 177–192.

[4] Trudgill, P., 1986. *Dialects in contact*. Oxford: Blackwell.

[5] Di Salvo, M., 2009. Interferenza fonetica in siciliani e lucani a Napoli: variazione interregionale e italiano regionale. *L'Italia dialettale*, 141–162.

[6] Giacchi, L., 2010. Per una ridefinizione dei rapporti tra italiano e dialetto in un contesto immigratorio interno: analisi in una località toscana. Università di Firenze.

[7] Nese, N., 2016. Le affricate dentali nel collegio universitario 'Giasone del Maino'. Un approccio sociofonetico. Tesi di Laurea triennale, Università di Pavia.

[8] Nese, N., & Meluzzi, C., 2017. Accomodamento ed emergenza di varianti fonetiche: le affricate dentali intermedie a Pavia e Bolzano. In Bertini, C., Celata C., Lenoci G., Meluzzi C.(Eds.), *Fattori sociali e biologici nella variazione fonetica*. Milano: Officinaventuno, 67–82.

[9] Auer, P., & Hinskens, F., 2005. The role of interpersonal accommodation in a theory of language change. In Auer, P., Hinskens, F., Kerswill, P. (Eds.), *Dialect Change*. Cambridge: Cambridge University Press, 335–357.

[10] Ascoli, U., 1979. *Movimenti migratori in Italia*. Bologna: Il Mulino.

[11] Vietti, A., 2019. Phonological Variation and Change in Italian. In *Oxford Research Encyclopedia of Linguistics*, Oxford: Oxford University Press.

[12] Cerruti, M., 2011. Regional varieties of Italian in the linguistic repertoire. *International Journal of the Sociology of Language*, 210, 9–28.

[13] Della Valle V. & Patota, G. (2007) *L'italiano.* Milano: Sperling & Kupfer.

[14] Giannelli, L., & Savoia, L. M., 1978. L'indebolimento consonantico in Toscana I. *Rivista italiana di dialettologia*, 2, 25–58.

[15] De Pascale, S., & Marzo, S., 2016. Gli italiani regionali. Atteggiamenti linguistici verso le varietà geografiche dell'italiano. *Incontri*, 31, 1, 61 – 75

[16] De Pascale, S., Marzo, S., & Speelman, D., 2017. Evaluating regional variation in Italian: towards a change in standard language ideology? In Cerruti, M., Crocco, C., & Marzo, S. (Eds.), *Towards a New Standard*. Berlin: De Gruyter, 118–142.

[17] Boersma, P., & Weenink, D., 2023. *Praat: doing phonetics by computer* [Computer program].

[18] Harrington, J., 2010. Acoustic Phonetics. In W., Laver, J., & Gibbon, F. (Eds.), *The Handbook of Phonetic Sciences*. Hardcastle, Hoboken: Wiley-Blackwell, 81–129.

[19] Di Canio, P., 2013. [Online]. Available: https://www.acsu.buffalo.edu/~cdicanio/
scripts/Time_averaging_for_fricatives_4.0.praat

[20] Sorianello, P., 2003. Spectral characteristics of voiceless fricative consonants in Florentine Italian. In Solé, M. J., Recasens, D., and Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions, 3081–3084.

[21] Auer, P., Barden, B., & Grosskopf, B., 1998. Subjective and Objective Parameters Determining 'Salience' in Long-term Dialect Accommodation. *Journal of Sociolinguistics*, vol. 2, no. 2, 163–187.
[22] Nycz, J., 2016. Awareness and Acquisition of New Dialect Features. In Babel, A. M. (Ed.), *Awareness and Control in Sociolinguistic Research*, Cambridge: Cambridge University Press, 62–79.
[23] Marotta, G., 2014. New parameters for the sociophonetic indexes: Evidence from the Tuscan varieties of Italian. In Celata, C., & Calamai, S. (Eds.), *Studies in Language Variation*. Amsterdam: John Benjamins Publishing Company, 137–168.
[24] Cravens, T. D., 2000. Sociolinguistic subversion of a phonological hierarchy. *WORD*, 51, 1, 1–19.
[25] Rys, K., & Bonte, D., 2006. The role of linguistic factors in the process of second dialect acquisition. in *Studies in Language Variation*, Hinskens, F. L. (Ed.), Amsterdam: John Benjamins Publishing Company, 201–215.
[26] Avano, G., 2021. L'attrito dialettale: uno studio esplorativo sulla Gorgia Toscana. Università di Bologna.
[27] Lee, J., Idemaru, K., & Vaughn, C., 2023. Effects of Sociolinguistic Factors on Second Dialect Acquisition of North Korean Refugees in Seoul. In Skarnitzl, R., Volín, J. (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*. Prague: Guarant International, 3671–3675.

11. Rime merger in Beijing Mandarin retroflex suffixation: expanding the typology of incomplete neutralisation

Yutong Wang[1], Mitko Sabev[2,1]
[1]University of Cambridge, [2]Universität des Saarlandes

Incomplete neutralisation (IN), wherein phonologically neutralised categories yield phonetically non-neutralising outputs, has been documented in various languages. Frequently-discussed evidence includes American /t/-flapping [1] and German final devoicing [2]. This paper presents a clear case of IN using data from Beijing Mandarin (BM). In Beijing retroflex suffixation (BRS), a suffix *er* (transcribed as [ɚ] or [ɻ]) is attached to the rime to denote diminutiveness, as exemplified in (1), which substantially changes the quality of the rimes by adding r-colouring to the rime while deleting certain segments. This results in the loss of distinctiveness between some rime categories.

(1) [pʰai] + *er* → [pʰa˞ɻ] 'tag'; [pʰan] + *er* → [pʰa˞ɻ] 'plate'

We conducted a production experiment involving 11 native BM speakers. The experiment elicited 136 pairs of unsuffixed and suffixed forms, embedded in carrier sentences, as shown in (2). The speakers were free to choose between the affirmative or negative form of the carrier sentence as a response to the prompt: 'do you know what the meaning of <target word> is?' in a naturalistic way.

(2) 我（不）知道 <target word> 的意思是什麼。
I (don't) know the meaning of <target word>.

Using the Montreal Forced Aligner [3] and manual correction, acoustic data, including the first three formant frequencies (F_1, F_2, F_3) and f_0 values, were extracted at 21 equidistant points in the rime using Praat scripts [4, 5]. To capture the subtle acoustic differences, the study employed generalised additive mixed models (GAMMs, [6]) to analyse dynamic formant trajectories and f_0 contours across all 34 rime × 4 tonal conditions in BM. In total, 251,332 data points were obtained, and 77 GAMMs were fitted to model dynamic changes in formant trajectories and tonal contours. The last five points from each trajectory smooth were removed to reduce co-articulatory effects from the following segment. Since the output of GAMMs can be complex and lengthy (e.g., [7, 8]), we reported and visualised the predicted measurements from the models, focusing on pairwise comparisons of suffixed vs. unsuffixed rimes, as well as groups of potentially merging suffixed rimes, as exemplified in Figure 1.

The results show that suffixation causes spectral changes across all rime types, with varying degrees of vowel raising, lowering, and retraction, depending on the type of nuclear vowel. Low-nucleus rimes like [a], [ai] and [an] exhibit significant spectral change and complete merger after suffixation. High-nucleus rimes also demonstrate substantial changes: lowering and retraction are observed for the high-nucleus group, and openness and more backness are found in suffixed [ie, ye]. Therefore, suffixed [ie] and [ye] remain distinguishable within their respective groups, as the quality of [e] seems to be preserved. The mid nuclear vowel [ɤ] exhibits less quality change and remains distinct in its group after suffixation. Tone contour shapes are preserved after suffixation, with a significant decrease of overall F0 values but no differences within the merging suffixed groups.

Thus, some of the merging suffixed rime categories are spectrally distinguishable, but these differences seem unlikely to be large enough to be perceptible, which remains to be confirmed. These findings of phonetic incompleteness or gradience challenge the prevailing assumption that phonologically neutralised contrasts, which are expected to be categorical, do not exhibit gradient surface distinctions. This poses challenges to the traditional modular feed-forward phonology-phonetics interface [9-11], which posits that phonetics only has access to the discrete phonological output that does not contain any gradient phonetic information.

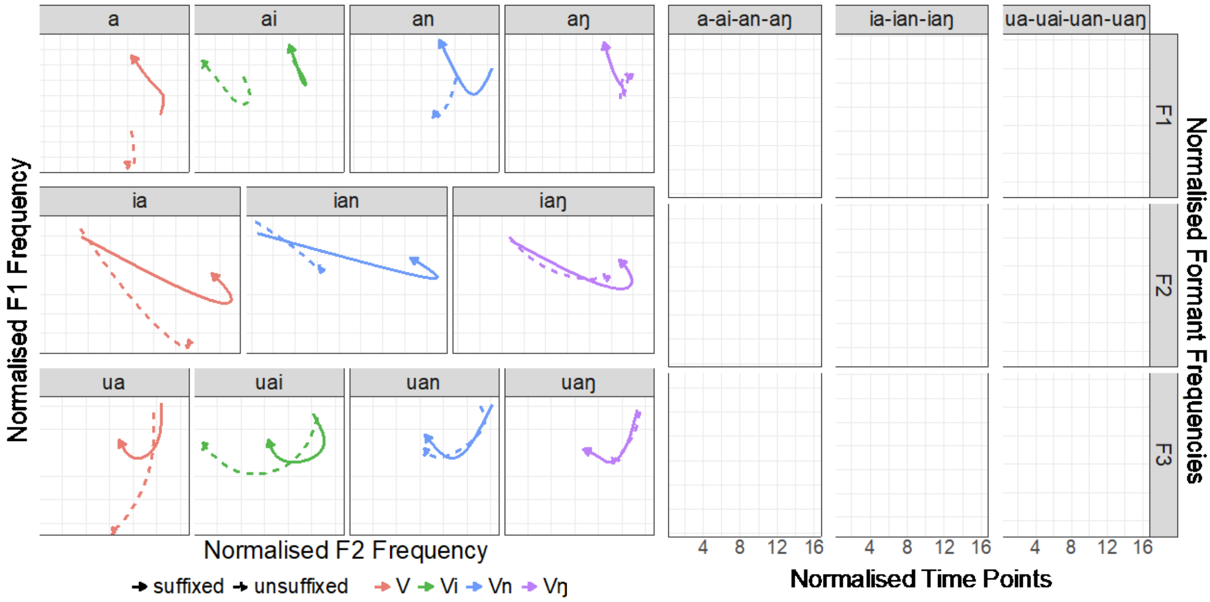


Figure 1. *Left: Average predicted GAMM trajectories for rimes with low nuclei based on the interaction term between suffixation and rime type. Right: Plots of predicted GAMM trajectories for suffixed normalised formants of rimes with low nuclei. Solid lines: formant trajectories of suffixed rime; dashed lines: unsuffixed rimes.*

[1] A. Braver, 'Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping', *Lingua*, vol. 152, pp. 24–44, Dec. 2014, doi: 10.1016/j.lingua.2014.09.004.

[2] M. van Oostendorp, 'Incomplete devoicing in formal phonology', *Lingua*, vol. 118, no. 9, pp. 1362–1374, 2008, doi: https://doi.org/10.1016/j.lingua.2007.09.009.

[3] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, 'Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi', in *Interspeech*, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:12418404

[4] J. Stanley, 'A Tutorial on Extracting Formants in Praat'. 2017. [Online]. Available: https://joeystanley.com/blog/a-tutorial-on-extracting-formants-in-praat/

[5] C. Kaland, 'Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours', *Journal of the International Phonetic Association*, vol. 53, no. 1, pp. 159–188, 2023, doi: 10.1017/S0025100321000049.

[6] T. Hastie and R. Tibshirani, 'Generalized Additive Models', *Statistical Science*, vol. 1, no. 3, pp. 297–310, 1986.

[7] K.-A. Butcher, 'The MOAN-MOWN and MOAN-GOOSE mergers in Lowestoft English: Perception and production', Ph.D, University of Cambridge, 2022.

[8] J. A. Stanley, M. E. L. Renwick, K. I. Kuiper, and R. M. Olsen, 'Back Vowel Dynamics and Distinctions in Southern American English', *Journal of English Linguistics*, vol. 49, no. 4, pp. 389–418, 2021, doi: 10.1177/00754242211043163.

[9] N. Chomsky and M. Halle, *The sound pattern of English*, 1st MIT Press pbk. ed. Cambridge, Mass: MIT Press, 1991.

[10] R. Bermúdez-Otero, 'Diachronic phonology', in *The Cambridge Handbook of Phonology*, P. deEditor Lacy, Ed., in Cambridge Handbooks in Language and Linguistics. ,

Cambridge University Press, 2007, pp. 497–518. doi: 10.1017/CBO9780511486371.022.
[11] J. B. Pierrehumbert, ‘Word-specific phonetics’, in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds., Berlin, New York: De Gruyter Mouton, 2002, pp. 101–140. doi: doi:10.1515/9783110197105.1.101.

12. Regional variation of pre-boundary lengthening in a corpus of German dialects

Nadja Spina, Alfred Lameli
Forschungszentrum Deutscher Sprachatlas

Pre-boundary lengthening (PBL), the lengthening of segments immediately preceding prosodic boundaries, is an important correlate of prosodic phrasing ([1]). As such, PBL has received intense research interest for half a century. For this time, mainly experimental studies focused on cross-linguistic variation in PBL from a language typological perspective (e.g., [2]). Language-specific variation in PBL *across* languages is hence well-studied, whereas little is known about cross-regional variation *within* one language. Only few studies have addressed this research gap. Rao reports regional variation in PBL between Cuban, Ecuadorian and European Spanish ([3]) and calls for further corpus studies investigating variation in PBL between dialects and "macrolects" (79) of other languages to validate the results for Spanish.

In German, local dialects and the more standard-targeted "regiolects" ([4]) are well studied on the segmental level but poorly regarding prosody. The present study investigated data from the REDE corpus (project "regionalsprache.de" [5]), which contains recordings of regional varieties from 150 locations in Germany. Nine locations were selected for analysis, and all speakers provided for each location were included in the analysis. In total, 10 Low German, 14 West Central German and 20 Upper German speakers were analyzed. The speakers heard 40 sentences that consisted of several intonation phrases (IPs), such as in (1), once in Standard German and once in dialect. Their task was to repeat the sentences, and one time target their native dialect (dialect register) and another time Standard German (regiolect register). Each speaker thus produced 80 sentences. Dialect and regiolect registers were hence compared within the same speakers. Segment and word boundaries were automatically annotated using WebMAUS [6]. Syllable boundaries and word stress were manually annotated in Praat. The distance of each segment and syllable from a prosodic boundary was annotated using an automated procedure. Phrase-final words were manually annotated by labelling boundary tones after the GToBI standards [7]. The duration of each segment was measured. Linear mixed effects models accounting for DURATION as a function of BOUNDARY and REGION were fitted to the data in R. SPEAKER, LEMMA and SENTENCE were included as random effects.

Figure 1 shows a highly significant PBL effect ($p<0.001$) in the (stressed) penultimate and ultimate syllables of disyllabic words across regions. Figure 2 shows the difference between final and non-final mean segment duration in percent. It clearly illustrates that PBL starts on the stressed syllable and increases towards the boundary. This confirms experimental results for Standard German, which show a progressive increase in PBL towards the prosodic boundary ([8]). Regional differences in PBL could be observed in terms of a highly significant register effect for the ultimate syllable in Low German ($p<0.001$). That is, PBL differs in Low German between speech registers in the syllable where PBL is strongest and therefore likely to be perceptually most salient ([8]). This result has important implications for the status of Low German. Thus, Low German varieties differ considerably from High German (i.e., Central and Upper German) varieties as they developed differently with regard to late West Germanic forms. For example, compared to High German dialects, Low German has a preference for simpler consonant clusters and syllable structure as well as a tendency towards monosyllabicity ([9]). As a consequence, Low German differs so much from Standard German that it could be regarded as an independent language rather than a variety of German. The stylistic choice of register in the Low German area therefore resembles bilingualism in that speakers strictly distinguish between dialect and standard forms ([10]). In the High German area, however, speakers mix both forms. The present results suggest that there are different extents of PBL in Low German and Standard German and that Low German speakers distinctly produce this difference in dialect- and standard-targeted speech styles.

(1) [Im Winter]$_{IP}$ [fliegen die trockenen Blätter]$_{IP}$ [durch die Luft herum]$_{IP}$.
'In winter, dry leaves fly about through the air.'
[Es hört gleich auf]$_{IP}$ [zu schneien]$_{IP}$, [dann wird das Wetter wieder besser]$_{IP}$.
'It will soon stop to snow, then the weather will become better again.'
[Der gute alte Mann]$_{IP}$ [ist mit dem Pferd durchs Eis gebrochen]$_{IP}$ [und in das kalte Wasser gefallen]$_{IP}$.
'The good old man has broken through the ice with his horse and fallen into the cold water.'

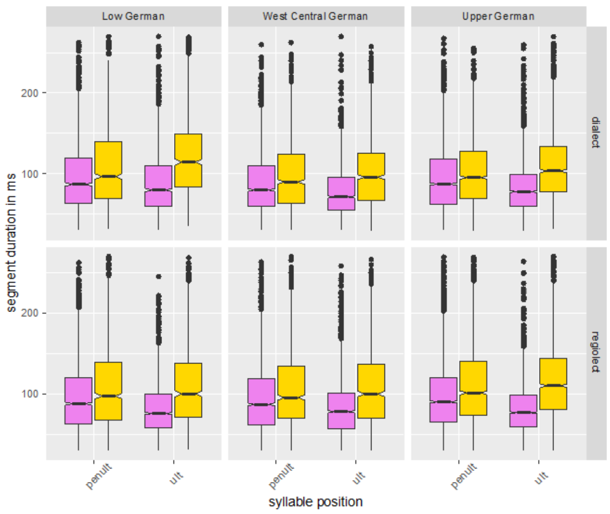


Figure 1: *PBL effect in milliseconds compared over regions and registers in disyllabic words (e.g., WOchen, 'weeks') across speakers*

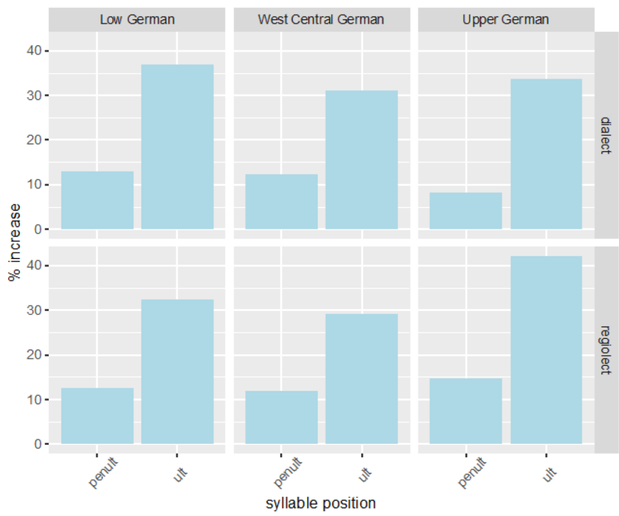


Figure 2: *Strength of PBL effect in percent compared over regions and registers in disyllabic words (e.g., WOchen, 'weeks') across speakers*

[1] Frazier, L., Carlson, K. & Clifton, C. 2006. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 10/6, 245-249.

[2] Nakai, S., Kunnari, S., Turk, A., Suomi, K. & Ylitalo, R. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37, 39-45.

[3] Rao, R. 2010. Final Lengthening and Pause Duration in Three Dialects of Spanish. In: Ortega-Llebaria, M. (Ed.): *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*. Somerville, MA: Cascadilla Proceedings Project, 69-82.

[4] Schmidt, J. E., Herrgen, J. 2011. Sprachdynamik. Eine Einführung in die moderne Regionalsprachenforschung. Berlin: Schmidt.

[5] Schmidt, J. E., Herrgen J., Kehrein, R., Lameli, A., & Fischer, H. 2020 (Eds.): *Regionalsprache.de (REDE III). Forschungsplattform zu den modernen Regionalsprachen des Deutschen*. Marburg: Forschungszentrum Deutscher Sprachatlas.

[6] The Munich Automatic Segmentation System (MAUS), web service, version 3.12, URL: https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic.

[7] Grice, M. & Baumann, S. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191, 267-298.

[8] Schubö, F. & Zerbian, S. 2020. Phonetic content and phonological structure affect pre-boundary lengthening in German. *Proceedings of the 10th International Conference on Speech Prosody* (Tokyo, Japan), 111-115.

[9] Lameli, A. 2022. Syllable structure spatially distributed. Patterns of monosyllables in German dialects. *Journal of Germanic Linguistics* 34/3: 241-287.
[10] Lameli, A. 2024. Gesprochenes Deutsch in den Regionen. Eine Standortbestimmung für die Bundesrepublik Deutschland. In Dannerer, M., Deppermann, A, Proske, N. & Weber, T. (Eds): *Gesprochenes Deutsch. Struktur, Variation, Interaktion*. Berlin, Boston: De Gruyter.

13. The intersection of four phonetic variables in Yucatan Spanish: A "variable swarm" analysis

Jim Michnowicz, Allyson Gee
North Carolina State University

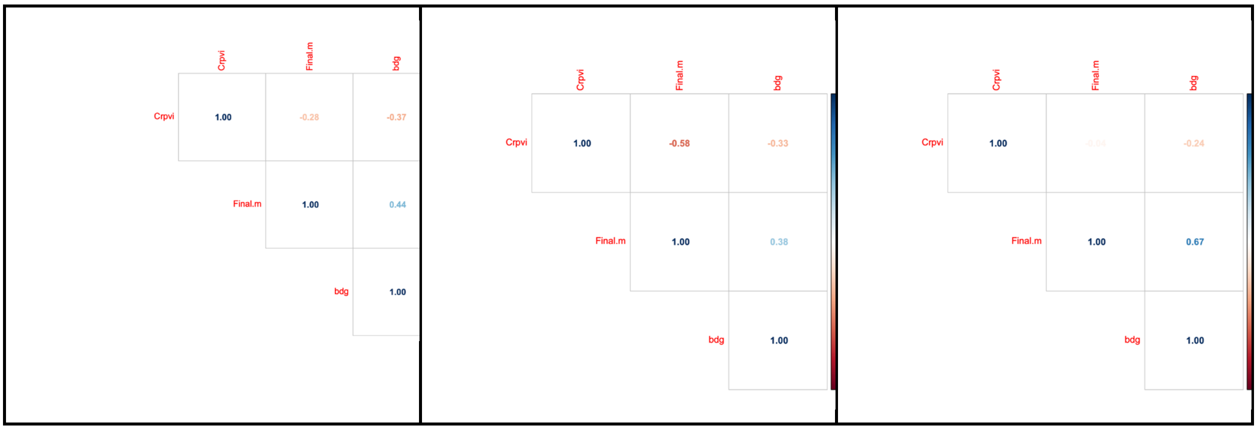
Previous research has argued that sociolinguistic variables are best approached as part of a "variable swarm," where multiple features are studied simultaneously for the same set of speakers [1], [2]. A swarm recognizes that "[e]ach linguistic variable can reveal new meanings and social patterns" [1] that are not apparent through the study of a single variable. The present study conducts a swarm analysis on Yucatan Spanish (YS) data in both sociolinguistic real and apparent time to better understand the patterns of use of traditional YS variables over time.

YS is a regional variety in contact with Yucatec Maya and is one of the regions in Latin America with the greatest influence of an indigenous language [3]. Previous studies have shown that many traditional features of YS, directly or indirectly related to possible Mayan influence, are undergoing a rapid process of standardization in urban areas, such as the capital city, Mérida [4].

The present study builds on previous research by examining four traditional YS phonetic variables: the realization of final /n/ as [m], e.g., [ˈpam] for [ˈpan] [5]; intervocalic stop realizations of /bdg/, e.g., [ˈbi.bo] for [ˈbi.βo] as measured by intensity difference with the following vowel [6], [7]; aspirated /ptk/ as measured by VOT [8]; and vocalic and consonantal prosodic rhythm, measured with Vnpvi and Crpvi, respectively [9], [10]. Data were collected through sociolinguistic interviews at two different points in time (2005 and 2016), which allowed studying ongoing change in YS both in real time and in apparent time [11]. The study uses data from 27 YS speakers, divided by age group, sex, and linguistic background (Spanish monolinguals, Mayan-Spanish bilinguals), and statistical patterns across speakers and variables were analysed using correlations and logistic/linear regressions.

The correlation analysis between three initial variables (VOT analysis is ongoing) for all speakers finds several weak correlations, one of which approaches moderate strength (rates of final -m and stronger realizations of /bdg/), suggesting that the two variables might be related within the speech community (Figure 1). Results by linguistic background show stronger correlations, with Spanish monolinguals showing a moderate negative correlation between Crpvi and final -m (Figure 2), while Mayan bilinguals show a moderately strong correlation between final -m and stop /bdg/ (Figure 3), indicating continued bilingual-based distinctions in the community.

Additional analyses will be conducted to further elucidate possible patterns of change in YS, including the addition of the VOT data, with a particular focus on real-time and apparent differences, and on trends that would only be visible through a "swarm" approach.



1. 2. 3.

Figure 1. *Initial correlations, all speakers*
Figure 2. *Initial correlations, Spanish monolinguals*
Figura 3. *Initial correlations, Maya-Spanish bilinguals*

[1] Thomas, E. (2015). What a swarm of variables tells us about the formation of Mexican American English. Paper presented at LAVIS IV. North Carolina State University Raleigh, NC.

[2] Michnowicz, Jim, Rebecca Ronquest, Sarah Chetty, Georgia Green & Stephanie Oliver. 2023. "Spanish in the Southeast: What a swarm of variables can tell us about a newly-forming bilingual community". Languages 8, 168.

[3] Klee, C. A. (2009). Migrations and Globalization: Their Effects on Contact Varieties of Latin American Spanish. In M. LaCorte & J. Leeman (Eds.), *Español en Estados Unidos y otros contextos de contacto: Sociolingüística, ideología y pedagogía* (39- 66). Madrid: Vervuert/Iberoamericana.

[4] Michnowicz, J. (2015) "Maya-Spanish contact in Yucatan, Mexico: Context and sociolinguistic implications". In S. Sessarego and M. González Rivera (Eds.). New Perspectives on Hispanic Contact Linguistics in the Americas. Madrid: Iberoamericana/Vervuert. 21-42.

[5] Michnowicz, J. (2021). Apparently real changes: Revisiting final/-m/in Yucatan Spanish. In M. Díaz-Campos (Ed.), The Routledge Handbook of Variationist Approaches to Spanish (pp. 249-262). London: Routledge.

[6] Michnowicz, J. (2023). "Real and apparent (time) changes in Yucatan Spanish: the case of /bdg/". In Fernández Cuenca, S., Judy, T. & Miller, L. (Eds.) Innovative Approaches to Research in Hispanic Linguistics: Regional, Diachronic, and Learner Profile Variation. Amsterdam: John Benjamins. 130-151.

[7] Hualde, J. I., Simonet, M., & Nadeu, M. (2011). Consonant lenition and phonological recategorization. *Laboratory phonology*, *2(2),* 301-329.

[8] Michnowicz, J., & Carpenter, L. (2013). "Voiceless stop aspiration in Yucatan Spanish: a sociolinguistic analysis". Spanish in Context 10:3. 410-437.

[9] Michnowicz, J., & Hyler, A. (2020). "The changing accent of Yucatan Spanish". In R. Rao (Ed.). Spanish phonetics and phonology in contact: Studies from Africa, the Americas, and Spain. Amsterdam: John Benjamins. 36-62.

[10] Low, E. L., & Grabe, E. (1995). Prosodic patterns in Singapore English. In ICPhS 95, Volume 3 (636–639), Stockholm.

[11] Chambers, J. K. (2009). Sociolinguistic theory: Linguistic variation and its social significance. rev. ed. Oxford: Wiley-Blackwell.

14. Underlying vowel asymmetry in Okinawan word-initial preglottalization

Antoine Heintzmann
CNRS - Paris VIII University (UMR 7023 SFL); Lyon III University

Ryukyuan languages constitute a subset of the Japonic family, whose common intermediary ancestry is usually known as proto-Ryukyuan (pR). A common thread throughout the branch is the apparent neutralisation of pR mid-vowels (*e, *o) with their high counterparts. Though the surface melodies of both reflexes coincide, the etymologically high vowels (EHV) tend to exhibit some behaviours that distinguish them structurally from corresponding etymologically mid vowels (EMV). In the Northern branch in particular, EHVs tend to undergo "pre-glottalization" in the intermediate phase between the onset and the nucleus [1].

Okinawan, one of the most spoken and better documented Ryukyuan languages, belongs to the Northern branch. The source used in this study is a dictionary compiled for Okinawan by the NINJAL (沖縄語辞典, *Okinawa-go jiten*) in 1963, counting 14543 entries transcribed using a notation making conspicuous pre-glottalized variants of phonemes.

In Okinawan, "pre-glottalization" manifests in the form of word-initial glottal stop insertion, occurring in three precise contexts: before initial vowels with empty onsets, initial glides and moraic nasals. The occurrence of this phenomenon in Okinawan stands out cross-linguistically from other occurrences since it is contrastive, glottal insertion being the unmarked issue in the case of empty onsets and marked in the other two (see minimal pairs in (1)).

After a comprehensive spanning of the corpus though, we find that etymological vowel height does not directly affect pre-glottalization; instead, glottal insertion is mainly inhibited by etymological glides, which have syncopated in mainland Japanese. However, etymological height does play a part in the formation of other pre-glottalized segments such as the initial nasal mora. In fact, those segments tend to arise at places previously occupied by EHV. Furthermore, the emergence of this segment is also dependent on the etymological height of the following vowel, since this kind of change only arises if the EHV is followed by an EMV in the following mora/syllable.

In a metrical framework, the height difference in an earlier form may have impacted the heading of the (moraic) foot in cases like (1c), as mid vowels are higher in the sonority hierarchy and thus more susceptible to attract head status [2]; the iambic configuration then survived even as the contrast became neutralized. However, the (syllable containing) the head mora not being aligned with the left edge of the prosodic word is a violation of its role as a "stronger edge" (represented as ALIGN-HEAD-L(SYL, PRWD)). This discrepancy is solved by apheresis, which is followed by initial gemination [3], in order to keep the minimum of 2 morae (documented in other Japonic languages in [4], represented here as $*[\mu]_\omega$). The overriding of the typological markedness of initial geminates (*#Cː) is represented here in (2) for the word ˀɴni 'rice paddy'.

Government phonology [5–7] on the other hand, attributes this chain of events to an instance of proper government between the nuclei, which implies a difference in charm between EHV and EMV, a segmental property conditioning a segment's ability to govern, directly derived from its internal (elemental) structure. The nasal insertion here is here not analyzed as a geminate but rather as a spreading of the nasal element (represented as |L| since [8]) to the preceding V slot (see (3)).

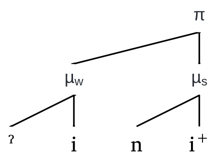
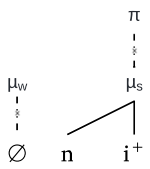
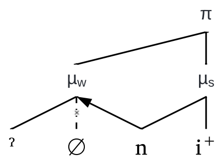
Overall, the choice of the framework used for the representation of the emergence of preglottalized segments yields deep implications on the overall shape of Okinawan phonology, since, unlike the metrical explanation, a government-based analysis suggests that the Ryukyuan vowel "neutralisation" is merely superficial with each etymological vowel leading to a distinct feature structure.

(1) a. **ˀijuɴ** *'to enter'* ≠ **ʲijuɴ** *'to sit'*
ˀutu 'sound' ≠ **ʷutu** 'husband'

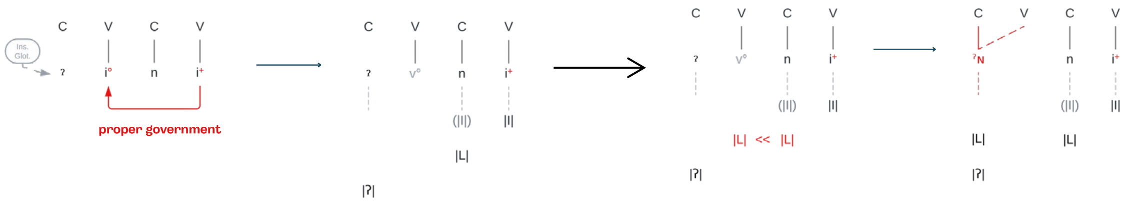
b. **ˀja**: *'you'* ≠ **ja**: *'house'*
ˀwi: 'top' ≠ **wi**: 'handle, knob'

c. **ˀɴ**na 'feces' ≠ **ɴ**na 'everyone'
ˀɴd͡ʒi 'Igei (place name)' ≠ **ɴ**d͡ʒi 'spike'

(2)

*ine > *ini > [ˀ***n***ni]	Align-Head-L(Syl, PrWd)	*[μ]$_{\omega}$	*#C:	Max
a. 	*!			
b. 		*!		*(*)
☛ c. 			*	*

(3)



[1] Pellard T. 2 Ryukyuan and the reconstruction of proto-Japanese-Ryukyuan. In Frellesvig B, Kinsui S, editors. Handbook of Historical Japanese Linguistics, De Gruyter; 2024, p. 39-68. http://doi.org/10.1017/CBO9781614512851-003.
[2] Kager R. Feet and metrical stress. In: Lacy P de, editor. The Cambridge Handbook of Phonology, Cambridge: Cambridge University Press; 2007, p. 195-228. http://doi.org/10.1017/CBO9780511486371.010.
[3] Davis S. On the Representation of Initial Geminates. Phonology 1999; 16:93-104.
[4] McCarthy J, Prince A. Prosodic Morphology. John J. McCarthy 1995.
[5] Kaye J, Lowenstamm J, Vergnaud J-R. The Internal Structure of Phonological Elements: A Theory of Charm and Government. Phonology Yearbook 1985; 2:305-28.
[6] Kaye J, Lowenstamm J, Vergnaud J-R. Constituent structure and government in phonology. Phonology 1990; 7:193-231. https://doi.org/10.1017/S0952675700001184.

[7] Scheer T. A Lateral Theory of Phonology: What is CVCV, and why Should it Be? Walter de Gruyter; 2004.
[8] Backley P. An Introduction to Element Theory. Edinburgh University Press; 2011.

15. Loss of gender differences in coda rhotic articulations in US Puerto Rican Spanish

Mary Elizabeth Beaton, Zhe Wang,
Denison University

This study examines rhotic manner of articulation across three sociolinguistic generations (G1, G1.5, G2) of Puerto Rican Spanish (PRS) speakers in Lorain, Ohio. The bilingual and multi-dialectal environment in Lorain offers a setting to examine the factors influencing language change. Coda /r/ shows considerable variation in PRS, making it an ideal feature to explore potential shifts in a multilingual context. Previous studies on PRS both on the island and in the diaspora in the United States find articulations ranging from taps [ɾ], fricatives [ř], deletions, aspirations [h], geminates, trills [r], and approximant sounds ranging from an American English-like [ɹ], to neutralization with laterals [l], to productions somewhere between [ɹ] and [l], sometimes represented as [lɹ] ([1], [2], inter alia). While most studies have focused on the approximant realizations and the possibility of (incomplete) neutralization of /r/ to [l], the present study considers articulations more broadly to examine possible influences of English and of other dialects of Spanish on Puerto Rican coda /r/.

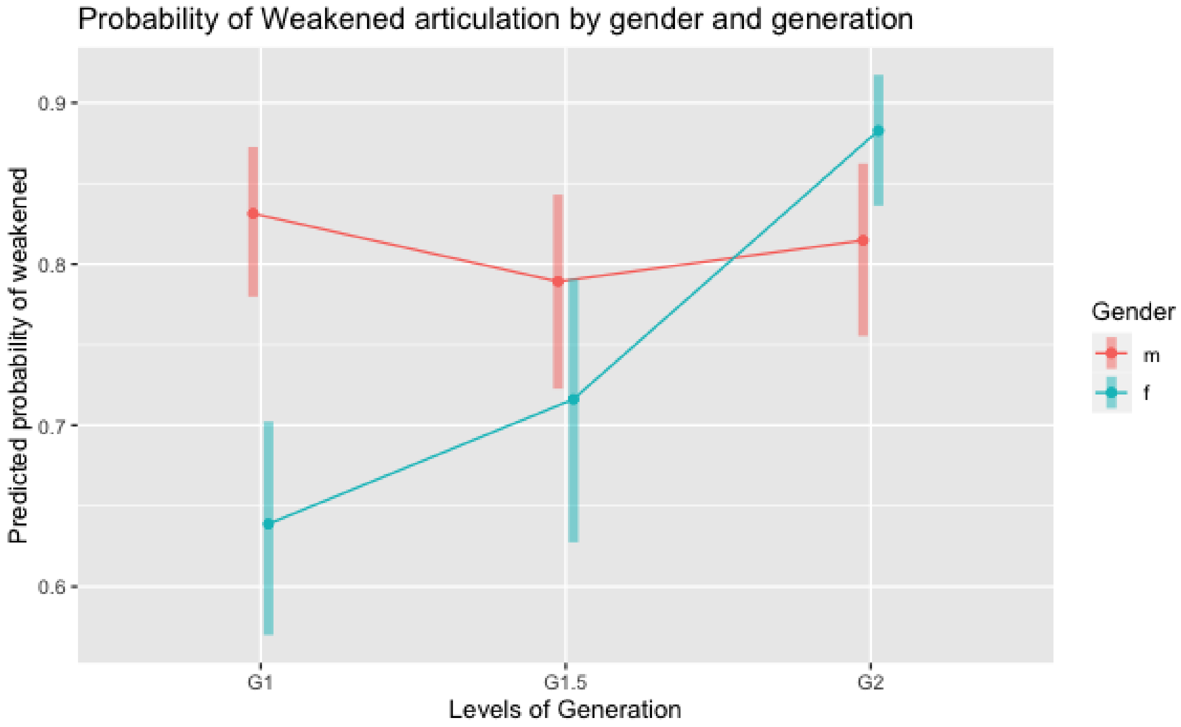
Due to steel industry recruitment beginning in the 1940s, 30% of Lorain's population is Hispanic, of which nearly 80% is of Puerto Rican heritage. Puerto Ricans in Lorain interact daily with a smaller but still sizable Mexican population. Intense contact with Mexican varieties of Spanish, which feature 'strong' tap [ɾ] and fricative [ř] articulations, could condition a shift toward more of these articulations across sociolinguistic generations. However, contact with English [ɹ] could further PRS's tendency for coda lenition. The central research questions are: a) Are there generational changes in coda /r/ articulation in Lorain PRS? b) If so, is the directionality of the change towards the 'strong' pronunciations in Mexican Spanish or toward a greater number of 'weak' /r/ common in Caribbean varieties and English? c) How is the change stratified socially in terms of gender, age, and origin (family history from the capital San Juan vs. other areas of the island)?

To answer these questions, we utilized 16 interviews from a corpus created by the first author in Lorain. Starting five minutes into each interview, coda rhotics (N = 2,363) were classified according to manner of articulation using evidence from the spectrogram and waveform in Praat. Using a logistic regression mixed effects model in R, we explored the conditioning of 'strong' (taps and fricatives) and 'weak' (approximants and deletions) /r/ by the following independent variables: vowel preceding rhotic, manner of articulation of sound following rhotic, lexical stress, lexical frequency, position in word (medial or final), sociolinguistic generation, age, gender, and origin. In addition, we considered models with interactions between generation and the other social variables (age, gender, and origin).

Our main finding is that rhotic articulations change across sociolinguistic generations in the direction of more weakening, but only in women. G1 women differ significantly from men, using more 'strong' /r/ articulations. However, women in subsequent generations produce more 'weak' /r/ like men, erasing the gendered differences seen in the first generation (see Figure 1). In both genders, the conditioning of following sound is lost by G2, but there is no generational change in the influence of the vowel preceding the rhotic.

Our findings suggest that contact with English plays a stronger role in change than does dialect contact. Other factors, such as shifts in the social meaning vis à vis the prescriptive status and identity marking of /r/ articulations in the contact environment are explored. Our study bolsters previous claims that majority language influence tends to be stronger than dialect contact ([3]) and that social conditioning in heritage languages may be less robust than in first generation (im)migrants ([4]).

Figure 1



[1] Navarro Tomás, T. 1948. *El español en Puerto Rico*. San Juan: Editorial de la Universidad de Puerto Rico Río Piedras.

[2] Simonet, M., Rohena-Madrazo, M. & Paz, M . 2008. Preliminary Evidence for Incomplete Neutralization of Coda Liquids in Puerto Rican Spanish. In Laura Colantoni & Jeffrey Steele (eds.), *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 72–86. Somerville, MA: Cascadilla Proceedings Project.

[3] Erker, D. & Otheguy, R. 2016. Contact and coherence: Dialectal leveling and structural convergence in NYC Spanish. *Lingua* 172. 131–146.

[4] Kang, Y. & Nagy, N. 2016. VOT merger in Heritage Korean in Toronto. *Language Variation and Change* 28(2). 249–272.

16. Vowel centralization in highland and lowland Peruvian Spanish

Patricia de Crignis[1] Conceição Cunha[2]
[1]*LMU Munich,* [2]*LMU Munich*

Spanish is generally considered a syllable-timed language with stable vowels (Szczepaniak 2009) and several acoustic studies show minimal differences in formant values between stressed and unstressed vowels in both Peninsular Spanish (Almeida 1990, Ortega-Llebaria & Prieto 2010) and Hispano-American Spanish (Vaquero & Guerra de la Fuente 1992, Delforge 2009). However, vowel lenition processes, including centralization, have been reported in some *tierras altas* (highland) varieties in Hispano-American Spanish and explained by an L1-to-L2 transfer from indigenous languages (Nahuatl in the Valley of Mexico, Quechua and Aimara in the Andes; Canellada & Zamora 1960, Hundley 1986, Delforge 2009). Since vowel centralization has hardly been studied in the Peruvian *tierras altas* varieties of Spanish (Hundley 1983, Delforge 2008), we ask, first, whether vowel weakening is prevalent in Peruvian Spanish, and second, whether centralization patterns in L1 and L2 speakers indicate a language contact phenomenon.

To answer these questions, data was collected via field research in the Peruvian department of Arequipa from a total of 50 participants, both in a village in the *tierras altas* (n=20), where a language shift from Quechua to Spanish can be observed, and in a village in the coastal *tierras bajas* (lowlands; control group; n=30), where Spanish has developed without language contact. The participants in the *tierras altas* either speak Spanish as L1 (n=7) or are bilingual with Spanish as L2 and Quechua as L1 (n=13). The data was collected from three generations and two genders (f/m). All informants were asked to name 53 pictures in which the Spanish vowels occur in tonic, pretonic and posttonic positions. The data was recorded with a Zoom H4 mobile digital audio recorder in WAVE format with a sampling rate of 44100 Hz. The data was then segmented in Praat (Boersma/Weenink 2018), imported into an Emu database (Tcl Core Team 2018) and analyzed using R (R Development Core Team 2018). A total of 6,253 vowels (2,083 stressed and 4,170 unstressed vowels) were analysed.

The analysis shows that vowel centralization is not only common among L1 and L2 speakers in the *tierras altas*, but, surprisingly, also among L1 speakers in the *tierras bajas*. Unstressed /a/ is clearly centralized towards schwa in all three speaker groups. Unstressed /i/, /e/ and /o/ are less centralized: In L1 speakers (*tierras altas* and *tierras bajas*), /e/ moves slightly towards [ɪ] and /o/ towards /ʊ/, /i/ centralizes slightly towards the middle of the vowel space (towards [ɨ]). In L2 speakers, /e/ and /o/ are less centralized and more raised (see Figure 1). A comparison by word position revealed that centralization in L1 speakers (*tierras altas* and *tierras bajas*) is more prominent syllable finally. Furthermore, for L2 speakers, unstressed /e/ and /o/ are raised only in final position, but not in other word positions. Thus, vowel raising in this corpus primarily affects unstressed final vowels. This may be attributed to a parsing error: L2 speakers may perceive the centralized-raised vowels [ɪ] and [ʊ] produced by L1 speakers in final position as [i] and [u]. This interpretation is further supported by the fact that L2 speakers are quite capable of articulating unstressed /e/ and /o/ in other word positions. The formant values of stressed and unstressed /e/ hardly differ, and /o/ and /a/ even appear hypercorrectly strengthened (see Figure 2).

Overall, the data do not indicate an L1-to-L2 transfer, as centralization is also widespread in the control group in the *tierras bajas*. Furthermore, the data rather indicate a non-target like articulation of the L2 speakers. To complete the picture, data on the Quechua contact variety spoken in the *tierras altas* of Arequipa should also be analyzed.

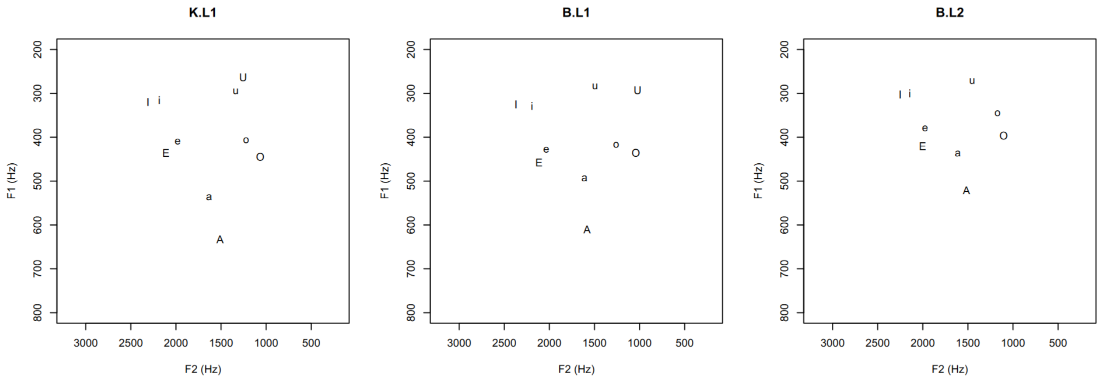


Figure 1. *Vowel quality of stressed (uppercase) and unstressed (lowercase) vowels: L1 Spanish in the tierras bajas (K.L1) and tierras altas (B.L1), and L2 Spanish in the tierras altas (B.L2).*

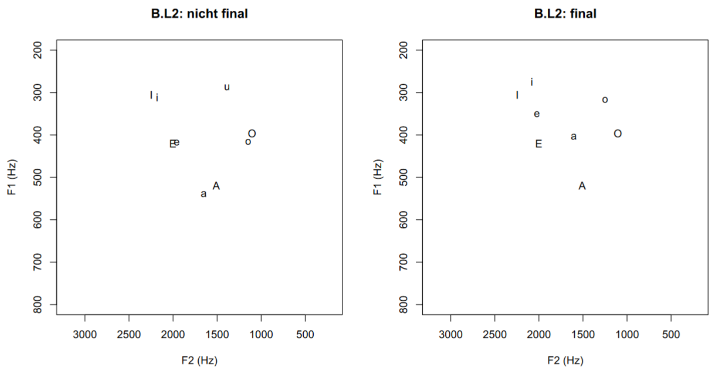


Figure 2. *Vowel quality of stressed (uppercase) and unstressed (lowercase) vowels in non-final (B.L2: nicht final) and final positions (B.L2: final) in L2 Spanish in the tierras altas*

[1] Almeida, M. 1990. El timbre vocálico en español actual. *Filología Románica* 7, 75–85.
[2] Boersma, P., & Weenink, D. 2018. *Praat: doing phonetics by computer*. Computer program.
[3] Canellada, M. & Zamora Vicente, A. 1960. Vocales caducas en el español mexicano. *Nueva Revista de Folología Hispánica* 14, 222–241.
[4] de Crignis, P. 2018. *Vokalschwächung im peruanischen Spanisch*. Munich: LMU München dissertation.
[5] Delforge, A.M. 2009. The Rise and Fall of Unstressed Vowel Reduction in the Spanish of Cusco, Peru: A Sociophonetic Study. Davis, CA: University of California at Davis dissertation.
[6] Hundley, J.E. 1983. *Linguistic Variation in Peruvian Spanish: Unstressed Vowel and /s/*. Ann Arbor: University Microfilms International.
[7] Ortega-Llebaria, M. & Prieto, P. 2010. Acoustic Correlates of Stress in Central Catalan and Castilian Spanish. *Language and Speech* 54(1), 73–97.
[8] R Development Core Team (2018): R. Computer program.
[9] Szczepaniak, R. 2009. Wortsprachliches Deutsch und silbensprachliches Spanisch. Ein phonologisch-typologischer Vergleich. *Estudios filológicos alemanes* 17, 251-267.
[10] Tcl Core Team (2018): The EMU Speech Database System. Computer program.
[11] Urban, M. 2019. Is there a Central Andean Linguistic Area? A View from the Perspective of the ‘Minor’ Languages. *Journal of Language Contact* 12, 271–304.
[12] Vaquero, M. & Guerra de la Fuente, L. 1992. Fonemas vocálicos de Puerto Rico. *Revista de Filología Española* 72(3/4), 555–582.

17. F0 variation in Italian-dialect bilingualism: close-up on Campanian varieties

Giovanni Leo, Giuseppe Magistro
Ghent University

Background: In this study, we focus on Italian-dialect bilingualism, which features the co-existence of a national official language (Italian) and Italo-romance dialects, whose contact gives rise to regionally/locally-tinted varieties of Italian [1] [2]. Italian and dialects co-exist in a "dilalic" situation, e.g., they are highly overlapped in informal speech [2], in other words the boundary between the high variety (Italian) and the low variety (dialects) is less-clear cut, also because of their structural similarity. This leads to the formation of an Italian-dialect continuum, resulting in "dialectisation" of Italian and "Italianisation" of dialects [1]. We focus on two linguistic areas in Campania (South of Italy), i.e., Naples and Cilento, respectively corresponding to an urban area and to a rural area.

Research questions/aims: Research has shown a link between F0 configurations and specific language varieties in contact, i.e., in sequential bilingualism [3] [4] [5]. Long-term distributional parameters are usually used to study F0 variation either across the whole utterance [4] or by targeting specific points in the utterance [5], showing that bilingual speakers exhibit F0 manipulations depending on the language being currently spoken. The extent of such manipulations is also reported to be linked to the degree of competence in the speaker's repertoire [5]. We set out to investigate **H1)** whether bilingual Italian-dialect speakers exhibit a manipulation of F0 properties sensitive to the variety being used, **H2)** whether the degree of these putative F0 manipulations (in terms of pitch span and level) depend on the different sociolinguistic properties of Naples (urban) and Cilento (rural).

Corpus & speakers: We rely on two linguistic groups: bilingual speakers (9 M) of Neapolitan (N) and Neapolitan Italian (NI), and bilingual speakers (6 M) of Cilentan and Cilentan Italian (CI), both featuring young men aged 19 to 30 with a technical-vocational education. Each group conducted comparable Discourse Completion Tasks (DCTs) in both varieties on a range of types of utterances, for a total of 1035 utterances in the Neapolitan group and 685 in the Cilentan group.

Method: By using *rPraat* [6], we extracted pitch span and pitch level over the entire utterance. These values have been normalised per speaker and scaled. In order to detect meaningful differences, a *k-nearest neighbours* algorithm was leveraged with the aim of classifying the observed data into the two varieties, more specifically creating a distinctive boundary accounting for the observed variation (cf. [7] for a likewise application of *kNN* within Italian-dialect prosody). This procedure was repeated for each linguistic group: N-NI, C-CI.

Results: As seen in Fig. 1, the dividing boundary is orthogonal to the F0 mean axis, indicating that the latter is a decisive feature for separating Italian and the respective dialect. For both linguistic groups, dialectal utterances are found with higher F0 means. At the same time, a closer inspection into the clusters reveals a more complex picture. While the Italian utterances of Naples are consistently found within the low F0 area (71%), the dialectal ones are more scattered along the entire F0 mean axis (54% for the high area and 46% for the low area). The high F0 mean of the Cilentan group is occupied by dialectal utterances, however the majority of the latter is also found in the low F0 mean, showing that both varieties consistently employ a lower level. More dynamicity is suggested in the Neapolitan group, which falls in line with what suggested by [8] with Neapolitan exhibiting phonetic enhancement of prominence **(H1)**. The rural area (Cilento) exhibits a more cohesive behaviour inasmuch as the use of different F0 levels seems not to be strongly correlated with the use of either variety **(H2)** (see Fig. 2 for a comparison across groups). Therefore, the different processes of dialect-to-Italian advergence [9] call for a deeper inquiry with the aim of unravelling the mapping of acoustic mechanisms and the sociolinguistic connotations of the areas.

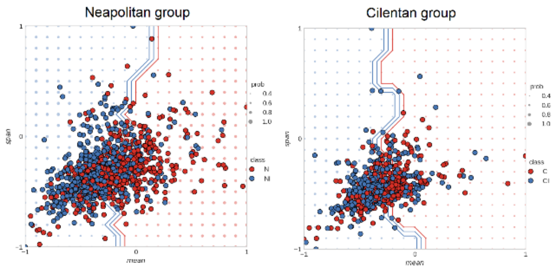


Figure 10. *kNN (Neapolitan Group on the left, Cilentan Group on the right).*

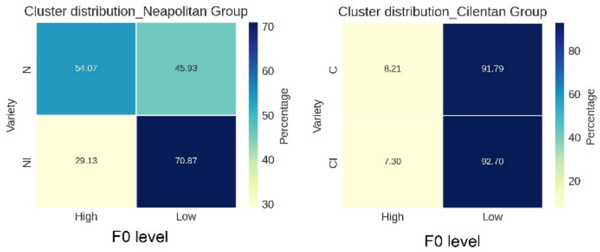


Figure 11 . *Cluster distribution per variety (Neapolitan Group on the left, Cilentan Group on the right).*

[1] Telmon, Tullio. 1993. 'Varietà Regionali'. In *Introduzione All'italiano Contemporaneo. La Variazione e Gli Usi*, 93–149. Laterza.

[2] Berruto, Gaetano. 2008. 'Le Varietà Del Repertorio'. In *Introduzione All'italiano Contemporaneo*, 3–36. Editori Laterza.

[3] Andreeva, Bistra, Grażyna Demenko, Bernd Möbius, Frank Zimmerer, Jeanin Jügler, and Magdalena Oleskowicz-Popiel. 2014. 'Differences of Pitch Profiles in Germanic and Slavic Languages'. In *Fifteenth Annual Conference of the International Speech Communication Association*.

[4] Meer, Philipp, and Robert Fuchs. 2019. 'Pitch Level, Range, and Dynamism in Trinidadian English: A Comparative Study with Other Varieties of English'. In *Proceedings of ICPhS*, 19:329–33.

[5] Cataldo, Violetta and Orrico, Riccardo and Savy, Renata. 2017. 'Phonetic Variation of f 0 Range in L1 and L2 : A Comparison between Italian, English and Spanish Native and Non-Native Speakers'. In *SOCIAL AND BIOLOGICAL FACTORS IN SPEECH VARIATION*, edited by Bertini, Chiara and Celata, Chiara and Lenoci, Giovanna and Meluzzi, Chiara and Ricci, Irene, 3:227–56. Studi AISV. Pisa: Officinaventuno. {http://doi.org/10.17469/O2103AISV000012}.

[6] Bořil, Tomáš, and Radek Skarnitzl. 2016. 'Tools rPraat and mPraat: Interfacing Phonetic Analyses with Signal Processing'. In *Text, Speech, and Dialogue: 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016, Proceedings 19*, 367–74. Springer.

[7]Crocco, Claudia, G Magistro, B Gili Fivela. 2022. 'Comparing Dialectal and Italian Prosody: The Case of Venetian'. *STUDI AISV*, 51–74.

[8] Crocco, Claudia, Barbara Gili Fivela, and Mariapaola D'Imperio. 2022. 'Comparing Prosody of Italian Varieties and Dialects: Data from Neapolitan'. In *Eleventh International Conference on Speech Prosody (Speech Prosody 2022)*, 140–44. International Speech Communication Association (ISCA).

[9] Auer, Peter. 2011. 'Dialect vs. Standard: A Typology of Scenarios in Europe'. In *The Languages and Linguistics of Europe*, 485–500. De Gruyter Mouton.

18. Differences in peak alignment between Central Castilian Spanish and Basque Spanish in declarative utterances in conversational speech

Gorka Elordieta[1], Magdalena Romera[2], Asier Illaro[1]
[1]Universidad del País Vasco (UPV/EHU), [2]Universidad Pública de Navarra

The present study aims to contribute to the growing work on prosodic aspects of languages or language varieties in contact, where prosodic features of one language or language variety are present in the other. It focuses on the variety of Spanish in contact with Basque (Basque Spanish), in parts of the provinces of Bizkaia, Gipuzkoa and Navarre, in northern Spain. [1]-[3] analyze the nuclear contours in conversational speech of information-seeking absolute interrogatives in Basque Spanish, showing that the prevailing contour is rising-falling: L+(¡)H* (H)L% (a rising nuclear pitch accent that can be upstepped, followed by a L% or HL% boundary tone). This type of contour is found in Basque, and contrasts with the rising final contour L* H% reported for Castilian Spanish in read speech (cf. references in [1]-[3]). [4] also finds a majority of rising final contours in Madrid Spanish conversational speech. [1]-[3] attribute the falling contour of Basque Spanish polar interrogatives to influence from Basque.

As for declaratives, [5] showed that in declarative utterances of speakers from the province of Gipuzkoa the majority of rising pitch accents had their peaks strictly aligned within the tonic syllable, L+H*. This contrasts with Castilian Spanish, where rising pitch accents have delayed peaks, L+<H* [6-9]. In Basque, the most common prenuclear and nuclear pitch accent in conversational speech is L+H* (cf. [10], [11]). Thus, a plausible hypothesis is that Basque L+H* is being transferred to Basque Spanish.

In the present study, we analyze the pitch accents in declarative utterances of conversational speech of 12 speakers from the province of Bizkaia, 6 from a city (Bilbao) and 6 from a small town (Lekeitio). They add to the 12 speakers from Gipuzkoa analyzed in [5], also from a city (San Sebastian) and a smaller town (Ibarra), for a total of 24 speakers (12 from two cities, and 12 from two small towns). The interest of considering urban and non-urban areas is that previous work on information-seeking polar questions in Basque Spanish has shown that there is a higher degree of presence of the Basque type of contour in areas where Basque is used in daily life consistently, compared to urban areas, where the use of Basque is very reduced (cf. [2]). 886 declarative utterances with 4,621 pitch accents have been analyzed, and compared to 1,051 pitch accents from 210 declarative utterances of 7 speakers of Central Castilian Spanish (from the city of Madrid) analyzed in [5]. The experimental methodology is the same for Basque Spanish and Madrid Spanish, so the results are directly comparable. We divide words in two types, depending on their information status: those that convey background or known information, and those that convey relevant information, responding to what was asked by the interlocutor. We call the accents in the two types *prenuclear* and *nuclear*, respectively (although the parallelism with the terms used in the literature on read speech is not exact). Three ToBI transcribers annotated the pitch accents and boundary tones (85% agreement).

Four separate logistic regression models were computed using the glmer function from the lme4 package [12] in RStudio [13] for the four most frequent tones: L+H*, L+<H*, H*, and L*. The results plotted in Figure 1 show that L+H* has a substantially higher probability of occurrence in non-urban and urban varieties of Basque Spanish compared to Madrid Spanish, with a higher probability in non-urban varieties. Since L+H* is the dominant pitch accent in Basque, the results would be compatible with the hypothesis that the use of Basque-like intonational features in Spanish is higher in areas where Basque is more present. Conversely, the predicted probability of L+<H* is much higher in Madrid Spanish. Finally, the predicted probability of L*is higher in Madrid Spanish than in Basque Spanish. The differences are less clear for H*.

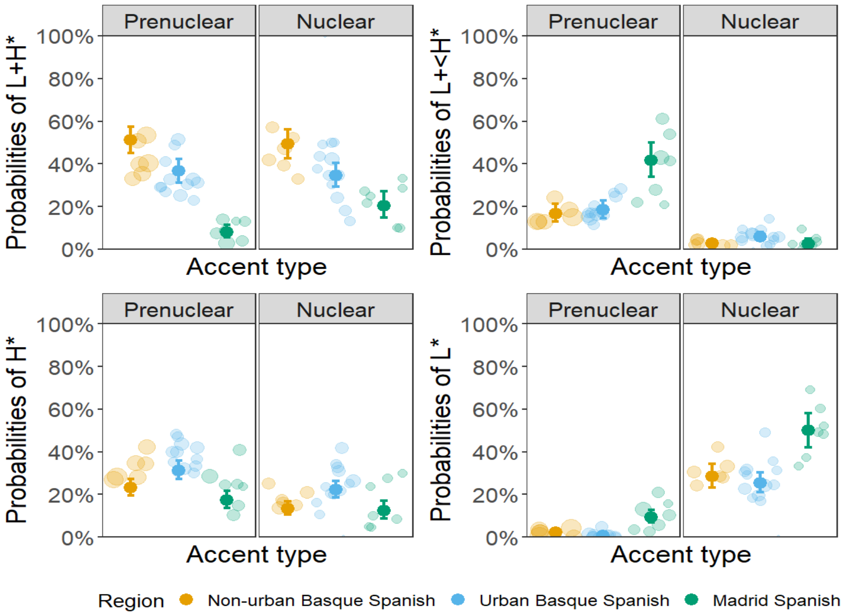


Figure 1. *Probabilities of pitch accents by accent type and region. Full dots indicate model estimates with 95% CIs, and lighter dots indicate individual speaker means.*

[1] Elordieta, G. & Romera, M. 2021. The influence of social factors on the prosody of Spanish in contact with Basque. *International Journal of Bilingualism* 25, 286–317.

[2] Romera, M. & Elordieta, G. 2020. Information-seeking question intonation in Basque Spanish and its correlation with degree of contact and language attitudes. *Languages* 5, 70.

[3] Elordieta, G. & Romera, M. 2024. Contornos ascendentes-descendentes en el castellano del País Vasco y Navarra. In Elvira-García, W. & Roseano, P. (Eds.), *Avances metodológicos en fonética y prosodia*. Madrid: UNED, 85-96.

[4] Elordieta, G. & Romera, M. 2020. The intonation of information-seeking absolute interrogatives in Madrid Spanish. *Estudios de Fonética Experimental* XXIX, 195-213.

[5] Elordieta, G., Romera, M. & Illaro, A. 2024. Pitch accents in Basque Spanish declarative utterances. *Proceedings of Speech Prosody 2024* (Leiden, The Netherlands), 627-631.

[6] Sosa, J.M. 1999. *La entonación del español*. Madrid: Cátedra.

[7] Face, T. 2008. The intonation of Castilian Spanish declaratives and absolute interrogatives. Munich: Lincom Europa.

[8] Estebas-Vilaplana, E. & Prieto, P. 2010. Castilian Spanish intonation. In Prieto, P. & Roseano, P. (Eds.), *Transcription of intonation of the Spanish language*. Munich: Lincom Europa, 17-48.

[9] Hualde, J.I. & Prieto, P. 2015. Intonational variation in Spanish, European and American varieties. In Frota, S. & Prieto. P. (Eds.), *Intonation in Romance*. Oxford: Oxford University Press, 350-391.

[10] Igartua, J. 2023. Ibarrako euskararen azentuari eta intonazioari hurbilpen bat. Undergraduate thesis, University of the Basque Country.

[11] Ezama, M. 2023. Azpeitiko euskara: Azentuaren eta intonazioaren azterketarako ekarpen txiki bat. Undergraduate thesis, University of the Basque Country.

[12] Bates, D., Maechler, M., Bolker, B. & Walker, S. 2015. Fitting linear mixed-effects models using lme4 [R package]. *Journal of Statistical Software* 67, 1-48.

[13] R Core Team. 2024. *R: A language and environment for statistical computing* [software]. Viena: R Foundation for Statistical Computing. Available: https://www.r-pro-ject.org.

19. VC timing in CVC-CVCC pairs: The role of vowel type and morphology

Anisia Popescu[1], Ioana Chitoran[2]
[1]LISN, CNRS, Université Paris Saclay ; [2]LLF, CNRS, Université Paris Cité

The present study investigates duration and F2 dynamics in vowel-consonant (VC) sequences in American English singleton (CVC) – cluster (CVCC) coda pairs as a function of vowel quality and morphological composition. Syllable-level timing patterns have been the focus of many studies since they were first described in Browman and Goldstein's (1988) seminal work. The focus has been on onset coordination patterns, which exhibit language- and cluster-specific variability. Much less focus has been given to coda coordination patterns, which for the most part conform to the predicted local coordination (i.e. coda consonants are sequentially added to the rhyme, without exhibiting a global stability point). Evidence suggests variability in coda timing as well. Marin and Pouplier (2010) observed a diverging timing pattern in American English coda clusters involving the lateral consonant /l/, resembling the pattern found in onset clusters. Similarly, Katz (2012) reported compression effects (acoustic vowel shortening) in cluster compared to singleton codas involving laterals and rhotics, but not nasals or obstruents. Both studies highlight the divergent behaviour of laterals in coda clusters, that triggers vowel shortening compared to singletons. However, both studies focused solely on monomorphemic tokens with lax vowel nuclei. In the present study, we ask:

1. Is this effect sensitive to vowel type? Data involving glides (Hsieh, 2017) and TaDA (Nam et al., 2004) simulations involving liquids (Popescu, 2019) suggest that vowel quality might have an influence on coda timing patterns.
2. Does the morphological composition of coda clusters play a role on timing patterns?

We address these questions by comparing a broader range of vowel nuclei (lax vowels, tense vowels, diphthongs) and monomorphemic vs. dimorphemic words).

Methods. A total of 11 native speakers of American English (6F, 6M, mean age: 31;7) were recorded producing three repetitions of 78 singleton (C)VC1 – cluster (C)VC1C2 pairs. In each pair V was either a lax vowel (/ɪ/), a tense vowel (/iː/) or a diphthong (/aɪ/ or /eɪ/), C1 was either a lateral or a nasal, followed by an obstruent C2. Cluster tokens were either monomorphemic or dimorphemic (past tense verbs or plural nouns). Target words were embedded in varying 4 syllable carrier phrases. We measured duration and F2 dynamics. Raw duration measures were normalized by the speech rate of the corresponding carrier sentence. The primary duration measure was the ratio between the cluster and the singleton pairs (duration-$VC_{cluster}$ / duration-$VC_{singleton}$), which reflects the shortening of the VC sequence in clusters compared to singletons. Ratios close to 1 indicate less shortening in the cluster token. Differences in ratios as a function of coda type (lateral, nasal), nucleus type (lax V, tense V, diphthong) and morphological complexity (monomorphemic, past, plural morpheme) were analyzed using linear mixed-effects models. To analyze F2 dynamics, F2 values were extracted using FastTrack (Barreda, 2021), and a Generalized Additive Mixed Model (GAMM) was used to test for binary differences between singleton and cluster pairs across different vocalic contexts.

Predictions: Based on previous findings we expect VC shortening effects for laterals but not for nasals. Lax vowels are expected to trigger more shortening than tense vowels and diphthongs. We consider the role of morphological complexity as an exploratory study, without specific predictions.

Results (from one repetition blocks so far) show that VC shortening occurs in cluster vs. singleton tokens for both laterals and nasals. However, this shortening is modulated by both vowel quality and the morphological composition of the cluster tokens: VC shortening is greater in lax vowel contexts and for monomorphemic tokens compared to dimorphemic ones. F2 dynamics confirm this pattern suggesting coda coordination patterns are more complex and multi-faceted as previously understood.

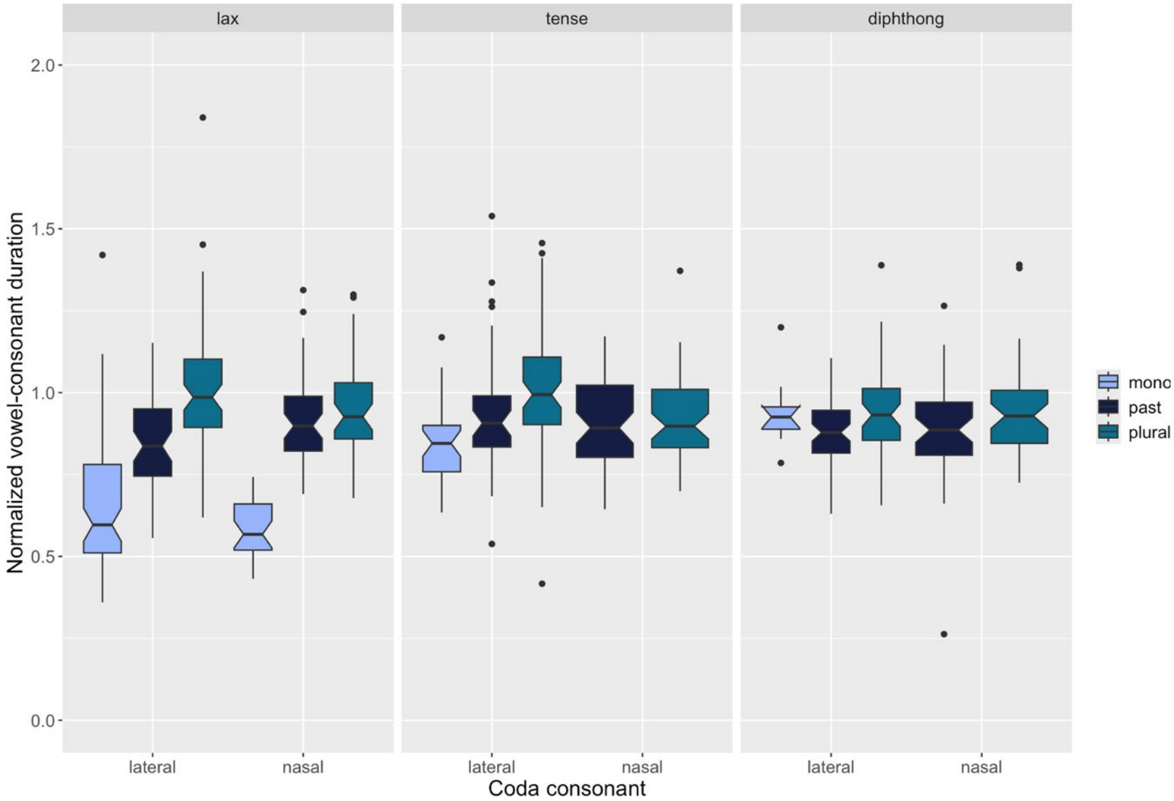


Figure 1. *Normalized Cluster-singleton VC duration ratio as a function of coda type (lateral, nasal), vowel nucleus type (lax V, tense V, diphthong) and morphological composition (mono-, past-, plural morpheme)*

[1] Barreda, S. (2021). Fast track: fast (nearly) automatic formant-tracking using praat. *Vanguard*, 7(1).

[2] Browman CP., & Goldstein L. (1988) Some notes on syllable structure in articulatory phonology. Phonetica; 45(2-4), 140-155

[3] Hsieh, F.-Y. (2017). A gestural approach to the phonological representation of English diphthongs. *PhD Dissertation*. University of Southern California.

[4] Katz, J. (2012). Compression effects in English. *Journal of Phonetics.* 40. 390-402.

[5] Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control, 14*(3), 380–407.

[6] Nam, H., Goldstein, L., Saltzman, E. & Byrd, D. (2004). TADA: AN enhanced, portable task dynamics model in MATLAB, *Journal of the Acoustical Society of America*. 115. 24-30.

[7] Popescu, A. (2019). Temporal organization of liquid consoantns in complex syllables: Implications for a dynamic articulatory model of the syllable. *PhD Dissertation.* University of Paris Diderot – Paris 7.

Poster session 2b

1. An exploratory study of phonetic accommodation of Cantonese sound change in human-human and human-AI interactions

Grace Wenling Cao[1], Ann To[2], Karen Ka Ki So[2] and Peggy Mok[2]
[1]University College Dublin [2] The Chinese University of Hong Kong

While previous studies found solid evidence of phonetic convergence in human-human interaction (Cao, 2024; Lin et al. 2021), Zellou and her colleagues start to navigate phonetic accommodation in human-AI interaction (Zellou et al. 2021) and found some evidence of humans converging towards AI on speech rate and F0. This study explores phonetic accommodation of human-AI interaction from the perspective of sound change with two implications. First, long-term sound change can be seen as the result of accumulation of short-term accommodation from the perspective of exemplar-based theories (Johnson, 2006; Pierrehumbert, 2003). Exploring how short-term accommodation works can shed light on long-term sound change. Second, the current theories of sound change are mainly developed based on human language input, it is unclear how well these theories can predict sound change in the new AI era when AI voice and speech also become part of input that trigger the changes. In this study, we are interested in the phonetic accommodation of human-human and human-AI interactions in Cantonese and how it might affect the ongoing Cantonese sound change in Hong Kong (Cheung, 2023).

Ten features of Cantonese sound change were selected: (1) n-l merger, (2) initial-ng deletion, (3) ng-m merger, (4) [l]-hypercorrection, (5-6) two conditions of ng-n merging at the coda position, (7) palatalisation of [s], (8) [kw]-[k] merging before rounded vowels, (9) pinjam and (10) high falling of tone 1. Each feature was represented by a single character (e.g. [naam]-male) and a disyllabic word (e.g. [naam jan]-men). Two native Hong Kong Cantonese speakers produced target words in both a conservative citation form (e.g. [n] in the n-l merger), and a new variant representing the variant adopted in the ongoing sound change (e.g. [l] in the n-l merger).

Two pilot studies were conducted. Pilots 1 and 2 included 12 (6F + 6M) and 8 (4F + 4M) young Hong Kong Cantonese speakers respectively. Participants read out loud a wordlist as the baseline. Following the error-repair paradigm (Cohn & Zellow, 2021; Zellou et al. 2021), participants in both pilots were instructed to read out the target word shown on the screen. They would then hear pre-recorded feedback with the target word either pronounced in a conservative form or a new form. Both variants appeared once in the experiment. Participants were asked to repeat what they had just said. Participants who use the conservative variants at the beginning were expected to converge towards new variants after they are primed with the new variants in the feedback. The human condition featured a human image or video during feedback, while the AI condition used a cartoon avatar. The experiment also tested the impact of visual cues, with Pilot 1 using static images and Pilot 2 using videos.

For data analysis, three native Cantonese speakers listened to each token of responses after the feedback and judged whether that was a convergence or divergence. The percentage of convergence and divergence in each condition was calculated and shown in Table 1. Greater accommodation, either a convergence or a divergence, occurred in the avatar condition during video presentations, while human-human interactions showed more change with static images. Convergence was higher in the video conditions than the image conditions. The ranking of accommodation across the ten features remained consistent across both conditions (from high to low degree): ng-m merger > ng-n merging at the coda > n-l merger > initial ng-deletion. It seems nasal features are more likely to change in phonetic accommodation. No change was found for tonal features. The preliminary results of these two pilot studies revealed the complexity of phonetic accommodation in human-human and human-avatar interactions when different visual cues were presented. A hierarchy of Cantonese sound change variants was found in phonetic accommodation.

		Image		Video	
		Converge	**Diverge**	**Converge**	**Diverge**
Single	Human	2.93%	2.34%	3.13%	0.39%
	Avatar	0.39%	1.17%	11.72%	7.81%
Disyllabic	Human	1.72%	1.56%	3.13%	3.13%
	Avatar	0.63%	0.00%	6.56%	3.44%

Table 1. *Percentage of accommodation.*

Cao, GW. (2024). Phonetic dissimilarity and L2 category formation in L2 accommodation. *Language and Speech*, 67 (2), 301-345.

Cheung, H. (2023). *Cantonese: since the 19th century.* The Chinese University of Hong Kong Press.

Cohn, M., & Zellou, G. (2021). Prosodic differences in human- and Alexa-directed speech, but similar local intelligibility adjustments. *Frontiers in Communication, 6*, 675704.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4), 485–499.

Lin, YH., Yao, Y. & Luo, J. (2021). Phonetic accommodation of tone: reversing a tone merger-in-progress via imitation. *Journal of Phonetics.* 87, 101060.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2–3), 115–154.

Zellou, G., Cohn, M., & Kline, T. (2021). The influence of conversational role on phonetic alignment toward voice-AI and human interlocutors. *Language, Cognition and Neuroscience, 36*(10), 1298–1312. https://doi.org/10.1080/23273798.2021.1931372

2. The Quiet Influencer: Investigating the Effect of Tajweed on the Acoustic Correlates of Pharyngealized /sˤ/ in Ammani Jordanian Arabic

Amenah Salman
University of South Carolina

This study investigates the effect of Tajweed on the acoustic correlates of the pharyngealized fricative /sˤ/ in Ammani Jordanian Arabic (AJA). Tajweed is the study of the phonetic and phonological rules of Arabic, typically learned for the prescriptive recitation of the Quran. Given the similarities between Tajweed and generative phonology in terms of features, sound classifications, and phonological processes [1], this study explores Tajweed not merely as a measure of religiosity but as a social phonetic training that influences language variation.

Pharyngealization is one of many phonological processes that Tajweed emphasizes. The sound /sˤ/, like other pharyngealized sounds in AJA, has two articulation points: a primary one in the anterior vocal tract [2], and a secondary one marked by a "retracted tongue root" (RTR), adding emphasis that spreads to adjacent vowels [3]. /sˤ/, as in صام /sˤaːm/ 'he fasted,' contrasts phonemically with its plain counterpart /s/, as in سام /saːm/ 'poisonous,' and is considered the most frequent pharyngealized sound in Arabic and in the Quran [4, 5]. Despite its importance, /sˤ/ exhibits considerable variation in AJA, with some speakers using the expected pharyngealized variant while others produce a weakened, lenited-pharyngealized (LP) variant [6].

Previous studies have shown that the first three formant frequencies (F1, F2, and F3) of adjacent vowels can serve as reliable acoustic cues for distinguishing between pharyngealized and plain sounds in JA, with pharyngealized sounds exhibiting higher F1 and F3 and lower F2 values than plain consonants [3]. While prior research has explored the influence of factors such as gender and age [3, 7] on these cues, the potential impact of Tajweed has not been studied. However, Tajweed has been found to have a statistically significant effect on the lenition of /sˤ/ ($p < 0.001$), with higher levels of Tajweed proficiency associated with a decreased likelihood of lenition [6]. Given this finding and the prevalence of Tajweed courses in Jordan, it is reasonable to predict that Tajweed may have an effect on the target cues.

The current study comprises 24 upper-class AJA speakers from West Amman, balanced for gender and Tajweed proficiency level (L0–L3), ranging in age from 18 to 72. Data were collected from 30-minute Zoom interviews, yielding 420 word-initial /sˤ/ tokens from 12 hours of free speech. Formant measurements were taken 10% into the following vowel in Praat [8], and formant values were normalized using the Nearey1 method. A linear mixed-effects model revealed a significant main effect for Tajweed ($p < 0.001$), with increased F1 and F3 values and decreased F2 values corresponding to higher Tajweed levels, indicating greater tongue retraction and stronger pharyngealization [9] with each Tajweed level. This is consistent with prior findings that pharyngealized sounds exhibit higher F1 and F3 and lower F2 than their plain counterparts [3]. No main effects were found for gender and age, but women and younger speakers were found to have lower F1 and F3 and higher F2 values compared to men and older speakers, signaling lenition or a weaker articulation

These findings support Salman and Dalola's assertion that analyses of sound variation in Amman, Jordan, should incorporate Tajweed as a locally defined social force driving sound variation [6]. This research highlights the importance of incorporating culturally and locally defined social categories, such as Tajweed, into sociophonetic studies to fully capture the dynamics of language variation and change within each society.

[1] Safaie, E. (2008). Muslim's contributions to modern linguistics: The descriptive framework of Tajweed of the Holy Qur'an and the generative phonology in contrast. Proceedings

from the International Language Conference 2008, International Islamic University Malaysia.
[2] Card, E. A. (1983). A phonetic and phonological study of Arabic emphasis. Cornell University.
[3] Alzoubi, A. A. S. (2017). The effect of social factors on emphatic-plain contrast in Jordan: A sociophonetic study of Arabic in Amman City (Doctoral dissertation). University of Utah.
[4] Hasan, M. R., & Mamun, F. E. (2022). Numerical statements of the Holy Quran based on verses, words, and letters in Makki and Madani Sura. International Journal of Cultural and Religious Studies, 2(2), 24-33.
[5] Madi, A. (2010). A study of Arabic letter frequency analysis. Intellaren. Retrieved from http://www.intellaren.com/articles/en/a-study-of-arabic-letter-frequency-analysis.
[6] Salman, A., & Dalola, A. (2024). Elusive Pharyngeals: Exploring social variation of pharyngealized /sˤ/ production in Jordanian Arabic. NWAV 52 Abstracts.
[7] Abudalbuh, M. (2010). Effects of gender on the production of emphasis in Jordanian Arabic: A sociophonetic study (Doctoral dissertation, University of Kansas).
[8] Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer [Computer program]. Version 6.2, retrieved 15 Nov 2021 from http://www.praat.org/.
[9] Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. Laboratory Phonology, 8(1), 28.

3. Social Salience and Dialect Contact in Los Angeles Spanish

Jack Carter, Ji Young Kim, Jenny Choi, Jesús Duarte, Joo Kyeong Kim
University of California, Los Angeles

This study explores the interplay of social salience and dialect contact in Spanish spoken in Los Angeles where Mexicans and Central Americans of the Northern Triangle (NT; Guatemala, El Salvador, and Honduras) constitute the majority of the Latinx population. In dialect contact settings, high salience in lower prestige variants has been implicated to favor leveling (i.e., loss of lower prestige variants), while low salience has been shown to favor convergence (i.e., increasing similarities across groups) [5]. We investigate whether and to what extent high and low salience NT Spanish features – coda /s/-aspiration and word-final /n/-velarization, respectively [9] – demonstrate dialect leveling and convergence in Los Angeles Spanish. Moreover, we explore the role of superstrate language by examining NT dialectal features that resemble sounds in English (i.e., /x/-glottalization and /ɟ/-approximantization).

Four groups of Spanish speakers – first-generation Mexican (M1), second-generation Mexican (M2), first-generation NT (NT1), and second-generation NT (NT2) – participated in this study. Speech data collection was carried out via Zoom and participants' speech was recorded on their computer using Audacity [1]. We extracted the target phones in Praat [2] and analyzed the acoustic cues to place or manner distinction associated with the dialectal features in question: formant transition (F3-F2) for /n/-velarization [10, 11], center of gravity (COG) for /s/-aspiration and /x/-glottalization [3, 4, 6, 7, 8], and cepstral peak prominence (CPP) for /ɟ/-approximantization [13]. We performed linear mixed-effects modeling in R [12] to test the effects of heritage and generation on each acoustic correlate.

Preliminary results of data obtained from 12 speakers (M1: 2, M2: 4, NT1: 2, NT2: 4; see Figure 1) showed that the first-generation speakers, particularly M1, produced /n/ with more alveolar-like realizations [11] ($\chi^2(1) = 4.467$, $p < 0.05$) than the second-generation speakers. In the case of /s/, across the speaker groups, the majority of the tokens were within the range of [s] [3, 4, 6, 7, 8]. Based on visual inspection of the data, NT1 were the ones that produced the most fronted /s/ realizations, although we did not find any effect of heritage or generation. With regard to the NT features that have English equivalents, the two features differed in that /ɟ/ demonstrated values resembling the Mexican variant [ɟ] [13], while /x/ showed values similar to the NT/English variant [h] [3, 7]. While no effect of heritage or generation was found in the production of /x/, there was an effect of generation on /ɟ/ realizations, in which the second-generation speakers produced more approximant [j]-like patterns ($\chi^2(1) =4.532$, $p < 0.05$).

Based on our findings, we conjecture that, regardless of heritage and immigrant generation, the speakers adhere to higher prestige variants when producing /n/ and /s/; lower prestige variants are rarely used and are even avoided, as suggested in the NT1 /s/ data. On the other hand, lower prestige features that do have English equivalents (i.e., /ɟ/-approximantization and /x/-glottalization) are more likely to survive, as English, a superstrate language, exerts an effect on dialect contact in Spanish, a substrate language. Speech data of 22 additional speakers (M1: 2, M2: 15, NT1: 2, NT2: 3) has been collected and more data collection is underway. With more data, a clearer cross-generational pattern is expected to emerge.

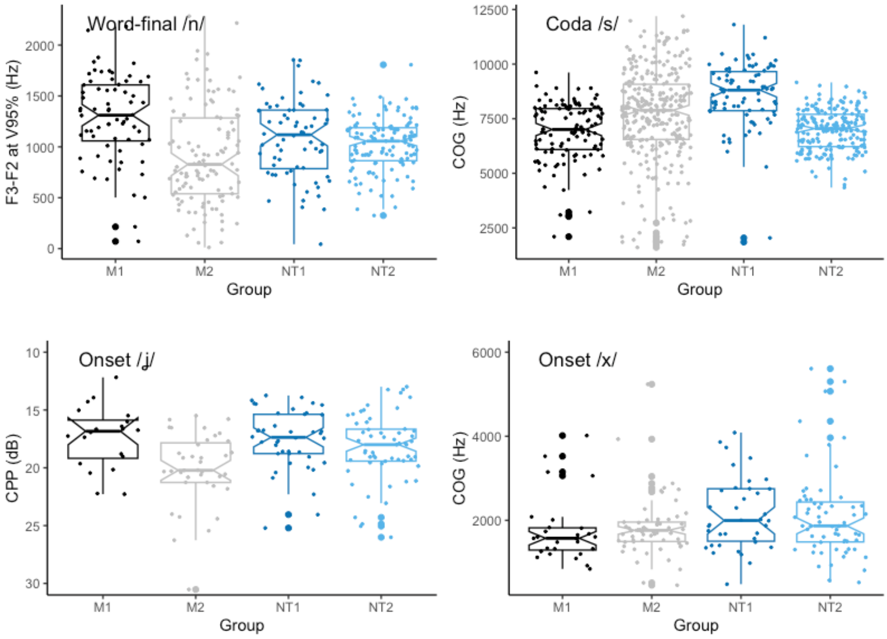


Figure 1. *Acoustic correlates of word-final /n/, coda /s/, onset /ʝ/, and onset /x/ (higher values indicate more Mexican-like features.*

[1] Audacity Team. (2021). *Audacity(R): Free Audio Editor and Recorder* [Computer application].

[2] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

[3] Brogan, F. D., & Bolyanatz, M. A. (2018). A sociophonetic account of onset/s/weakening in Salvadoran Spanish: Instrumental and segmental analyses. *Language Variation and Change*, *30*(2), 203–230.

[4] Erker, D. (2012). Of categories and continua: relating discrete and gradient properties of sociophonetic variation. *Penn Working Papers in Linguistics*, 18(2), 11–19.

[5] Erker, D. (2017). The limits of named language varieties and the role of social salience in dialectal contact: The case of Spanish in the United States. *Language and Linguistics Compass*, 11(1), e12232.

[6] File-Muriel, R., & Brown, E. (2011). The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change,* 23(2), 223–243.

[7] Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32(2), 141–174.

[8] Gradoville, M., Brown, E., & File-Muriel, R. (2022) The Phonetics of Sociophonetics: Validating Acoustic Approaches to Spanish /s/. *Journal of Phonetics*, 91(1), 101125.

[9] Hoffman, M. F. (2004). Sounding Salvadorean: Phonological variables in the Spanish of Salvadorean youth in Toronto. Ph.D. dissertation.

[10] Kent, R. D., & Read, C. (2002). *Acoustic Analysis of Speech* (2nd ed.). Thomson Learning.

[11] Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *The Journal of the Acoustical Society of America*, 72(2), 379–389.

[12] R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

[13] Repiso-Puigdelliura, G., Benvenuti, I., & Kim, J. Y. (2021). Heritage Speakers' Production of the Spanish Voiced Palatal Obstruent /ʝ/: A Closer Look at Orthography and Universal Phonetic Principles. *Heritage Language Journal*, 18(1), 1–30.

4. Effects of Lexical Stress on Static and Dynamic Properties of Ukrainian Vowels

Olga Maxwell, Adele Gregory, Marija Tabain
Univeristy of Melbourne

Ukrainian is an East Slavic language spoken widely in Ukraine and across the globe. Despite its large speaker base, limited instrumental work has been conducted on its phonetics, particularly the acoustic properties of Ukrainian vowels. Prior research on vowels has variously relied on auditory analysis (e.g., [1]), faced challenges in phonetic description of Ukrainian in terms of IPA due to inconsistencies across transcription systems in the literature, or was based on a few speakers. The IPA illustration for Ukrainian, for instance, is derived from a single speaker from Bukovina, Southwestern Ukraine, who received academic education in Kyiv [2]. However, Ukrainian exhibits significant regional variation across its three main dialect groups: Northern, Southwestern, and Southeastern [3]. Additionally, all descriptions of the vowel system in Ukrainian to date are based on static analyses (i.e., F1/F2 estimates measured at the midpoint). While static methods have been acknowledged for their efficacy, dynamic measures provide deeper insights and more robust evidence of temporal changes in vowel spectral characteristics [4].

To address these gaps, the present study investigates the spectral and durational characteristics of Ukrainian vowels, focusing on their realisation in stressed and unstressed syllables. Ukrainian is described as having six vowel phonemes: /i ɪ ɛ a ɔ u/, with no phonemic vowel length [2, 5]. Stress in Ukrainian is free and can occur on any syllable of a word, with unstressed vowels exhibiting a tendency to reduce to less peripheral articulations [2]. In contrast to [2], Danylenko and Vakulenko [5] suggest minimal perceivable reduction in both vowel quality and duration for most vowels, except /ɛ/, /ɪ/, and /ɔ/, noting that unstressed /ɛ/ and /ɪ/ are not distinguished phonetically, with "their pronunciation and acoustic properties halfway between the respective stressed vowels" (p.5). These two vowels are also reported to vary in unstressed contexts [6].

Speech data were obtained from twelve native Ukrainian speakers (9F, 3M, ages 19-32) who grew up speaking Ukrainian and predominantly use it in daily life. To control for regional variation, participants were recruited from Dnipro (southeastern dialectal group) and northern Ukraine (mostly Kyiv). The corpus comprises wordlist recordings made with a Zoom recorder by the first author and self-recordings made by participants on their phones, a method validated in phonetic studies [7]. Audio files were force-aligned using the WebMAUS [8] and manually corrected in Praat [9]. A database was created in the EMU system [10] for further analysis in R [11]. The analysis includes a) a static approach—F1 and F2 values extracted at vowel midpoints and linear mixed effect regression used to examine the vowel positions in the F1/F2 space; and b) a dynamic approach—formant values extracted at five points throughout the vowel (e.g., [12]) and generalised additive mixed models used to model F1/F2 trajectories [13].

Preliminary findings based on three speakers reveal that stressed vowels are longer than their unstressed forms. Stressed vowels /i, u, a/ occupy more peripheral positions in the F1/F2 vowel space in relation to their unstressed counterparts, with the low vowel showing the greatest magnitude of reduction. However, stressed /ɪ/ has a lower F2, indicating retraction relative to the more fronted unstressed /ɪ/. The vowel /ɛ/ shows substantial overlap between stressed and unstressed contexts, with its trajectory characterised by a downward shift (increase in F1) in the F1/F2 space. No dynamicity is observed for the back vowels /u/ and /ɔ/. The findings of our study have implications for the typology of vowel reduction across languages, demonstrating that while unstressed vowels in Ukrainian do not undergo phonological reduction [14], they show a degree of phonetic reduction, as evidenced by spectral and temporal changes—patterns also observed in other Slavic languages, such as Polish [15]. More importantly, closer examination of individual vowels reveals greater nuance.

[1] Zilyns'kyj, I. 1979. *A phonetic description of the Ukrainian language.* Cambridge, MA: Harvard University Press.
[2] Pompino-Marschall, B., Steriopolo, E. & Żygis, M. 2017. Ukrainian. *Journal of the International Phonetic Association*, *47*(3), 349–357. https://www.jstor.org/stable/26352393
[3] Del Gaudio, S. 2018. *An Introduction to Ukrainian Dialectology.* Wiener Slawistischer Almanach - Sonderbände, Volume 94. Peter Lang
[4] Cox, F., Penney, J. & Palethorpe, S. 2024. Australian English Monophthong Change across 50 Years: Static versus Dynamic Measures. *Languages*, *9*(3), 99. https://doi.org/10.3390/languages9030099
[5] Danylenko, A. & Vakulenko, S. 1995. *Ukrainian.* Munich: Lincom Europa.
[6] Pugh, S. & Press, I. 2005. *Ukrainian: A Comprehensive Grammar*. New York: Routledge.
[7] Zhang, C., Jepson, K., Lohfink, G. & Arvaniti A. 2021. Comparing acoustic analyses of speech data collected remotely. *The Journal of the Acoustical Society of America, 149*(6), 3910. doi: 10.1121/10.0005132.
[8] Kisler, T., Reichel, U. & Schiel F. 2017. Multilingual processing of speech via web services. *Computer Speech & Language, 45*, September 2017, 326–347.
[9] Boersma, P., & Weenink, D. 2024. *Praat: doing phonetics by computer*. Computer program.
[10] Winkelmann, R., Jänsch, K., Cassidy, S., and Harrington, J. 2021. emuR: Main package of the EMU Speech Database Management System [Version 1.4.1]. Available: https://cran.r-project.org
[11] R Core Team, 2021 R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
[12] Renwick, M. E. L. and Stanley, J. A. 2020. Modeling dynamic trajectories of front vowels in the American South. *The Journal of the Acoustical Society of America*, *147*(1), 579–595, 2020.
[13] Sóskuthy, M., 2021. Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics, 84*, 101017. https://doi.org/10.1016/j.wocn.2020.101017
[14] Gordon, M. 2011. Stress: Phonotactic and phonetic evidence. *The Blackwell companion to phonology*, *2*, 924-948.
[15] Rojczyk, A. 2019. Quality and duration of unstressed vowels in Polish. *Lingua, 117*, 80-89. https://doi.org/10.1016/j.lingua.2018.10.012

5. Correlation between vowel dispersion and the size of vowel inventories

Giovanna Di Paola
Universidad de Sevilla

Languages vary in the size of their phonological inventories and in the distribution in the vocalic space. Some authors maintain that a larger vowel inventory implies a larger acoustic space and a greater vowel dispersion [1] [2] [3], and vowel variability is inversely correlated with the inventory size, because distances between vowels in the acoustic space decrease with the increase of the inventory size [1] [4]. Others sustain that the inverse correlation between vowel dispersion per vowel type and the inventory size is confirmed in part because it could depend on the condition of Word, Isolation or Syllable [3] or not confirmed because, in the case of some languages, vowel variability is not inversely correlated with the inventory size [5].

This study aims to compare the correlation between vowel dispersion and the size of vowel inventories. We assume that greater vowel variability implies a wider vowel space, therefore a greater vowel dispersion which should be associated to smaller vowel inventories.

To do so, we firstly selected vowels from 4 phonetic databases: UCLA Phonetics Lab Archive [6], PHOIBLE [7], The Illustrations of the IPA [8] and Vox Angeles Corpus [9]. The vowels were manually annotated in Praat (in the case of Vox Angeles Corpus using their available transcriptions) and we used Praat [10] to extract F1 and F2 data. Then, we normalized the data considering the same number of point vowels for each language (i, a, u) and computed their repulsive force [11]. To compute vowel dispersion, we used vowel's repulsive force [11] which is calculated in R studio. The repulsive force is defined as the inverse sum of the square distances between vowels and, given that different languages will have different number of vowels, we report the mean distance between vowels. Using this method, we analysed the vowel charts of 120 languages and 6433 vowels. Although, in the case of the correlation between vowel dispersion and the inventory size, our analysis is limited to selected languages of The Illustrations of the IPA [8]. Our results for these data suggest that there is not a correlation between vowel dispersion and the size of vowel inventories. Therefore, since these data are very preliminary, more languages need to be analysed and other measures of vowel dispersion could be employed, to enhance the discussion. We are currently calculating the correlation between vowel dispersion and vowel inventories of Vox Angeles dataset. Moreover, a further analysis will be carried out calculating dispersion of other vowels of the triangle instead of point vowels, and the correlation with inventory size.

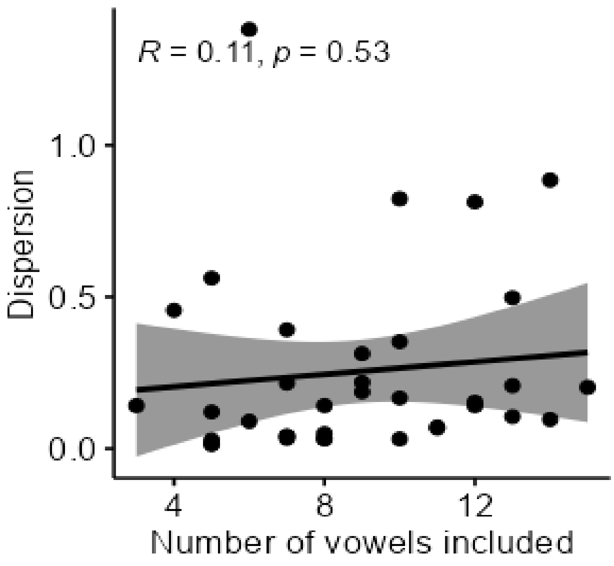


Figure 1. *Correlation between vowel dispersion and the inventory size based on Illustrations of the IPA.*

[1] Lindblom, B. 1986. "Phonetic universals in vowel systems". In Experimental Phonology, J.J. Ohala and J.J. Jaeger (eds.), p.13-44. New York, Academic Press.

[2] Neel, A. T. 2008. Vowel space characteristics and vowel identification accuracy.

[3] Al-Tamimi, J. E., & Ferragne, E. 2005. Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *Interspeech* (pp. 2465-2468).

[4] Engstrand, O., & Krull, D. 1991. Effect of inventory size on the distribution of vowels in the formant space: preliminary data for seven languages. *PERILUS*, *13*, 15-18.

[5] Lee, W. S. 2011. Dispersion and variability of vowels of different vowel inventory sizes. In *Proceedings of Meetings on Acoustics* (Vol. 14, No. 1). AIP Publishing.

[6] Ladefoged, P., Blankenship, B., Schuh, R. G., Jones, P., Gfroerer, N., Griffiths, E., Harrington, L., Hipp, C., Jones, P., Kaneko, M., Moore-Cantwell, C., Oh, G., Pfister, K., Vaughan, K., Videc, R., Weismuller, S., Weiss, S., White, J., Conlon, S., Lee, WS. J., and Toribio, R. 2009. The UCLA Phonetics Lab Archive. Los Angeles, CA: UCLA Department of Linguistics. http://archive.phonetics.ucla.edu.

[7] Moran, Steven & McCloy, Daniel (eds.) 2019. PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History. (Available online at http://phoible.org, Accessed on 2024-10-04.)

[8] Baird, L., Evans, N., & Greenhill, S. J. 2022. Blowing in the wind: Using 'North Wind and the Sun'texts to sample phoneme inventories. *Journal of the International Phonetic Association*, *52*(3), 453-494.

[9] Chodroff, E., Pažon, B., Baker, A., & Moran, S. 2024. Phonetic Segmentation of the UCLA Phonetics Lab Archive. *arXiv preprint arXiv:2403.19509*.

[10] Boersma, P., & Weenink, D. 2008. *Praat: doing phonetics by computer*. Computer program

[11] Wright, R. 1997. Lexical competition and reduction in speech: A preliminary report. *Research on spoken language processing progress report*, *2*.

6. Relating the effect of beat gestures in second language stress perception to individual differences in working memory

Luise Böttcher, Marieke Hoetjes, Lieke van Maastricht
Centre for Language Studies, Radboud University Nijmegen

Acquiring the prosody of a second language (L2), i.e., its rhythm, intonation, and stress, is challenging for learners in both production and perception [1]. Determining which syllable is the most prominent in a word (i.e., where the lexical stress is) is essential for speech comprehension because it affects the deactivation of competitor items in the mental lexicon [2-3] and because minimal lexical pairs can result from different stress patterns [1]. For example, Spanish verb forms of the first person singular present tense and the third person singular preterit can only be distinguished by their lexical stress, e.g., 'aCEPto' ('I accept') vs. 'acepTÓ' ('(s)he accepted'), capital letters indicating lexical stress. A plausible approach to support learners in acquiring L2 stress is using beat gestures because they serve as visual stress cues and increase the perceived prominence of syllables, at least in the native language (L1) [4-5]. There is little research on the benefits of gestures in L2 prosody perception, and findings for L2 production are inconclusive: Some studies found that L2 learners benefited from beat gestures, while others did not, e.g. [6-8]. These diverging findings may stem from the fact that beat gestures do not universally benefit all learners.

Individual differences in working memory (WM) might affect the benefit of gestures [9-10]. However, there are different components of WM, namely phonological, visuospatial, and executive working memory [11], but no study to date has attempted to disentangle their effects on multimodal L2 stress perception. Arguably, better phonological and executive WM may benefit L2 acquisition in general [11], while visuospatial WM may be especially important in multimodal gesture-speech processing [9-10]. This study relates the effects of beat gestures on L2 stress perception to individual differences in three different WM components. In the present study, 24 Dutch natives performed an L2 Stress Perception Task in which they saw videos of a Spanish native speaker producing Spanish verbs, with or without a beat gesture co-occurring with the stressed syllable. After a short audiovisual distraction task that should place additional load on the visuospatial and phonological WM, participants indicated which syllable in the word that they initially had heard was stressed (cf. Figure 1). Both reaction times and accuracy scores were recorded. Participants also completed additional standard tests for visuospatial, phonological, and executive WM.

Mixed-effects modeling of participants' accuracy and reaction times in identifying stressed syllables revealed that participants did not perform better in the gesture condition than in the no gesture condition, a surprising finding considering prior L1 research. In line with earlier research on the isolated role of phonological WM in L2 acquisition [11], participants with better phonological WM correctly identified more stressed syllables. However, no main effects of executive and visuospatial WM were found on accuracy nor reaction times, nor were there signification interactions of any of the three WM components with gesture condition. This aligns with the variability observed in previous L2 studies regarding the benefit of beat gestures [6-8] and the individual differences in WM reported in [10], for example. The present study's added value lies in highlighting the importance of incorporating measures of cognitive individual differences into experimental designs. The results suggest that such measures may have a greater impact on outcomes than the gesture effect itself, providing a critical direction for future L2 research on multimodal stress perception.

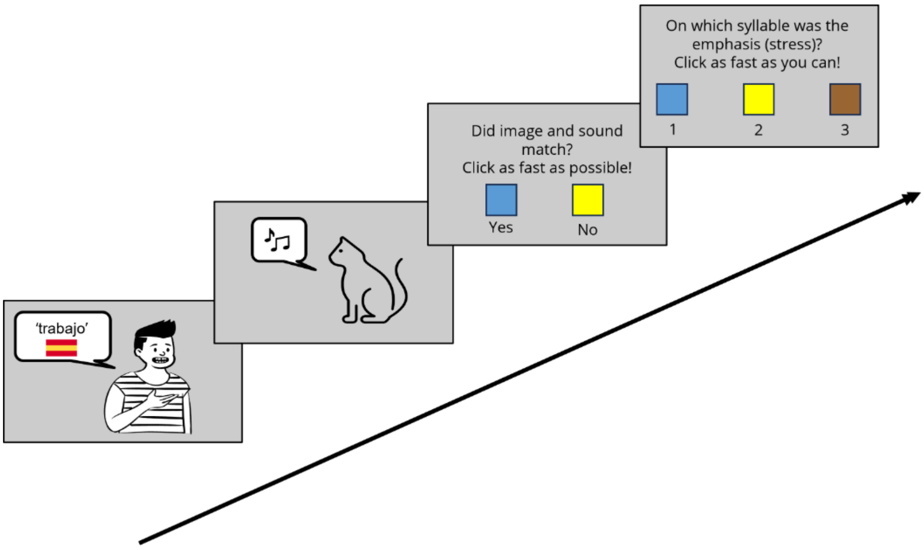


Figure 1. *L2 Stress Perception Task*

[1] Zhu, Y., & Mok, P. (2022). The role of prosody across languages. In T. M. Derwing, M. J. Munro & R. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking*. Routledge Taylor & Francis Group.

[2] Jesse, A., & McQueen, J. M. (2014). Suprasegmental lexical stress cues in visual speech can guide spoken-word recognition. *Quarterly Journal of Experimental Psychology* (2006), 67(4), 793–808.

[3] Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology (2006)*, 63(4), 772–783.

[4] Bosker, H. R., & Peeters, D. (2021). Beat gestures influence which speech sounds you hear. *Proceedings of the Royal Society B*, 288(1943), 20202419.

[5] Rohrer, P. L., Bujok, R., van Maastricht, L., & Bosker, H. R. (2024). The timing of beat gestures affects lexical stress perception in Spanish. *Proceedings of Speech Prosody 2024*.

[6] Llanes-Coromina, J., Prieto, P., & Rohrer, P. L. (2018). Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task. *Speech Prosody 2018*, 498–502.

[7] van Maastricht, L., Hoetjes, M., & van Drie, E. (2019). Do gestures during training facilitate L2 lexical stress acquisition by Dutch learners of Spanish? *Proceedings of the 15th International Conference on Auditory-Visual Speech Processing (AVSP2019)*.

[8] van der Heijden, L. (2021). *Gestures in learning foreign language prosody: The importance of considering task and learner characteristics* [Master's Thesis]. Radboud University.

[9] Özer, D., & Göksun, T. (2020). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Language, Cognition and Neuroscience*, 35(7), 896–914.

[10] Wu, Y. C., & Coulson, S. (2014). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychologica*, 153, 39–50.

7. The interplay between prosody and gestures to process phrasal ambiguities in children with Developmental Language Disorder

Albert Giberga[1], Ernesto Guerra[2], Nadia Ahufinger[1], Alfonso Igualada[1], Mari Aguilera[3], Núria Esteve-Gibert[1]
[1]*Estudis de Psicologia i Ciències de l'Educació, Universitat Oberta de Catalunya, Barcelona, Spain,* [2]*Centro de Investigación Avanzada en Educación, Universidad de Chile, Santiago, Chile,* [3]*Departament de Cognició, Desenvolupament i Psicologia de l'Educació, Secció Cognició, Facultat de Psicologia, Universitat de Barcelona, Barcelona, Spain*

This study explores the role of prosodic and multimodal cues in helping children with Developmental Language Disorder (DLD) interpret phrasally ambiguous sentences. Prosody is found to aid adult's resolution of phrasal ambiguities ([1]), but previous results on when children are able to do that were not conclusive ([2]). Moreover, co-speech gestures further facilitate this process in adults ([3]), and this gestural boosting effect might be even more crucial in development. This study examines whether combining prosodic and gestural cues provides additional support to children with DLD, who often face challenges in structural language processing and with phonology, and for whom the additional support might be even more crucial. Seventy-nine Catalan-speaking children (34 with DLD, 45 Typically Developing) aged 5-10 years participated in the study. Children were divided into two age subgroups: a younger group (age range 5:0–7:11; average 6:2) and an older group (age range 8:0–10:11; average 8:9). Using a visual-world eye-tracking paradigm, children were exposed to phrasally ambiguous sentences such as *La dona balla amb la nena descalça* ('the woman dances with the girl barefoot'), which Catalan listeners preferently interpret as the girl being barefoot but which has an alternative less-frequent high-attachment interpretation (the woman being barefoot) ([4]). Sentences were presented in three conditions: baseline (prosodic cues to default interpretation: [La dona balla]L+H*H- [amb la nena descalça]L+H*H%), prosody-only (prosodic breaks cues to less frequent high-attachment interpretation: [La dona balla amb la nena]L+H*H- [descalça]L+H*H%), and multimodal (same prosodic pattern as in the prosody-only condition signaling the less frequent high-attachment interpretation, but accompanied by a manual beat gesture and a head nod aligned with the prosodic breaks). We analyzed the proportion of trials in which participants selected offline the less-frequent high-attachment interpretation using Generalized Linear Mixed Models (GLMMs), as well as their gaze patterns during the task. The model included Experimental Condition (baseline; prosody-only; multimodal), Linguistic Group (DLD; TD), and Age Group (younger; older) as fixed factors, with random effects for Participants and Items. Offline responses (Figure 1) revealed both DLD and TD children used prosodic cues to interpret the high-attachment meaning ($z = 2.921$, $p < 0.01$), and gestures did not further contribute to enhancing this alternative interpretation ($z = -0.355$, $p = 0.72$). All materials and detailed results are on OSF. Online results (see Figure 2) showed that, after the end of the sentence, the multimodal condition led to a significant increase in looks toward the high-attachment interpretation in the younger subgroups (both TD and DLD) and the older DLD group (sum t TD = 36.5; sum t DLD = 26.5 for younger, and sum t = 57.3 for older DLD). No such effect of multimodality was observed in older TD children (sum t = 0). Notably, TD children temporarily fixated more on the speaker in the center of the screen upon the unfolding of gestural cues indicating the alternative phrasal structure (sum t = 149.8), compared to children with DLD (sum t = 23.6). The fact that TD children briefly fixated on the speaker when gestural cues were present, whereas children with DLD did not show this effect, could indicate that children with DLD may experience difficulties in processing multimodal information to phrasal meaning. These findings provide insights into how prosodic and multimodal cues influence language comprehension in children with DLD and suggest that prosodic information is sufficient to guide the children's interpretation of phrasal structures, independently of whether they face difficulties with structural language, that gestures might have a limited contribution to sentence comprehension, and that we need to investigate further how children with DLD process online multimodal information.

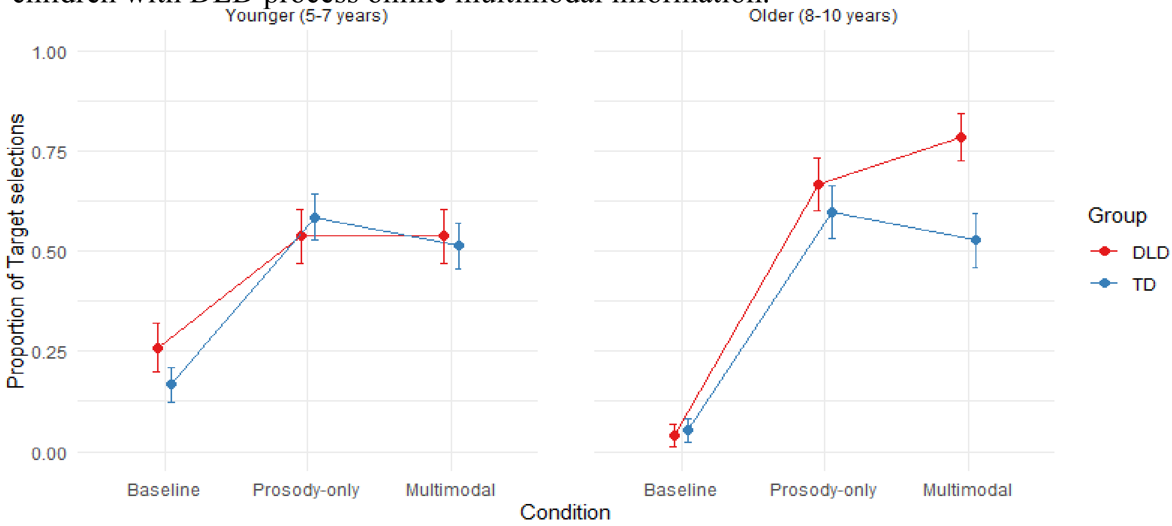


Figure 1. *Accuracy in target selection for younger (left panel) and older children (right panel), across Linguistic Groups (DLD: red; TD: blue). The X axis shows the three experimental conditions: baseline (left columns), prosody (centre columns) and multimodal (right columns).*

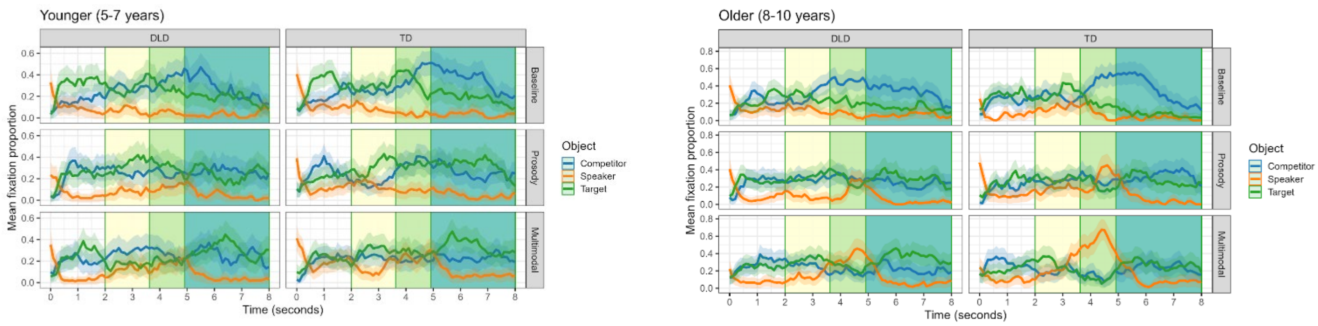


Figure 2. *Mean fixation proportion of fixations to Target, Competitor and Speaker across Conditions, and Linguistic Groups by the younger (left panel) and older (right panel) groups. The 0 in the X axis represents the start of the trial. The three coloured areas represent the time window 1 (yellow area), time window 2 (green area) and time window 3 (blue area).*

[1] Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238-299. https://doi.org/10.1016/j.cogpsych.2004.03.001

[2] Choi, Y., & Mazuka R., (2003). Young Children's Use of Prosody in Sentence Parsing. *Journal of Psycholinguistic Research* 32 (2): 197–217.

[3] De La Cruz-Pavía, I., Werker, J. F., Vatikiotis-Bateson, E., & Gervain, J. (2020). Finding Phrases: The Interplay of Word Frequency, Phrasal Prosody and Co-speech Visual Information in Chunking Speech by Monolingual and Bilingual Adults. *Language and Speech*, *63*(2), 264-291. https://doi.org/10.1177/0023830919842353

[4] Prieto, P. (1997). «Prosodic manifestation of syntactic structure in Catalan». In: Martínez-Gil, Fernando; Morales-Front, Alfonso (eds.). Issues in the Phonology and Morphology of the Major Iberian Languages. Washington, D.C.: Georgetown University Press, pp. 174-194.

8. Beat gestures align with vowel onsets in word-initial syllables and affect speech acoustics in Cantonese monosyllabic utterances

Patrick L. Rohrer [1], Yitian Hong [2], Hans Rutger Bosker [1]
[1] *Donders Institute for Brain, Cognition, and Behaviour,* [2]*The Chinese University of Hong Kong*

Manual co-speech gestures are generally produced in temporal coordination with prominent syllables in speech [1], [2]. In laboratory settings, pointing and beat gestures have specifically been shown to most stably align with pitch peaks in pitch-accented syllables in multiple languages such as English [3], Catalan [4], and French [5]. However, only three studies to our knowledge have set out to assess gesture-speech coordination in tonal languages. One study on spontaneous speech in Medʉmba [6] found that gesture apexes tend to stably align with vowel onset, and that this timing was modulated by vowel duration (longer vowels led to later apexes) and by tone (apexes occurred later in low-tone syllables). Additionally, words with gestures tended to have longer and louder vowels than those without gesture. Investigating Mandarin monosyllabic CV productions in a lab setting, [7] similarly found apexes to stably align with vowel onset, with the precise timing modulated by consonant duration (longer consonants led to earlier apexes). Additionally, words produced with gesture showed a significantly higher F0 and vowel intensity. Finally, one study on Cantonese [8] used a pointing and picture-naming task and found that speakers' pointing gesture apexes occurred early within the mono- or disyllabic target word regardless of tone or focus conditions. However, this study does not report the precise timing of gesture with respect to syllables or individual segments. Thus, the current exploratory study adds to this body of literature by seeking further cross-linguistic evidence about gesture-speech integration in tonal languages. Specifically applying the methods in [7] to Cantonese, we assess (a) the precise temporal coordination of beat gestures with both mono- and disyllabic words, and (b) the effect gesture production has on speech.

Five native speakers of Cantonese were asked to produce 30 monosyllabic words in isolation from tonal minimal pairs between tone 1, a high flat tone, and tone 3, a mid-level flat tone (e.g., /si1/, meaning *poem* and /si3/, meaning *attempt*). Each word was once produced with and once without a beat gesture. Speakers also produced 12 disyllabic words (2 words for each of the 6 Cantonese lexical tones), where each syllable within the word was produced with the same tone (e.g., /sɐi1 kʷa1/, meaning *watermelon*; /sɵi2 sɐu2/, meaning *sailor*, etc.). Praat was used to transcribe the audio clips and extract acoustic data (i.e., F0, intensity, duration). Video clips were annotated in ELAN for gesture apexes (i.e., the maximal downward extension of the hand). All scripts and data are openly available at [9].

With regard to timing, in monosyllabic utterances, we found that the most stable landmark for gesture production was vowel onset (occurring 31 ± 79 *ms* after vowel onset). This timing was modulated by consonant duration, with longer consonants leading to earlier apexes (β = -26.435, p < .001; Figure 1, upper panel). With regard to the effect of gesture production on speech, monosyllables produced with a gesture had a significantly higher mean F0 in semitones than those produced without gesture (β = 0.599, p = .022; Figure 2, upper panel), and were slightly shorter (β = -0.036, p = .008). In disyllabic target words, the gesture always occurred during the first syllable and were similarly most closely associated with the vowel onset (occurring 71 ± 79 *ms*). However, neither tone nor vowel or consonant duration modulated the precise timing of the gesture. Regarding the acoustic effects on speech, gesture production did not significantly affect the mean F0 (Figure 2, lower panel), intensity, or duration of either syllable, possibly due to lower n. Taken together, these results show similar patterns to what has been previously shown for Mandarin monosyllables, while the results from disyllables may suggest that precise timing or acoustic effects may become less robust as speech becomes more complex. This highlights the need for more cross-linguistic work in more naturalistic contexts.

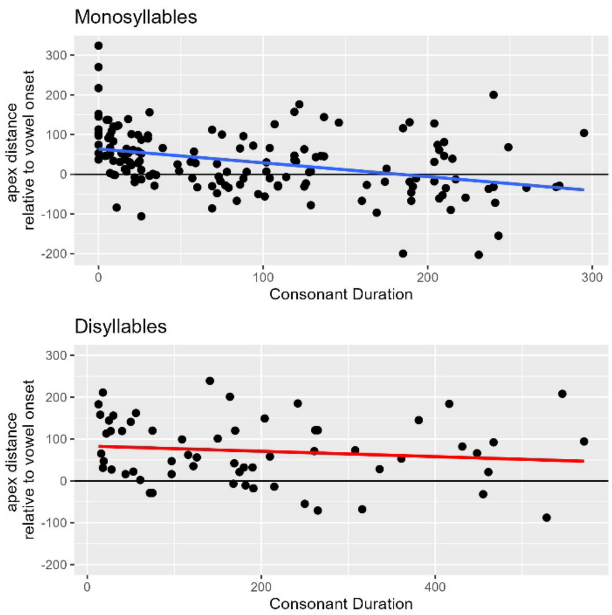


Figure 1. *Scatterplot showing apex distance to vowel onset as a function of consonant duration in monosyllables (upper panel) and disyllables (lower panel).*

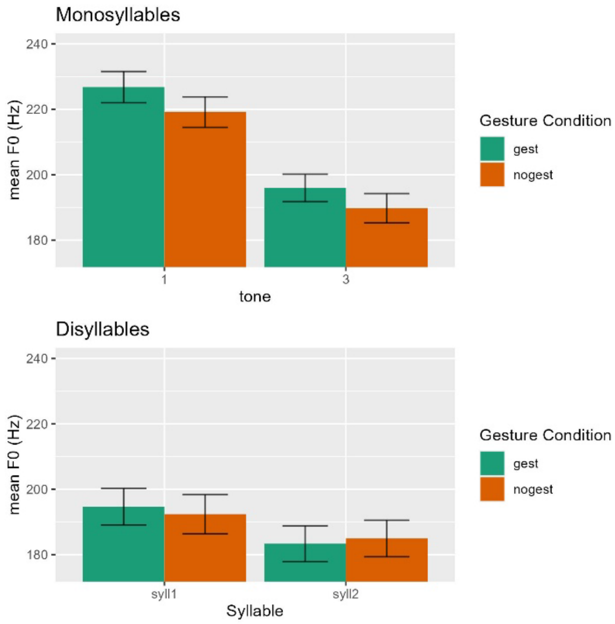


Figure 2. *Mean F0 (Hz) as a function of when utterances were produced with gesture (green) or without (orange). Upper panel shows monosyllables divided by tone, lower panel shows disyllables divided by syllable.*

[1] Loehr, D. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3(1), 71-89.

[2] Shattuck-Hufnagel, S. & Ren, A. 2018. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Front. Psychol.* 9, 1514.

[3] Leonard, T. & Cummins, F. 2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes* 26(10), 1457–1471.

[4] Esteve-Gibert, N. & Prieto, P. 2013. Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements. *J Speech Lang Hear Res* 56(3), 850–864.

[5] Roustan, B. & Dohen, M. 2010. Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. *Speech Prosody 2010*, 110.

[6] Franich, K. & Keupdjio, H. 2022. The influence of tone on the alignment of speech and co-speech gesture. *Speech Prosody 2022*, 307–311.

[7] Rohrer, P.L., Hong, Y., & Bosker, H.R. 2024. Gestures time to vowel onset and change the acoustics of the word in Mandarin. *Speech Prosody 2024*, 866–870.

[8] Fung, H.S.H. & Mok, P.P.K. 2018. Temporal coordination between focus prosody and pointing gestures in Cantonese. *Journal of Phonetics* 71, 113–125.

[9] Rohrer, P.L., Hong, Y., & Bosker, H.R. 2025. Beat gestures align with vowel onsets in word-initial syllables and affect speech acoustics in Cantonese monosyllabic utterances. *Open Science Framework*, osf.io/fk24w.

9. How multimodal saliency helps the acquisition and generalization of verb-flection morphemes

Aina Bonet, Mònica Sanz, Núria Esteve-Gibert
Universitat Oberta de Catalunya (UOC)

Acquiring verb-flection morphemes is a challenging process ([1]): not only can they be very variable, but they also tend to be less salient in speech ([2]). Research on L1 ([3]) and L2 adults in typical and atypical populations ([4]) shows that linguistic stimuli are better acquired when presented saliently, either prosodically (via pitch, duration and/or intensity) or multimodally (e.g., with head or hand "beat" gestures accompanying speech). While the role of saliency has begun to be explored within the area of lexical and grammatical learning ([5]), it remains unknown how both prosodic and multimodal saliency can affect morphological learning, and particularly, morphemes related to verb flection.

This research aims to explore whether saliency contributes to acquiring and generalizing verb-flection morphemes more quickly and more accurately, and whether the type of saliency (prosodic or multimodal) leads to different learning effects.

To investigate this, we designed a study inspired by [6] in which we explore the acquisition and generalization of number in the verb flection of pseudowords (1st person singular vs. 1st person plural) while adding the variable of saliency (no saliency, prosodic or multimodal). A total of 78 Finnish L1 adult speakers (26 per condition), with no knowledge of Romance languages will take part in an experimental learning task. Finnish and Catalan (and other Romance languages) both have verb-flection morphemes for 1st person singular and plural but differ in lexical stress assignment (Finnish lexical stress always falls on the first syllable of the word and thus verb-flection morphemes are never stressed, while in Catalan the position of the lexical stress can vary). This allows us to specifically investigate the role of saliency in such acquisition process.

The experiment consists of three phases. In the first phase (exposition), participants watch a video of a person presenting 256 singular or plural verb forms that follow Catalan phonotactics (e.g., *dunt-ol* vs. *dunt-op*) matched with an image representing a particular action (e.g., person cycling vs. people cycling; see Figures 1 and 2). Additionally, 128 fillers are also included. In the second phase (acquisition), participants then hear the verb forms (e.g. *dunt-ol*) while seeing images of an action from the exposition phase (e.g. people cycling) and are asked to answer whether the verb form and the image of the action match (in this case, *dunt-ol* & people cycling [plural form] do not match). In the third phase (generalization), participants hear new stimuli with the same morphemes (e.g., *tilm-ol*) while seeing two new images of actions (e.g., person reading vs. people reading), and are asked to choose which of the two images matches the verb form. A final questionnaire is added to verify explicit knowledge of the verb-flection rules, and additional cognitive measures (phonological processing and working memory) will be gathered. Crucially, participants are divided into three different groups during the exposition phase, following a between-subject design: no saliency on the morpheme (*dunt-ol*+picture of person cycling), prosodic saliency on the morpheme (*dunt-**ol***+picture of person cycling) and multimodal (prosody + beat gesture) saliency on the morpheme (*dunt-**<u>ol</u>***+picture of person cycling) (see Figure 3).

The pilot of this experiment is currently being run. Reaction time (from video onset until participants' response) and accuracy in the acquisition and generalization phases will be the dependent variables in the models, and the fixed factor will be the condition (3 levels: no-saliency, prosodic, multimodal) to which learners were exposed during the exposition phase. We expect that learners exposed to the multimodal condition will acquire verb flection morphemes faster (lower reaction time) and more accurately (higher percentage of correct responses) than those exposed to the prosodic condition, and those in the prosodic condition will be faster and more accurate than those exposed to the no-saliency condition. This research will shed light on how saliency plays an essential role when it comes to language acquisition as a bridge between speech and gesture.

Figures 1 and 2. *Examples of the visual stimuli that learners had to learn to match with a verb form. Left figure: image representing the form dunt-ol (1st person singular); right image: image representing the form dunt-op (1st person plural).*

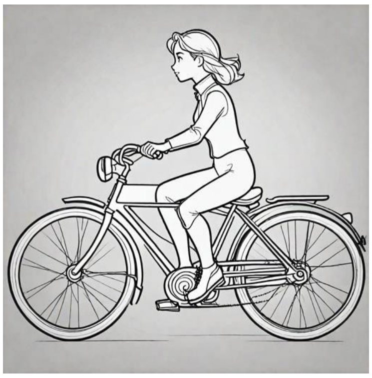


Figure 1. *Image matching the* dunt-ol *form (singular form).*

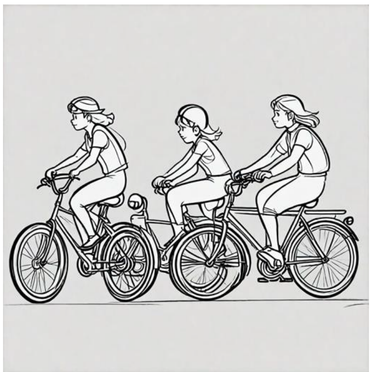
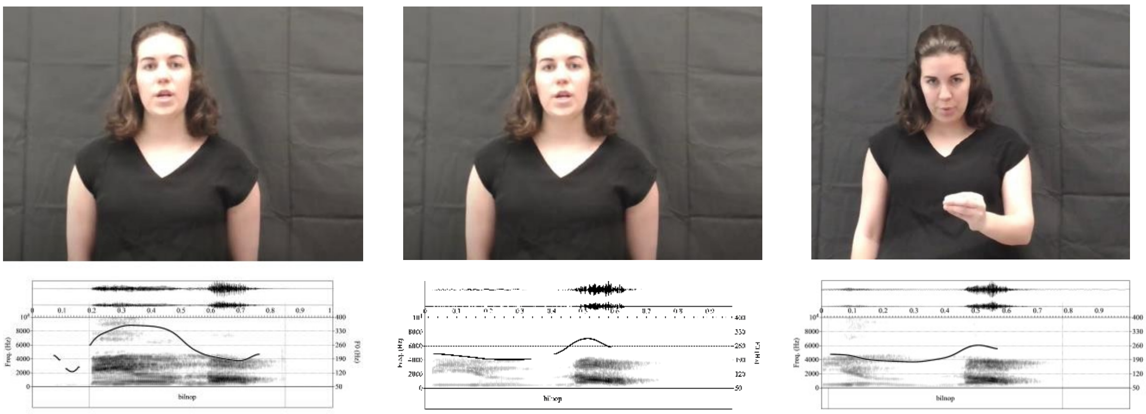


Figure 2. *Image matching the* dunt-op *form (plural form).*

Figure 3. Spectrograms and clip captions of the stimuli that listeners were exposed to in the no-saliency condition (left), prosodic condition (centre), and multimodal condition (right).



[1] Kodner, J., Khalifa, S. & Payne, S. (2023). Exploring Linguistic Probes for Morphological Generalization. *arXiv preprint arXiv:2310.13686*.
[2] Wang, Y., Seidl, A. & Cristia, A. (2015). Acoustic-phonetic differences between infant- and adult-directed speech: the role of stress and utterance position. *Journal of child language, 42*(4), 821-842.
[3] Ferry, A., Nespor, M. & Mehler, J. (2020). Twelve to 24-month-olds can understand the meaning of morphological regularities in their language. *Developmental psychology*, *56*(1), 40.
[4] Demuth, K., & Tomas, E. (2016). Understanding the contributions of prosodic phonology to morphological development: Implications for children with Specific Language Impairment. *First Language*, *36*(3), 265-278.

[5] Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, *134*(3), 406.
[6] Havas, V., Waris, O., Vaquero, L., Rodríguez-Fornells, A. & Laine, M. (2015). Morphological learning in a novel language: A cross-language comparison. *Quarterly Journal of Experimental Psychology, 68*(7), 1426-1441.

10. Speech meets reading: orthographic knowledge modulates speech categorization

Yue Qiu[1,2], Clara Martin[1,3], Antje Stoehr[1,3], Peng Li[1]

[1]Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastian, [2]Universidad del País Vasco (UPV-EHU), [3]Ikerbasque, Basque Foundation for Science, Bilbao

L1 speech categories are considered to develop primarily during infancy. However, recent research suggests speech categorization continues to evolve through adolescence [1] and its development is linked to reading acquisition [2–4]. Spanish provides a unique testbed for this link, as it features three plosive pairs with voicing contrasts (/b/-/p/, /d/-/t/, /g/-/k/) that differ in orthographic consistency. The /g/-/k/ contrast has inconsistent spellings at both ends (<g>, <gu> for /g/; <g> also represents /x/ in certain contexts; <c>, <k>, <qu> for /k/), posing decoding challenges for young readers. By contrast, /b/-/p/ has only one inconsistent endpoint (/b/ spelled as <b> or <v>), while /d/-/t/ is fully consistent.

This longitudinal study examined whether speech categorization continues to develop with reading experience and whether orthographic knowledge modulates this process. Sixty Spanish children were tested at ages 7 and 9 on three two-alternative forced-choice identification tasks (one per contrast), each employing a six-step voice-onset-time (VOT) continuum presented as pseudowords without any orthographic cues. Orthographic knowledge was assessed using a nonsense text reading task comprising words and pseudowords. While the focus was on perception, we also conducted three cued production tasks (for /k/, /t/, /p/) to provide supplementary data.

At age 7, children with stronger orthographic knowledge exhibited more precise /g/-/k/ categorization and faster /k/ identification, whereas no group differences emerged for /b/-/p/ or /d/-/t/. By age 9, performance on /g/-/k/ improved across all participants, and differences linked to reading proficiency diminished. Preliminary production measures (VOT and reaction times) also revealed a distinct pattern for /k/ relative to /t/ and /p/, aligning with perceptual outcomes.

These findings indicate that speech categorization continues maturing in the early reading years, with orthographic inconsistencies exerting a strong influence on the /g/-/k/ contrast. Recognizing this interplay between speech processing and orthography can inform strategies to support reading acquisition.

Keywords (3 to 5): speech categorization, orthographic knowledge, reading acquisition, orthographic consistency, voice onset time

[1] McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. Developmental Psychology, 54(8), 1472–1491.

[2] Burnham, D. (2003). Language specific speech perception and the onset of reading. Reading and Writing, 16(6), 573–609. https://doi.org/10.1023/A:1025593911070

[3] Hoonhorst, I., Medina, V., Colin, C., Markessis, E., Radeau, M., Deltenre, P., & Serniclaes, W. (2011). Categorical perception of voicing, colors and facial expressions: A developmental study. Speech Communication, 53(3), 417–430.

[4] Kolinsky, R., Navas, A. L., Vidigal de Paula, F., Ribeiro de Brito, N., de Medeiros Botecchia, L., Bouton, S., & Serniclaes, W. (2021). The impact of alphabetic literacy on the perception of speech sounds. Cognition, 213, 104687.

Appendix

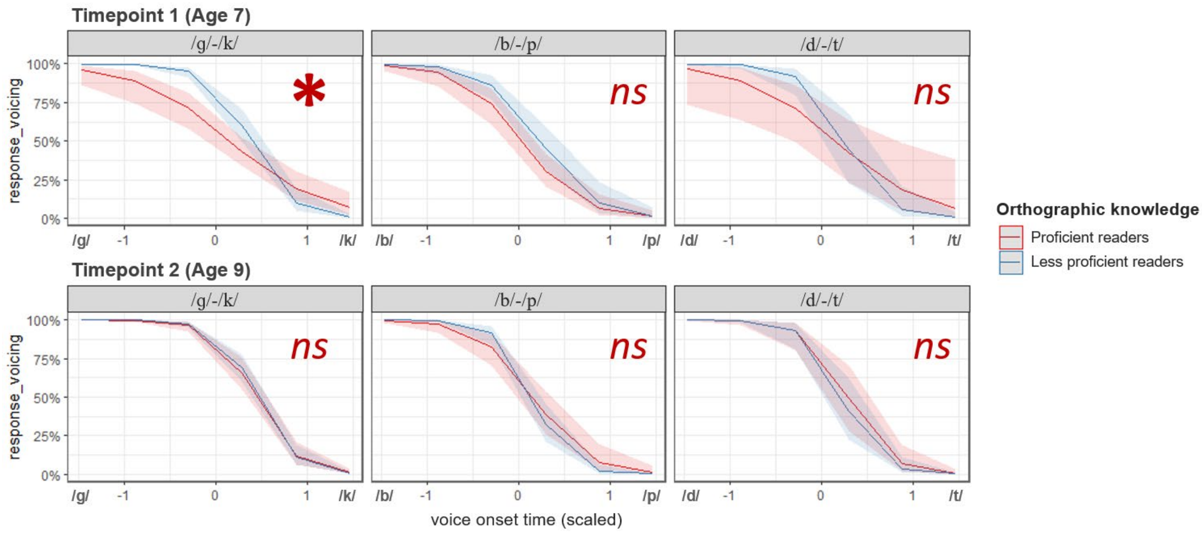


Figure 1. *Probability of voiced response (/g/, /b/, /d/) as a function of scaled Voice Onset Time (VOT) for proficient readers (blue line) and less proficient readers (red line) at Timepoint 1 (age 7) and Timepoint 2 (age 9). Results are shown for each phonemic contrast (/g/-/k/, /b/-/p/, /d/-/t/). Significant effects are highlighted with an asterisk (*), while "ns" indicates non-significant results.*

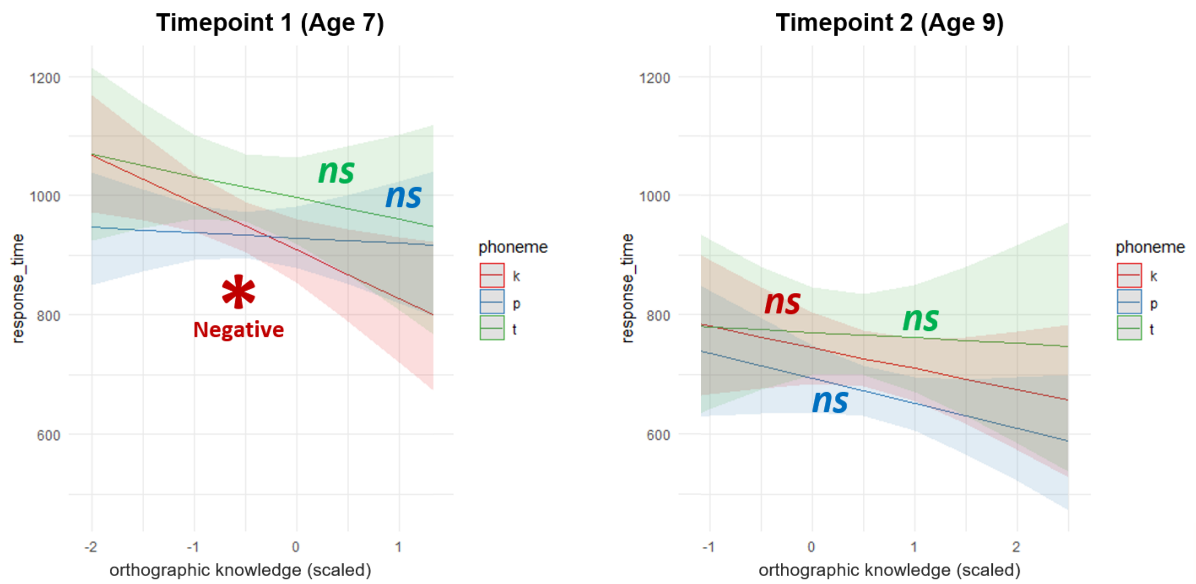


Figure 2. *Relationship between response time for correct identification of /k/, /p/, and /t/ and orthographic knowledge at Timepoint 1 (age 7) and Timepoint 2 (age 9). Significant effects are highlighted with an asterisk (*), while "ns" indicates non-significant results.*

11. Geography of children's articulation skills in Germany, with a special focus on rhotacism and kapacism

Eugen Zaretsky[1], Benjamin P. Lange[2], Christiane Hey[1]
[1]*Marburg University Hospital, Marburg, Germany,* [2]*IU International University, Berlin, Germany*

Introduction: The uneven geographical distribution of sociodemographic characteristics such as income has been shown to be associated with children's language competence [1–5]. Children who live or attend daycare centers in the districts or regions with a high percentage of non-migrants, of inhabitants with a high educational level, high income, with a low percentage of foreigners, immigrants, and unemployed tend to have advanced language competence. Also, language(-related) impairments and disorders (e.g., hearing disorder) can be found in such areas less often than in those with a comparatively low socioeconomic status of inhabitants [6]. The current study focused on the geographical distribution of the articulation skills of four-year-old children in the German language, especially on rhotacism and kapacism.

Methods: A total of 1,736 four-year-old children were tested in the German state of Hesse with the standardized, validated language screening "Kindersprachscreening" (KiSS.2) [7]. It includes subtests on the speech comprehension, vocabulary, articulation, grammar, and the phonological short-term memory as well as questionnaires for parents and daycare center teachers. Sociodemographic characteristics of the regions were correlated with children's articulation skills. Some sociodemographic characteristics were available not for the whole sample but only for a subsample of children living in the city Frankfurt/Main (n = 697).

Results: Children's articulation skills were better in the areas with a high percentage of non-migrants, a higher income, higher educational level, more ENT doctors and paediatricians, more children attending nursery schools and daycare centers, especially those with a high percentage of German children; with less unemployed and immigrants (Spearman's correlations). Children with rhotacism lived in the areas with a low socioeconomic status of inhabitants (Mann-Whitney U tests). On the contrary, children with kapacism lived in the areas with mixed sociodemographic characteristics: on the one hand, less highly qualified employees, low attendance of nursery schools and daycare centers, less paediatricians, but on the other hand, a low percentage of foreigners and immigrants, more living area per person, higher income, more employed inhabitants, less children living under conditions of poverty. Rhotacism occurred significantly more often among bi-/multilingual children, kapacism among monolingual German children (Chi-square tests).

Discussion: Age-appropriate articulation skills, including the absence of rhotacism, were found in the areas with a high socioeconomic status. Children in these areas had a higher quality and quantity of the German language input (e.g., through the early attendance of daycare centers), better access to healthcare services, and relatives with a high educational level. Since kapacism occurred more often among monolingual Germans than among bi-/multilinguals, it was associated with some characteristics of a high socioeconomic status. However, this status was clearly lower than that in the areas where children with age-appropriate articulation skills resided. Thus, kapacism seems to be characteristic of areas inhabited by German non-migrants with a low socioeconomic status (that is still higher than the status of most immigrants).

[1] Froyland, J. M., Powell, D. R., & Diamond, K. E. (2014). Relations among neighborhood social networks, home literacy environments, and children's expressive vocabulary in suburban at-risk families. *School Psychology International, 35*(4), 429–444.

[2] Zaretsky, E. (2019). Geographie der Wortschatzkenntnisse bei Vorschulkindern in Frankfurt am Main: Migrationshintergrund, Einkommen und andere demographische Einflussgrößen [Geography of vocabulary skills of pre-schoolers in Frankfurt/Main: Immigrant background, income, and other demographic influencign factors]. In C. Fölbes & L. A. Nefedova (Eds.), *Beiträge zur Interkulturellen Germanistik*. Vol. 13 (pp. 207–239). Gunter Narr.

[3] Rydland, V., Grøver, V. & Lawrence, J. (2014). The second-language vocabulary trajectories of Turkish immigrant children in Norway from ages five to ten: The role of preschool talk exposure, maternal education, and co-ethnic concentration in the neighborhood. *Journal of Child Language, 41*(2), 352−381.

[4] Da Rocha Neves, K., de Souza Morais, R. L., Teixeira, R. A., & Pinto, P. A. (2016). Growth and development and their environmental and biological determinants. *Jornal de Pediatria, 92*(3), 241−250.

[5] De Marco, A., & Vernon-Feagans, L. (2013). Rural neighborhood context, child care quality, and relationship to early language development. *Early Education and Development 24*(6), 792−812.

[6] Zaretsky, E., & Lange, B. P. (2016). The geography of language skills and language(-related) disorders: A case of Frankfurt/Main. In S. Grucza, M. Olpinska-Szkielko, & P. Romanowski (Eds.), *Advances in understanding multiculturalism: A global perspective* (pp. 181−197). Peter Lang.

[7] Holler-Zittlau, I., Euler, H. A., & Neumann, K. (2011). Kindersprachscreening (KiSS) – das hessische Verfahren zur Sprachstandserfassung [Kindersprachscreening (KiSS) – the Hessian language test]. *Sprachheilarbeit, 5*(6), 263–268.

12. Reading skills and their phonological and language precursors in children with oral and writing language difficulties

Eva Aguilar-Mediavilla, Lucía Buil-Legaz, Josep Pérez Castelló, Angélica Mateus, Àngels Esteller-Cano, Daniel Adrover-Roig

Institute for Educational Research and Innovation (IRIE), Investigation in Development, Education and Language (I+DEL), Universitat de les Illes Balears (UIB)

Developmental Language Disorder (DLD) causes persistent language delays affecting communication and/or learning without biomedical condition such as autism spectrum disorder, intellectual disability, neurological damage or hearing loss [1]. Reading Disorders (RD) involve specific reading decoding and/or comprehension impairments [2]. These two difficulties often overlap [3] as both share cognitive and linguistic challenges such as phonological working memory (PWM), verbal fluency and phonological awareness [6].Thus, children with RD often present oral language delays, and children with DLD can present RD [4], [5]. Previous studies, primarily in English, a non-transparent reading language, indicate phonological processing links to reading skills in children with DLD, with mixed results in transparent languages like Italian [7]. A prior study with bilingual Catalan-Spanish children with DLD showed initial decoding difficulties that evolved into comprehension difficulties in later primary school, linked to semantic fluency [8].

This study aims to 1) compare the reading decoding and comprehension skills in bilingual Catalan-Spanish children with DLD, RD and Typical Development (TD) at ages 10 and 14, 2) examining longitudinally the influence of PWM, phonological fluency, and oral language at age 10 on reading decoding and comprehension at 14 years. The sample included 29 TD children, 15 children with DLD and 34 children with RD, all Catalan-Spanish bilinguals, with a mean age of 10.5 (primary school) at the beginning of the study and 13.8 (secondary school) at the end. We assessed both phonological fluency (subtest verbal fluency) and PWM (non-word repetition task) with the NEPSY I [9], oral language with the CELF-IV (core language) [10] and decoding abilities (word reading/time) and reading comprehension (text comprehension) with the Spanish version of the PROLEC-R/SE [11].

Results (one way ANOVAs; *Figure* 1) show a significant main effect of Group in Decoding ($F(2, 74) = 18.24$, $p = <.001$), but not in Comprehension ($F(2, 75) = 2.58$, $p = .082$) at 10 years; and in Decoding ($F(2, 72) = 12.4$, $p = <.001$) and Comprehension ($F(2, 71) = 8.6$, $p = <.001$) at 14 years. *Post-hoc* analyses show significantly better Decoding skills in TD children than in both impaired groups, DLD ($p_{10y} = .003$ and $p_{14y} = <.006$) and RD ($p_{10y} = <.001$ and $p_{14y} = <.001$), showing the two impaired groups similar Decoding skills in each moment ($p_{10y} = .694$ and $p_{14y} = 1.0$). Differently, *post-hoc* analyses on Comprehension at 14 years show significantly poor Comprehension skills in DLD group compared to the other two groups ($p_{TD} = <.001$ and $p_{RD} = <.010$) which show equal Comprehension skills ($p_{14yTD\text{-}RD} = <.545$).

Correlational analyses show that only Phonological variables are related to Decoding (PWM, $r = .337$, $p = <.003$) and Comprehension skills at 10 years of age (Phonological fluency, $r = .224$, $p = <.049$). At 14 years, Decoding was not related ($p_s > .130$) with any of the precursors measured at 10 years (PWM, Phonological fluency and Language), meanwhile reading Comprehension was related to Language measured at 10 years ($r = .470$, $p = <.001$).

These findings suggest that, in transparent languages such as Spanish and Catalan, children with DLD struggle with decoding similarly to children with RD throughout school and only children with DLD show comprehension difficulties in secondary school. While reading skills (decoding and comprehension) are related to phonological skills at 10 years of age, at 14 years of age, reading comprehension is only related to oral language skills. Phonological skills seem crucial in early reading in transparent languages but as comprehension demands increase, language skills become more influential [7].

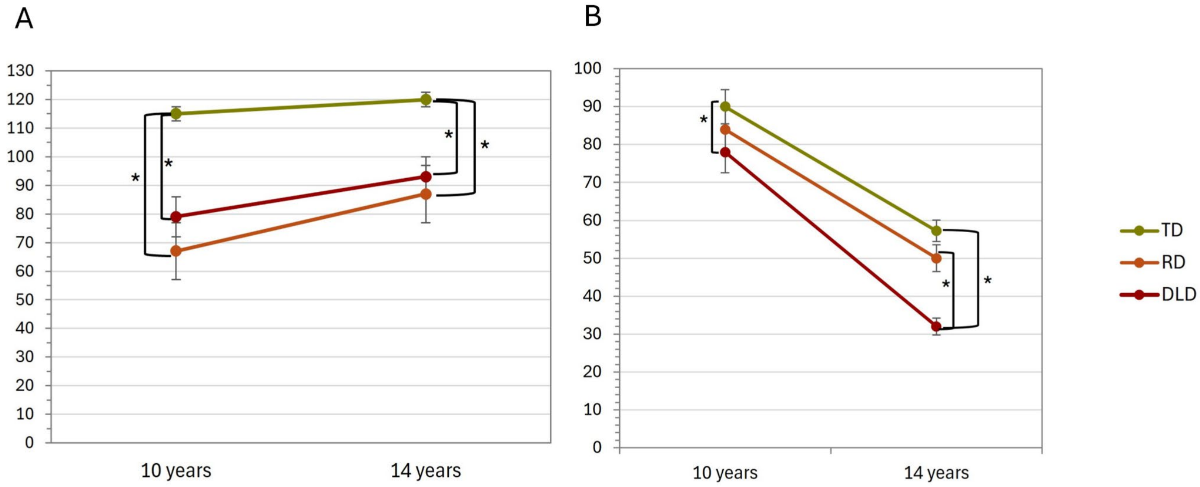


Figure 1. *Decoding (A) and Comprehension (B) score in TD, DLD and RD groups at 10 and 14 years. Reading has been measured with PROLEC-R at 10 years and PROLEC-SE at 14 years being similar the tasks of decoding in both texts; but the comprehension texts of PROLEC-SE were longer making them more difficult to children in all the groups.*

[1] D. V. M. Bishop, M. J. Snowling, P. A. Thompson, T. Greenhalgh, and CATALISE-2 consortium, 'Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology', *J. Child Psychol. Psychiatry*, vol. 58, no. 10, pp. 1068–1080, Oct. 2017, doi: 10.1111/jcpp.12721.

[2] A. Castles, K. Rastle, and K. Nation, 'Ending the reading wars: Reading acquisition from novice to expert', *Psychol. Sci. Public Interes.*, vol. 19, no. 1, pp. 5–51, 2018, doi: 10.1177/1529100618772271.

[3] M. J. Snowling, K. Moll, and C. Hulme, 'Language difficulties are a shared risk factor for both reading disorder and mathematics disorder', *J. Exp. Child Psychol.*, vol. 202, p. 105009, Feb. 2021, doi: 10.1016/j.jecp.2020.105009.

[4] H. W. Catts, 'Commentary: The critical role of oral language deficits in reading disorders: reflections on Snowling and Hulme (2021)', *J. Child Psychol. Psychiatry Allied Discip.*, vol. 62, no. 5, pp. 654–656, 2021, doi: 10.1111/jcpp.13389.

[5] M. J. Snowling and C. Hulme, 'Interventions for children's language and literacy difficulties', *Int. J. Lang. Commun. Disord.*, vol. 47, no. 1, pp. 27–34, 2012, doi: 10.1111/j.1460-6984.2011.00081.x.

[6] M. J. Snowling, H. M. Nash, D. C. Gooch, M. E. Hayiou-Thomas, and C. Hulme, 'Developmental Outcomes for Children at High Risk of Dyslexia and Children With Developmental Language Disorder', *Child Dev.*, vol. 90, no. 5, pp. 1–17, Sep. 2019, doi: 10.1111/cdev.13216.

[7] J. H. Y. Lam, M. A. Leachman, and A. S. Pratt, 'A systematic review of factors that impact reading comprehension in children with developmental language disorders', *Res. Dev. Disabil.*, vol. 149, no. August 2023, p. 104731, 2024, doi: 10.1016/j.ridd.2024.104731.

[8] L. Buil-Legaz, E. Aguilar-Mediavilla, and J. Rodríguez-Ferreiro, 'Reading skills in young adolescents with a history of Specific Language Impairment: The role of early semantic capacity', *J. Commun. Disord.*, vol. 58, pp. 14–20, 2015, doi: 10.1016/j.jcomdis.2015.08.001.

[9] M. Korkman, U. Kirk, and S. Kemp, *NEPSY: Developmental Neuropsychological Assessment*. San Antonio: Psychological Corp., 1998.

[10] E. Semel, E. H. Wiig, and W. A. Secord, 'Clinical Evaluation of Language Fundamentals®

- Fourth Edition', 2003.
[11] F. Cuetos-Vega, B. Rodríguez, E. Ruano, and D. Arribas, 'PROLEC-R Batería de Evaluación de los Procesos Lectores Revisada', *Madrid, TEA*, 2007.

13. Vocal expression of emotions in UVFP patients before and after medialization

Nicolas Audibert[1], Muriel Lalain[2], Alexia Mattei[2,3], Caterina Petrone[2]

[1]Laboratoire de Phonétique et Phonologie, CNRS & Sorbonne Nouvelle, Paris, FR ; [2]CNRS, LPL, UMR 7309, Aix-Marseille Université, Aix-en-Provence, FR ; [3] Hôpital La Conception, Marseille, FR

Prosody provides a powerful means to express emotions and, as such, it contributes to effective communication and social functioning [1]. While, in healthy adults, basic emotions are associated with systematic prosodic modulations, patients with unilateral vocal fold paralysis (UVFP) suffer from limitations in their capacity to produce emotional prosody [2], which partly explains the feelings of social isolation often reported by patients [2].

We explore emotional prosody in UVFP patients and evaluate the benefits of treatment, by comparing patients before and after medialization and healthy controls (HC). UVFP consists of an immobility of one of the vocal folds, resulting in dysphonia and instability in the vibratory pattern of the vocal folds. UVFP patients report weak voice, breathiness, roughness, diminished voice intensity, diplophonia, and air loss [3, 4]. Acoustically, UVFP results in higher values of jitter and shimmer, lower values of the harmonics-to-noise ratio (HNR), and a lower fundamental frequency (f0) range compared to HCs [5]. Vocal cord medialization is proposed for improving voice quality [6]. However, there are no studies on the functional improvement of the voice after medialization. We focus on the impact of medialization on the expression of anger and sadness, as these emotions are often conveyed by voice quality cues [7].

Five French UVFP patients (64.8 y.o; 4 women) and ten healthy controls (65.6 y.o; 5 women) were recorded so far. Eight short sentences with verbal neutral meaning and the same syntactic structure (e.g., *Il va rentrer chez lui*, “He is going back home”, [2]) were embedded in situational contexts intended to elicit three emotional states (neutral/sad/angry). Participants read all the contexts and target sentences silently, before producing the target sentences without reading. To facilitate the task, sentences were presented in three different blocks of neutral, sad, and angry emotional states. The intended emotion was indicated at the beginning of each block, and each block was preceded by a familiarization and a training phase. Within each block, sentences were presented in a random order. The same material was recorded twice for patients, before medialization on the same day (‘pre’ condition) and ~1 month later (‘post’).

We collected a total of 240 utterances for patients and 240 utterances for HCs. Acoustic measures were extracted at the midpoint of vowel /a/ of the word *va (“is going”)*. f0 within the /a/ was extracted using FCN-f0 [8]. HNR over 1kHz and smoothed cepstral peak prominence (CPPS) were extracted using a Praat script. Before medialization, patients had longer vowel duration than HCs, for anger ($t=3.33$, $p=0.007$) and neutral ($t=2.64$, $p=0.03$), while vowel duration was similar between HCs and patients in sadness. After medialization, vowel duration was similar to that of HCs for the three emotions ($p>.05$). However, other measures were less affected by medialization. CPPS values (Fig. 1) indicated that patients remain more dysphonic than HCs after medialization ($t=4.76$, $p < .001$). While no differences in HNR values were found for patients before vs. after medialization nor among emotions, HNR values in HC were higher in anger ($t=4.45$; $p< .001$) and in neutral ($t=2.97$; $p<.01$) than in sadness. Thus, HNR in HC showed a richer structure, helping distinguishing between the emotional states. An interaction between group and emotion ($\chi^2(2)=15.1$; $p<.001$) was found for f0. Sadness had similar f0 values as the neutral state, for both patients and HCs. Anger in patients was characterized by higher f0 values before medialization ($t=2.73$; $p=.01$), possibly because of the higher f0 variability before medialization. These preliminary results suggest that the improvement in voice quality and speech rate management brought by medialization partially allows for a greater modulation of emotional expression in UVFP patients.

Figure 1. Acoustic comparison of the realization of /a/ on the three intended emotions, for controls and UVFP patients in pre- and post-medialization recording sessions: (a) smoothed cepstral peak prominence (CPPS), (b) vowel duration, (c) harmonic-to-noise ratio (HNR) computed after high-pass filtering to retain only frequencies above 1kHz, (d) f0 in semitones compared to the neutral expression produced by the same speaker,

[1] Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, 40(1–2), 227–256.

[2] Haddad R, Mattei A, Petrone C, Cachi-Pouyenne M, Bogdanski E, Galant C, Giovanni A. (2024). Alterations in Emotional Expression Through Speech in Patients With Unilateral Vocal Fold Paralysis: A Preliminary Study. J Voice S0892-1997(24)00223-6

[3] Lotto, A. J., Holt, L. L., Kluender, K. R. (1997). Effect of Voice Quality on Perceived Height of English Vowels. *Phonetica*, 54(2):76–93.

[4] Jesus, L. M. T. , Martinez, J., Hall, A., Ferreira, A. (2015). Acoustic Correlates of Compensatory Adjustments to the Glottic and Supraglottic Structures in Patients with Unilateral Vocal Fold Paralysis. *Biomed. Res. Int.*, 704121.

[5] Hirano, M., Mori, K., Tanaka, S., Fujita, M. (1995). Vocal function in patients with unilateral vocal fold paralysis before and after silicone injection. *Acta Otolaryngol*, 115(4):553-9.

[6] Vij, S., Gupta, A. K., Vir, D. (2017).Voice Quality Following Unilateral Vocal Fold Paralysis: A Randomized Comparison of Therapeutic Modalities. Journal of Voice,31(6), 774.e9-774.e21.

[7] Gobl, C., and Chasaide, A. N. (2003). The Role of Voice Quality in Communicating Emotion, Mood and Attitude. *Speech Communication*, 40, 189–212.

[8] Ardaillon, L., Roebel, A. (2019) Fully-Convolutional Network for Pitch Estimation of Speech Signals. *Proc. Interspeech* 2019, 2005–2009.

14. Vowel Space as a Tool for Assessing Speech in Primary Progressive Aphasia

Wendy Elvira-García[1], Victoria Marrero Aguiar[2], Joan Borràs-Comes[1], Miguel Ángel Santos-Santos[3,4]

[1]Universitat de Barcelona, [2]Universidad Nacional de Educación a Distancia (UNED), [3]Sant Pau Memory Unit, Hospital de la Santa Creu i Sant Pau, [4]Biomedical Research Institute Sant Pau (Universitat Autònoma de Barcelona)

Primary Progressive Aphasia (PPA) is a neurodegenerative disease primarily characterized by the progressive deterioration of language [1], [2]. Despite the evolving characterization of PPA, it is currently divided into three subtypes: logopenic (lvPPA), semantic (svPPA), and non-fluent (nfvPPA), with challenges in distinguishing between them. Linguistic studies have increasingly focused on identifying phonetic features that may help differentiate these variants [3], [4], [5], [6]. One such characteristic is vowel space, [7] which has been used in clinical linguistics to assess deviant speech and Apraxia of Speech (AOS), a condition that often varies across PPA subtypes but is not usually tested. The performance of automatic classifiers has been used as a proxy for deviant speech in PPA [8], inferring that a lower accuracy of the automatic system can be interpreted as deviant speech. Considering that in PPA one of the differences between types is if they show AOS and therefore a reduced vowel space, using these techniques could assist in assessing PPA subtypes.

In this study, we aim to describe the vowel space across different PPA variants and examine which variants exhibit deviant vowel patterns. To do so, we used the Spanish version of *The North Wind and the Sun*, read by 20 PPA patients and 10 age-matched controls. The vowels were manually annotated using Praat, and the analysis was conducted on the central 70% of the vowel formants to exclude formant transitions. We excluded outliers and normalized the data [9]. We then calculated the median values for each vowel and conducted two analyses: a Naïve Bayes classifier to classify vowels, and an overlap analysis using distance metrics that capture different relationships within the data (Euclidean, Mahalanobis and Pillai) [10].

Our results show that, as expected, the classifier performs better for stressed vowels than for unstressed vowels. The controls achieved the best classification results, followed by the lvPPA, nfvPPA and svPPA (Table 1). Interestingly, svPPA showed better classification results for unstressed vowels than for stressed vowels (Figure 1), which contradicts the common understanding that both unstressed vowels systems and svPPA typically maintains relatively preserved speech.

Regarding the Euclidean distances, the diagnostic group was significant ($\chi^2(3) = 11.426$, $p = 0.01$), with significant effects observed for control > nfvPPA ($d = 1.134$, $p = 0.02$) and svPPA > nfvPPA ($d = 1.5$, $p = 0.03$). As for Mahalanobis, significant effects where found both for diagnostic group and stress ($\chi^2(3) = 228.747$, $p < 0.001$ and $\chi^2(1) = 217.774$, $p < 0.001$). Specifically, we found more separated vowel in controls (control > nfvPPA, $d = 0.429$, $p < 0.001$; control > svPPA, $d = 0.557$, $p < 0.001$), with no significant difference between nfvPPA and svPPA ($d = 0.127$, $p = 0.080$). Pillai values only showed stress effects ($\chi^2(1) = 37.950$, $p < 0.001$): group and the interaction between group and stress were not significant, however the numeric Pillai results were also consistent: control> lvPPA > nfvPPA > svPPA.

In conclusion, this study highlights the potential of vowel space analysis as a tool for differentiating PPA subtypes. The unexpected findings in svPPA suggest that further investigation is needed to fully understand the relationship between phonological impairments and speech patterns in this variant.

	Stressed			Unstressed		
Dataset	Accuracy	Recall	Specificity	Accuracy	Recall	Specificity
Control	0.85	0.75	0.92	0.81	0.70	0.92
lvPPA	0.82	0.72	0.91	0.80	0.68	0.91
nfvPPA	0.77	0.61	0.88	0.74	0.59	0.89
svPPA	0.73	0.57	0.89	0.76	0.60	0.89

Table 1. *Performance results for the naïve bayes classifier split by diagnostic and stress.*

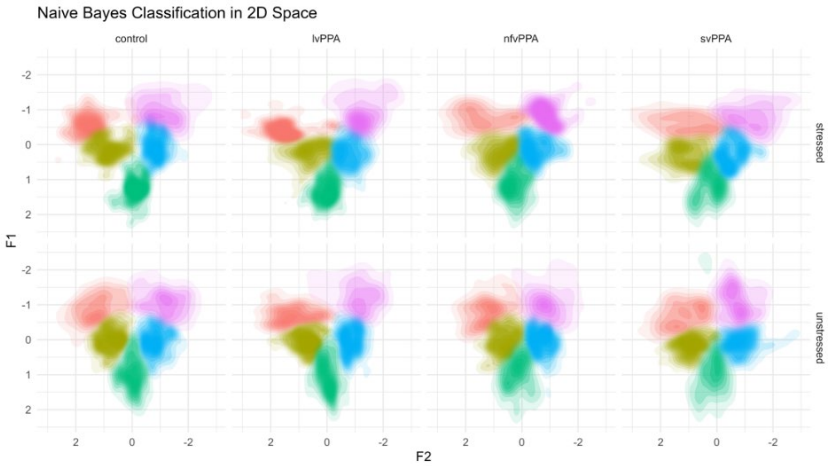


Figure 1. *Naïve Bayes classification of vowels by diagnostic group across stressed and unstressed conditions.*

[1] M. M. Mesulam. «Primary progressive aphasia», *Ann Neurol*, vol. 49, nº4, pp. 425-432, abr. 2001.

[2] A. G. Murley *et al*., «Redefining the multidimensional clinical phenotypes of frontotemporal lobar degeneration syndromes», *Brain J. Neurol*., vol 143, nº 5, pp. 1555-1571, may 2020, doi: 10.1093/brain/awaa097.

[3] K. L. Haley *et al*., «Speech metrics and samples that differentiate between nonfluent/agrammatic and logopenic variants of primary progressive aphasia», *J. Speech Lang Hear. Res*., vol. 64, nº3, 2021, doi: 10.1044/2020_JSLHR-20-00445.

[4] K. Croot, K. Ballard, C. E. Leyton, y J. R. Hodges, «Apraxia of speech and phonological errors in the diagnosis of nonfluent/agrammatic and logopenic variants of primary progressive aphasia», *J. Speech Lang Hear. Res*., vol 55, nº5, 2012, doi:10.1044/1092-4388(2012/11-0323).

[5] K. J. Ballard, S. Savage, C. E. Leyton, A. P. Vogel, M. Homberger, y J. R. Hodges, «Logopenic and Nonfluent Variants of Primary Progressive Aphasia Are Differentiated by Acoustic Measures of Speech Production», *PLOS ONE*, vol. 9, nº2, p. e89864, feb. 2014, doi: 10.1371/journal.pone.0089864.

[6] J. Stalpaert *et al*., «Language and Speech Markers of Primary Progressive Aphasia: A Systematic Review», *Am. J. Speech Lang. Pathol*, vol 29, nº4, pp. 2206-2225, nov 2020, doi: 10.1044/2020_AJSLP-20-00008.

[7] M. Shahin, B. Ahmed, J. McKechnie, K. Ballard, y R. Gutierrez-Osuna, «A comparison of GMM-HMM and DNN-HMM based pronunciation verification techniques for use in the assessment of childhood apraxia of speech», en *Interspeech 2014*, Isca, sep, 2014, pp. 1583-1587. Doi:10.21437/Interspeech.2014-377.

[8] J. A. Matias-Guiu *et al*., «Machine learning in the clinical and language characterisation of primary progressive aphasia variants», *Cortex*, vol. 119, pp. 312-323, oct. 2019, doi:10.1016/j.cortex.2019.05.007.
[9] R. A. Peterson, «Finding Optimal Normalizing Transformations via bestNormalize», *R J.* vol. 13, nº1, pp. 310-329, 2021.
[10] J. Nycz y L. Hall-Lew, «Best practices in measuring vowel merger», *Proc. Meet. Acoust*, vol. 20, nº1, p. 060008, ago. 2014, doi: 10.1121/1.4894063.

15. The Role of Prosody in Emotional Recognition Among Children With and Without Learning and Language Disorders

Oriol Verdaguer-Ribas[1,4], Mari Aguilera[1,3,4], Nadia Ahufinger[2,3], Llorenç Andreu[2,4], Mònica Sanz-Torrent[1,3,4], Ernesto Guerra[5]

[1]*Departament de Cognició Desenvolupament i Psicologia de l'Educació, Secció Cognició, Universitat de Barcelona (UB),* [2]*Estudis de Psicologia i Ciències de l'Educació, Universitat Oberta de Catalunya (UOC),* [3]*Institut de Neurociència, Universitat de Barcelona (UB),* [4]*NeuroDevelop eHealth Lab, eHealth Center, Universitat Oberta de Catalunya (UOC),* [5]*Centro de Investigación Avanzada en Educación (CIAE), Instituto de Educación, Universidad de Chile, Santiago, Chile*

Children with Specific Learning Disorders (LD), such as dyslexia and dyscalculia, and Developmental language disorder (DLD) often struggle to identify their own emotions as well as the emotions of others (Comti-Ramsden et al., 2019; Löytömaki et al., 2020). Recent empirical evidence suggests a strong interconnection between emotion and language during development, specifically pointing to a relevant role of prosody in emotional recognition in typically developing (TD) children and adults (Berman et al., 2016; Paulmann et al., 2012). This study aimed to evaluate the impact of emotional prosody on emotional identification in children with and without LD and DLD.

Two eye-tracking experiments were conducted with 31 school-age children with DLD, 88 with LD (dyslexia and/or dyscalculia) and 118 TD children, using the iMotions eye-tracker software. The study employed a language-mediated visual attention task, in which participants listened to sentences while inspecting a visual context featuring four faces from the Karolinska Directed Emotional Faces database (Lundqvist et al., 1998). Each face depicted different emotions—happiness, sadness, anger, fear, and neutral—expressed by the same person.

Participants were asked to listen to sentences containing a subject, verb and predicate. In Experiment 1, the emotions were stated explicitly (e.g., "Laura is happy"), while in Experiment 2, the emotions were conveyed implicitly, particularly driven by the verb (e.g., "Cristina escapes from the dog"). Half of the trials were prosody-content congruent, and the other half were incongruent. Participants had to click on the face expressing the emotion that matched what they heard. The emotional prosodies were created considering speech rate and pitch parameters, analysed by Praat software, described in previous literature for each emotion. All four prosodies showed statistically significant differences in speech rate, pitch rate, and mean pitch measures.

For the analysis of gaze patterns, we used the magnitude estimation approach with CI95%. Differences in fixation proportions across conditions were quantified by visualizing gaze patterns over time, with 95% confidence intervals computed every 50 ms to identify differences between conditions, objects, and groups. This method allows for precise examination of the timing and magnitude of fixations toward the target and competitors (Huettig et al., 2020).

Results indicated that in Experiment 1 there was a significant preference for the target and the semantically related face (phrase content, not prosody) for the three groups. In Experiment 2 there was also a significant preference for the target; however, when emotional meaning complexity increases (in Experiment 2's incongruent trials), there is an increasing effect of preferences of looking at the prosodic target, with the DLD group showing the strongest preference, followed by the LD group, and finally the TD group.

Therefore, TD children discarded prosodic cues to identify emotions and preferred the face congruent with the semantic content of the phrase. In children with LD and DLD, instead, prosodic cues interfered more with emotion identification. Prosody has a high impact when emotional meaning complexity increases, particularly for children with DLD and LD.

Supplementary data:

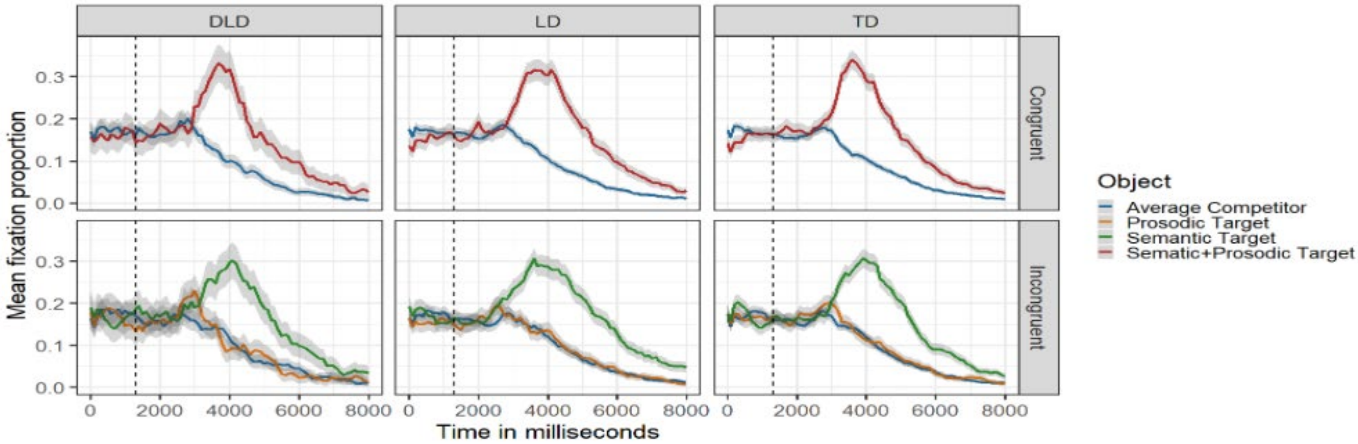


Figure 1. *Mean proportion of eye fixations for each group (Typical Development -TD-, Specific Learning Disorders -LD- and Developmental Language Disorder -DLD-) for each condition (congruent condition and non-congruent condition) for Experiment 1.*

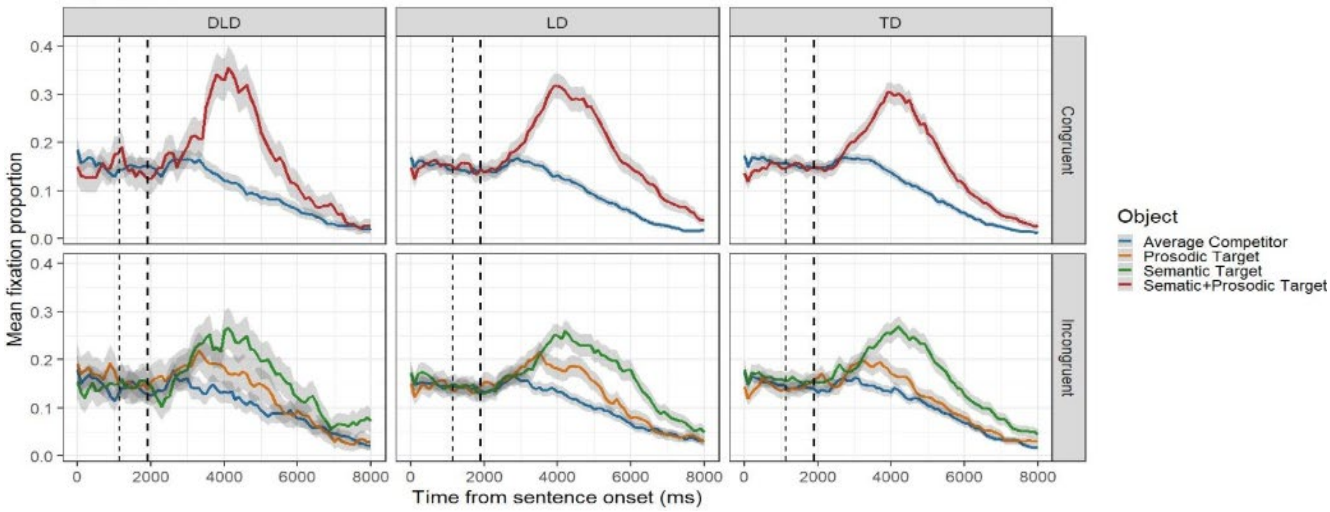


Figure 2. *Mean proportion of eye fixations for each group (Typical Development -TD-, Specific Learning Disorders -LD- and Developmental Language Disorder -DLD-) for each condition (congruent condition and non-congruent condition) for Experiment 2.*

[1] Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology, 142,* 391-399.

[2] Conti-Ramsden, G., Mok, P., Durkin, K., Pickles, A., Toseeb, U., & Botting, N. (2019). Do emotional difficulties and peer problems occur together from childhood to adolescence? The case of children with a history of developmental language disorder (DLD). *European Child & Adolescent Psychiatry, 28*(7), 993–1004. https://doi.org/10.1007/s00787-018-1261-6

[3] Huettig, F., Guerra, E., & Helo, A. (2020). Towards understanding the task dependency of embodied language processing: The influence of colour during language-vision interactions. *Journal of Cognition, 3*(1), 41. https://doi.org/10.5334/joc.135

[4] Löytömäki, J., Ohtonen, P., Laakso, M-L., Huttunen, K. (2020). The role of linguistic and cognitive factors in emotion recognition difficulties in children with ASD, ADHD or

DLD. *International Journal of Language & Communication Disorders, 55*(2), 231-242. https://doi.org/10.1111/1460-6984.12514
[5] Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: Evidence from eye movements. *Speech Communication, 54*(1), 92-107.

16. Analysing the teaching of English pronunciation: Secondary education textbooks in Galicia

Celia Veiga-Pérez
Universitat de les Illes Balears

The importance of teaching English pronunciation has fluctuated over the years from neglect to gaining relevance [1,2]. These fluctuations have been associated with shifts in language teaching methodologies, and current communicative trends have led to an increasing attention towards the instruction of English pronunciation in the English as a Foreign Language (EFL) classroom. English is used for different goals and contexts, like English as an International Language (EIL) or English as a Lingua Franca (ELF) [3], and thus it is important to focus on the features which promote comprehensibility among speakers of different L1s. However, this specific skill in the learning experience of a language remains an underexplored area in both the teaching context and in academic research [1,4].

Learners of EFL in the region of Galicia display a series of difficulties when pronouncing the English language. These are related both to the inherent obstacles of acquiring a foreign pronunciation [5,6,7] and to explicit contrasts between the target language and the learners' L1(s)—in this case, Galician, Spanish, or both—such as the difference in the number of phonemes or the grapheme-to-phoneme correspondence [8].

Previous research has shown that L2 learners can greatly benefit from pronunciation instruction [9,10]. Nevertheless, in Galicia, there appears to be a need for official regulation regarding the teaching of this skill, given that existing documents governing L2 teaching at secondary education lack specificity [11]. The *Common European Framework for Reference for Languages* [12,13] is a general guide and, thus, it does not focus on a particular language. Similarly, the *Currículo da Educación Secundaria Obrigatoria e Bacharelato* [14,15], the official document which rules academic organization of compulsory and post-compulsory education in Galicia, is imprecise. Teachers, then, usually resort to the textbooks employed in class as their sole guidance, which in turn raises the question of whether EFL textbooks provide adequate materials for teaching English pronunciation [11,16]. Nowadays, this is particularly relevant in this region, where knowledge of English phonetics is tested in the university entrance exam and, on these grounds, extra attention would be expected towards these sections of the textbook in the EFL classroom.

Against this backdrop, the present research aims to assess whether textbooks in secondary high schools in Galicia offer an adequate support for pronunciation instruction. To this purpose, a self-compiled selection of 24 textbooks from compulsory and post-compulsory education ('ESO' and 'Bacharelato') have been analysed in terms of (i) the presence or absence of pronunciation activities, (ii) the pronunciation features covered, (iii) their integration with other content units (isolated vs. integrated), and (iv) the type of exercise/task (e.g. listen-and-repeat, identification). The findings suggest that (i) despite the seemingly increase in the number of pronunciation exercises over the past decade [cf. 11], there is still a lack of consistency both across book series from different publishers and within the same series across different academic years. (ii) There is still a prevalence of segmental over suprasegmental instruction. Furthermore, (iii) there is also a tendency to isolate this skill, either in an appendix or a "pronunciation box", rather than to integrate it with other exercises in each unit, and (iv) a preference for the listen-and-repeat type of activity over other tasks such as identification—observations which align with previous studies [11,16].

The conclusions drawn from this analysis have important teaching implications: the scarcity of clear guidelines and the lack of consistent materials further handicap the already difficult task of teaching pronunciation.

[1] Pennington, M. C. 2021. Teaching pronunciation: The state of the art 2021. *RELC Journal* 52(1), 3–21.
[2] Levis, J. M. 2005. Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly* 39, 369-377.
[3] Snodin, N. & P. Resnik. 2019. WE, ELF, EIL and their implications for English language teacher education. In Steve Walsh & Steve Mann (eds.), *The Routledge handbook of English language teacher education*, 239–252. London: Routledge.
[4] Ketabi, S. & Fateme S. 2015. Pronunciation teaching: Past and present. *International Journal of Applied Linguistics & English Literature* 4(5), 182–189.
[5] Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language researche*, 233-277. Timonium, MD: York Press.
[6] Best, C. T. & M. D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O. S. Bohn (eds.), *Second language speech learning: The role of language experience in speech perception and production*, 13-34. Amsterdam, John Benjamins.
[7] Kuhl, P. K, B. T. Convoy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, & N. Tobey. 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society* 363, 979-1000.
[8] Veiga-Pérez, C. 2023. *Perception and production skills in Spanish university EFL learners: The role of the learning context*. Vigo: Universidade de Vigo. PhD dissertation.
[9] Messum, P. & Roslyn Y. 2021. Teaching students to pronounce English: A motor skill approach in the classroom. *RELC Journal* 52(1), 169–178.
[10] Monpean, J. A. & Fouz-González J. 2021. Phonetic symbols in contemporary pronunciation instruction. *RELC Journal* 52(1), 155–168.
[11] Calvo-Benzies, Y. J. 2016. *The teaching and learning of English pronunciation in Spain. An analysis and appraisal of students' and teachers' views and materials.* Santiago de Compostela: University of Santiago de Compostela. PhD dissertation.
[12] Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
[13] Council of Europe. 2018. *Common European Framework of Reference for Languages: Learning, teaching, assessment - Companion volume*. Strasbourg: Council of Europe. Available at www.coe.int/lang-cefr.
[14] Decree 156/2022, of 15 September. Diario Oficial de Galicia, 26 September 2022, no. 183, 50010–50542. https://www.xunta.gal/dog/Publicados/2022/20220926/AnuncioG0655-190922-0002_es.html (17 Nov, 2024.)
[15] Decree 157/2022, of 15 September. Diario Oficial de Galicia, 26 September 2022, no. 183, 50543–51110. https://www.xunta.gal/dog/Publicados/2022/20220926/AnuncioG0655-190922-0003_es.html (17 Nov, 2024.)
[16] Lindade, Carlos J. 2022. *A comprehensive analysis on how English language teaching coursebooks promote teaching pronunciation in Portuguese public schools*. Vigo: University of Vigo, PhD dissertation.

17. On featural bias and consistency in MFCC representations of acoustic distance

Charles Redmon
Department of Language and Linguistics, University of Essex

While much of the research in acoustic phonetics has focused on the study of discrete parameters with a clear articulatory or auditory link to phonological features such as voicing, sibilance, or nasalisation, the use of a more theoretically agnostic parameterisation, such as via Mel-frequency cepstral coefficients (MFCCs) has played a similarly critical role in some areas of speech analysis, such as in automatic speech recognition [2, 6]. This latter use has largely been motivated by practical concerns with an eye toward prediction; however, more recently researchers have begun to use MFCCs in an inferential way as an approximate measure of phonetic distance [see, for example, 1, 3, 4]. However, despite their wider practical use and increasing application in a range of speech science research, there remain critical gaps between our use of the MFCC parameterisation and our understanding of the degree to which it reflects different underlying phonetic or phonological distinctions used in the encoding of the speech signal.

This study uses two large speech databases (1F, 1M) of single word productions from the MALD project [5] to test how different contrasts are reflected in measures of acoustic distance based on MFCC representations of the signal. In particular, 6500 words were selected from each database, which comprise over 95% of the cumulative frequency of the set (based on measurements from several different written and spoken corpora; see [5] for details). Among these items 7542 minimal pair distinctions were identified (here we restrict contrasts to those with the same phonological length) and compared for acoustic distance along several different parameterisations of MFCCs. The HTK parameterisation [6] was utilised as a baseline upon which further testing of the effects of frequency range (upper-limit: 8, 10, 12 kHz), step size (5, 10, 15 ms), and window size (15, 25, 35 ms) was done, as well as the effect of inclusion of *delta* and *delta-delta* coefficients.

All models had the mean log amplitude of the power spectrum included in the coefficient set. MFCC matrices were then aligned using dynamic time warping (DTW), and then compared via the aggregate normalised distance from the DTW alignment. For simplicity, the default parameterisation used in the *dtw* R package (Euclidean distance, fixed edges, symmetric alignment, no windowing) was used for the alignment of signals, though in a future study the impact of different parameterisations of the alignment algorithm will also be tested (alongside wider analysis of how temporal information is encoded; e.g., in hidden Markov model, HMM, or long short-term memory, LSTM, states). In total 50 unique MFCC-DTW alignments were run and analysed.

Two main analyses of the acoustic distance measures according to different signal parameterisations were done: (1) a comparison of the distances measured for different featural distinctions; and (2) a comparison of the relative stability of distance measures by contrast for different parameterisations.

Figure 1 shows the mean and variance (+/- 1σ) of acoustic distance measures for different featural distinctions as derived from the baseline, delta, and delta-delta models. Figure 1 illustrates the expected result that those contrasts that have greater spectral information are better reflected in the MFCC parameterisation than are contrasts that rely to a greater extent on temporal information (e.g., noise duration) or dynamic spectral information (e.g., formant transitions). In terms of general effects of parameterisation, the frequency range has little apparent effect, and even declines slightly at 12 kHz (*norm. dist.* = 37.9) as compared to 8 (*d* = 38.4) and 10 kHz (*d* = 38.3). Window size has a greater effect, predictably showing reduced discriminability at larger sizes (from 39 to 37.5 from 5 to 20 ms), while on the other hand discriminability increases with larger step sizes (from 37.5 to 39 from 15 to 35 ms). These results are then discussed in the wider context of encoding and acoustic modeling, including relating to alternative approaches, such as self-supervised learning, which have different spectro-temporal properties and constraints.

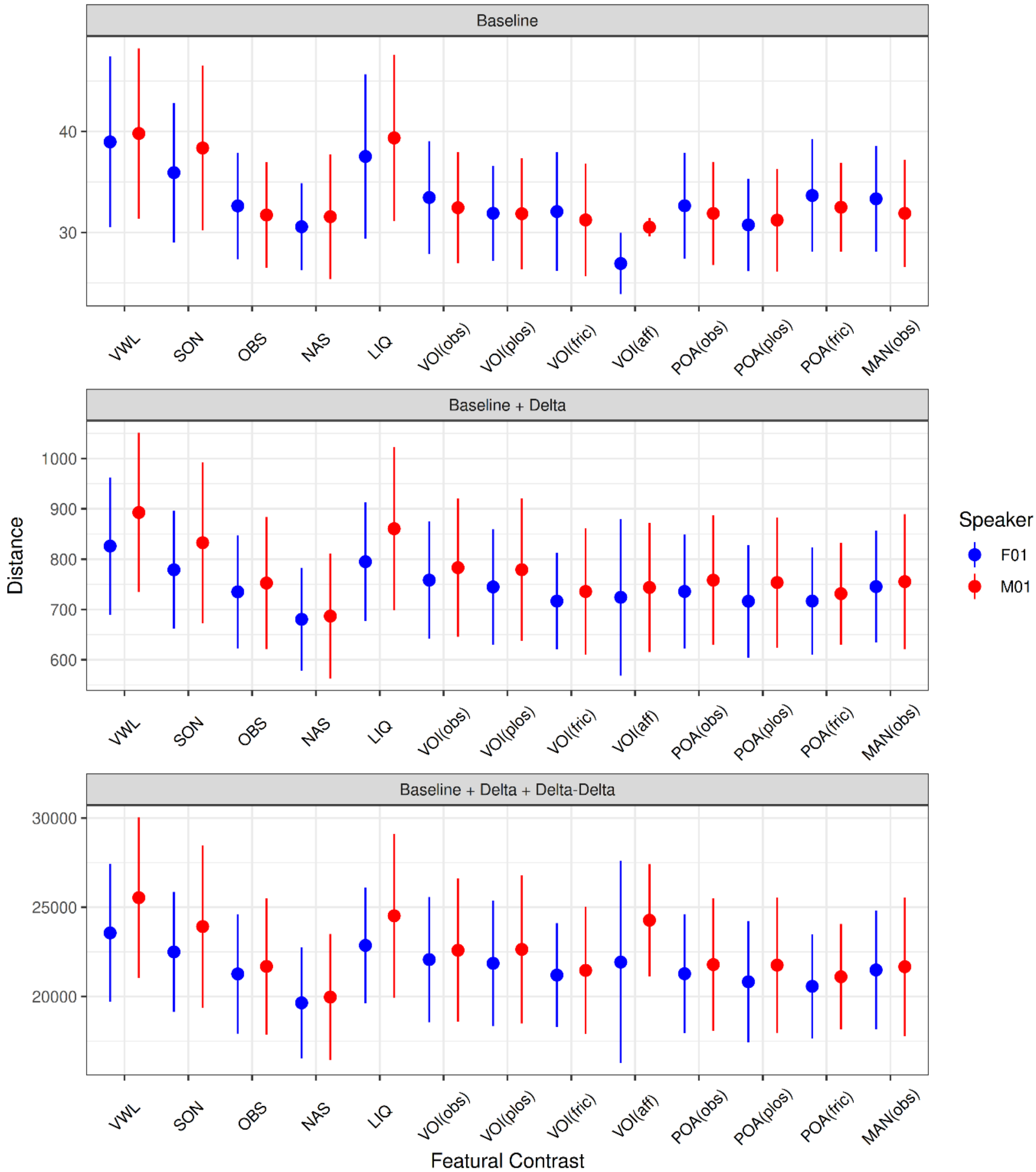


Figure 1. *Normalised distance between minimal pairs of different featural distinctions in the female (F01) and male (M01) lexical databases (point ranges represent +/- 1σ).*

[1] Bartelds, M., Richter, C., Liberman, M., and Wieling, M. (2020). A new acoustic-based pronunciation distance measure. *Front. Artif. Intell., 3*(39).

[2] Kelly, A.C., & Gobl, C. (2011). A comparison of mel-frequency cepstral coefficient (MFCC) calculation techniques. *Journal of Computing, 3*(10), 62-66.

[3] Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua, 122*, 145-163.

[4] Miodonska, Z., Marcin D. Bugdol, M.D., & Krecichwost, M. Dynamic time warping in phoneme modeling for fast pronunciation error detec*tion. Computers in Biology and Medicine, 69*, 277–285.
[5] Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The massive auditory lexical decision (MALD) database. *Behavior research methods*, *51*, 1187-1204.
[6] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., ... & Woodland, P. (2002). The HTK book. *Cambridge University Engineering Department, 3*(175), 12.

18. Finding the Human Voice in AI: Insights on the Perception of AI-Voice Clones from Naturalness and Similarity Ratings

Julia Schwarz[1], Linda Bakkouche.[1], Charles McGhee[1], Emily Lau[1], Stephan Cooper[1], Xinbing Luo[1], Madeleine Rees[1], Kai Alter

[1]University of Cambridge, [2]Newcastle University

AI-generated voice clones are becoming an indispensable tool in educational and clinical applications [1]. How human listeners perceive AI-Voice clones, however, remains largely untested. This may be particularly important when AI-generated voices fall short in replicating natural human prosody. We address this gap by comparing listeners' perception of natural human speech, three state-of-the-art AI-Voice models, and speech with deliberately reduced F0 variation. We present findings from two behavioural tasks, focusing on listeners' ratings of naturalness and similarity (MUSHRA tests: Multiple Stimuli with Hidden Reference and Anchor). The stimuli were three second samples of speech, categorised into six conditions: **Natural human speech** produced by three different speakers (two female and one male); three AI-voice clones based on the natural voices (**Elevenlabs, Style TTS-2, XTTS-v2**) [2], [3], [4]; a controlled prosodic manipulation retaining **30% F0** variation of the natural sample, to tap into the specific role of F0 variation on human perception [5], [6]; and a control condition that was designed to be the most unnatural/dissimilar condition (**Anchor**). In task 1 (Naturalness), participants were asked to judge how natural each sample sounded on a scale of 0 (very unnatural) to 100 (very natural). In task 2 (Similarity), participants were asked to judge how similar each sample sounded on a scale of 0 to 100, compared to a reference sample of the same voice producing a different sample. There were 21 trials in each task, and each trial presented the six conditions in randomised order.

The tests were implemented in Gorilla experiment builder. Data were collected via Prolific from 66 adult native English speakers without any hearing or language impairments. After passing a headphone screening test, participants completed a basic demographic questionnaire, either the Naturalness or the Similarity test (no participant completed both), a questionnaire about use of AI-generated speech, and a feedback form on task difficulty, concentration, and technical issues. Participants who failed to identify the anchor (Score=0) in more than 33% of trials, or who assigned a score of 0 to conditions other than the anchor in more than 33% of trials, were excluded (Naturalness: N=2; Similarity: N=9). Due to deviation from normality (a large number of values around 0 and 100), the data were analysed with Friedman tests followed by pairwise comparisons with Bonferroni correction.

The results revealed significant differences between AI-Voice models in the Naturalness task ($\chi^2(5)=131.1$, $p<.001$) and in the Similarity task ($\chi^2=117.1$, $p<.001$; **Figure 1**). In the Naturalness task, ratings for the Elevenlabs samples were high (*M(SD)*=86.2(18.3), *Mdn*=95), showing no statistically significant difference to natural human speech ratings (*M(SD)*=82.5(23.0), *Mdn*=95; W=101, p=.312). Conversely, XTTS and StyleTTS were rated significantly lower compared to natural human speech (p<.001). A similar pattern of results was found in the Similarity task. Interestingly, reducing F0 variation to 30% also resulted in significantly lower ratings on both tasks, suggesting that reducing intonational variation results in speech that is not only perceived as less natural (*M(SD)*=42.7(27.1), *Mdn*=35), but also less similar to speech derived from the same speaker (*M(SD)*=43.4(29.5), *Mdn*=40).

Our study provides insight into listeners' perception of three state-of-the-art AI-Voice clones. Among the models evaluated, only the Elevenlabs model was indistinguishable from natural human speech for listeners. The findings advance our understanding of naturalness and authenticity in AI speech, and inform the next phase of our research line which investigates the neural encoding of F0 variation and AI-Voice clones. Characteristics of speaker voices, individual differences, and implications for AI-powered applications will be discussed.

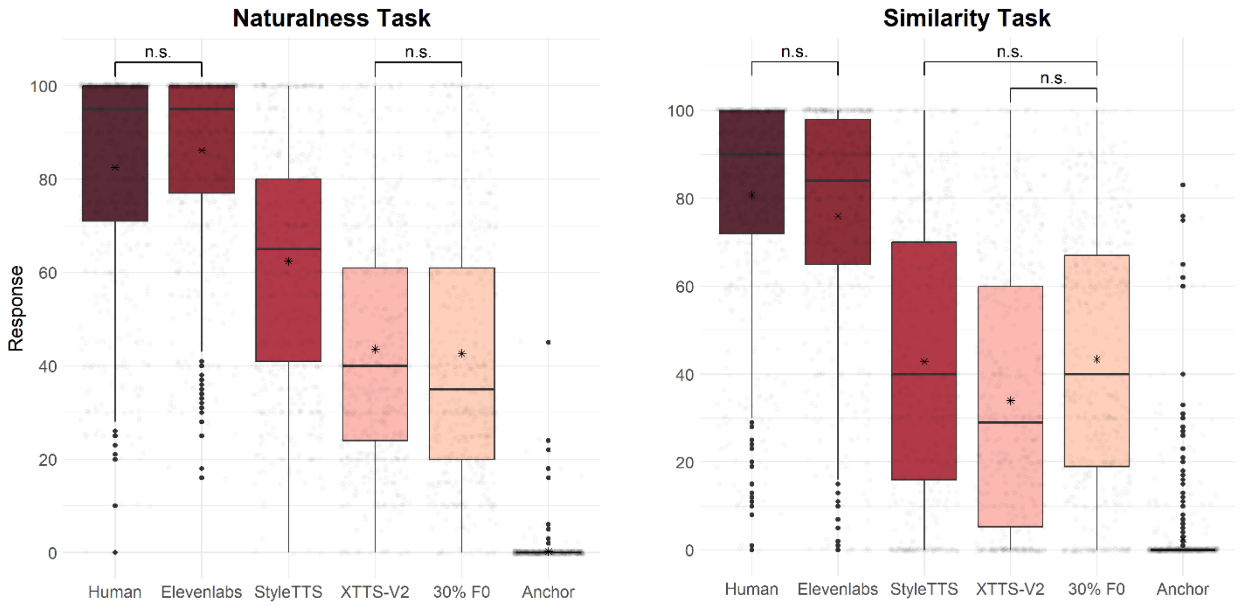


Figure 1. *Responses (Ratings 0-100) in the Naturalness Task (left panel) and Similarity Task (right panel) across Conditions. Response means are indicated by asterisks. All pairwise comparisons significant at p < .05 unless indicated otherwise (n.s.).*

[1] Zou, B., Reinders, H., Thomas, M., & Barr, D. (2023). Editorial: Using artificial intelligence technology for language learning. *Frontiers in Psychology, 14*, 1664–1078.

[2] Elevenlabs. 2024. Voice Cloning. https://elevenlabs.io/voice-cloning

[3] Li, Y. A., Han, C., Raghavan, V., Mischler, G., & Mesgarani, N. (2023). StyleTTS 2: Towards human-level text-to-speech through style diffusion and adversarial training with large speech language models. *Advances in Neural Information Processing Systems, 36*.

[4] Casanova, E., Davis, K., Gölge, E., Göknar, G., Gulea, I., Hart, L., ... & Weber, J. (2024). XTTS: a Massively Multilingual Zero-Shot Text-to-Speech Model. *Interspeech 2024*, 4978–4982.

[5] Boersma, P., & Weenink, D. 2008. Praat: doing phonetics by computer. Computer program.

[6] Corretge, Ramon (2012-2024). Change pitch median and variation. Praat Vocal Toolkit. https://www.praatvocaltoolkit.com

19. The inner composition of diphthongs: First insights from Tyrolean

Marianne Pouplier[1], Aleese Block[1], Alessandro Vietti[2]
[1]Institute of Phonetics, LMU Munich, Germany, [2]Alpine Laboratory of Phonetic Sciences, University of Bozen-Bolzano, Italy

In our current study, we investigate the diphthong inventory of Tyrolean, a variant of German belonging to the Southern Bavarian dialect group and spoken in Austria and the province of South Tyrol in Northern Italy. The larger goal of the current study is to further our understanding of the compositionality of diphthongs. Diphthongs are often described as sequences of independently available monophthongs, or, opposingly, as dynamic, independent entities of the vowel system. There has thus been a long-standing debate on the whether the onset/offset points, the transition of a diphthong, or both should be considered its defining characteristic. Thereby the onset frequency is often argued to be of primary importance [1-4]. Tyrolean is well suited to further our understanding of diphthongs since it features, as typical for Southern Bavarian dialects, a rich diphthong inventory.

We collected data from 35 native speakers of Tyrolean as spoken in the city of Meran, Italy. In a first step we provide a phonetic description of the present-day vowel inventory, since the last published work on this dialect dates back to 1936 [5]. For our corpus, we collected 4-6 words for all vowels in stressed position in laboratory speech sentences, rendering in total more than 8000 tokens across speakers and vowels. We measured the first two formants using FastTrack [6] at 20 and 80% of the vowel duration for the diphthongs and at 50% of a vowel's duration for the monophthongs.

We confirm that the present-day Meran variant features nine core diphthongs, traditionally transcribed as /ai, ia, au, ua, ea, oa, ei, ou, ui/ [5]. For this presentation, we will focus on the opening and closing diphthongs only, setting /ui/ as the only fronting diphthong aside. There are fourteen contrastive monophthongs in the form of short and long /a, ɒ, e, ɛ, o, i, u/. As a first step we determine the relationship of diphthong onset and offset to the monophthong inventory. Figure (1) gives the vowel inventory by gender, separately for opening and closing diphthongs. Our Pillai-score based analyses suggest that the traditional transcription represents the diphthong onsets fairly accurately, but the offsets are less well captured by the transcription. In particular, /ia, ua/ cover a much shorter formant distance between onset and offset than expected and are not symmetric to /ai, au/ as the transcription would suggest.

In order to test whether the unexpected offset qualities may be due to durationally conditioned truncation of the later part of the diphthong, we test whether duration predicts formant variation at onset and offset of the diphthong. While duration is a statistically significant predictor of formant variation within a given diphthong, the effect size is very small (3-8% variance accounted for per vowel). This suggests that the unexpected vowel quality at diphthong offset is not just durationally conditioned undershoot. When comparing the by-speaker standard deviation of the second formant at diphthong onset and offset, there is a statistically significant interaction between measurement point (onset, offset) and diphthong. This means that it is diphthong specific whether either onset- or offset frequency is more variable, or whether they are equally variable. For example, for /ou, ua/ the onset frequency is more variable, while for /au, oa/ it is the offset frequency. This does not globally support the hypothesis of the onset frequency of a diphthong being more stable than the offset frequency. We further plan to analyze how the naturally occurring durational variation in our data affects the dynamics of the formant transition.

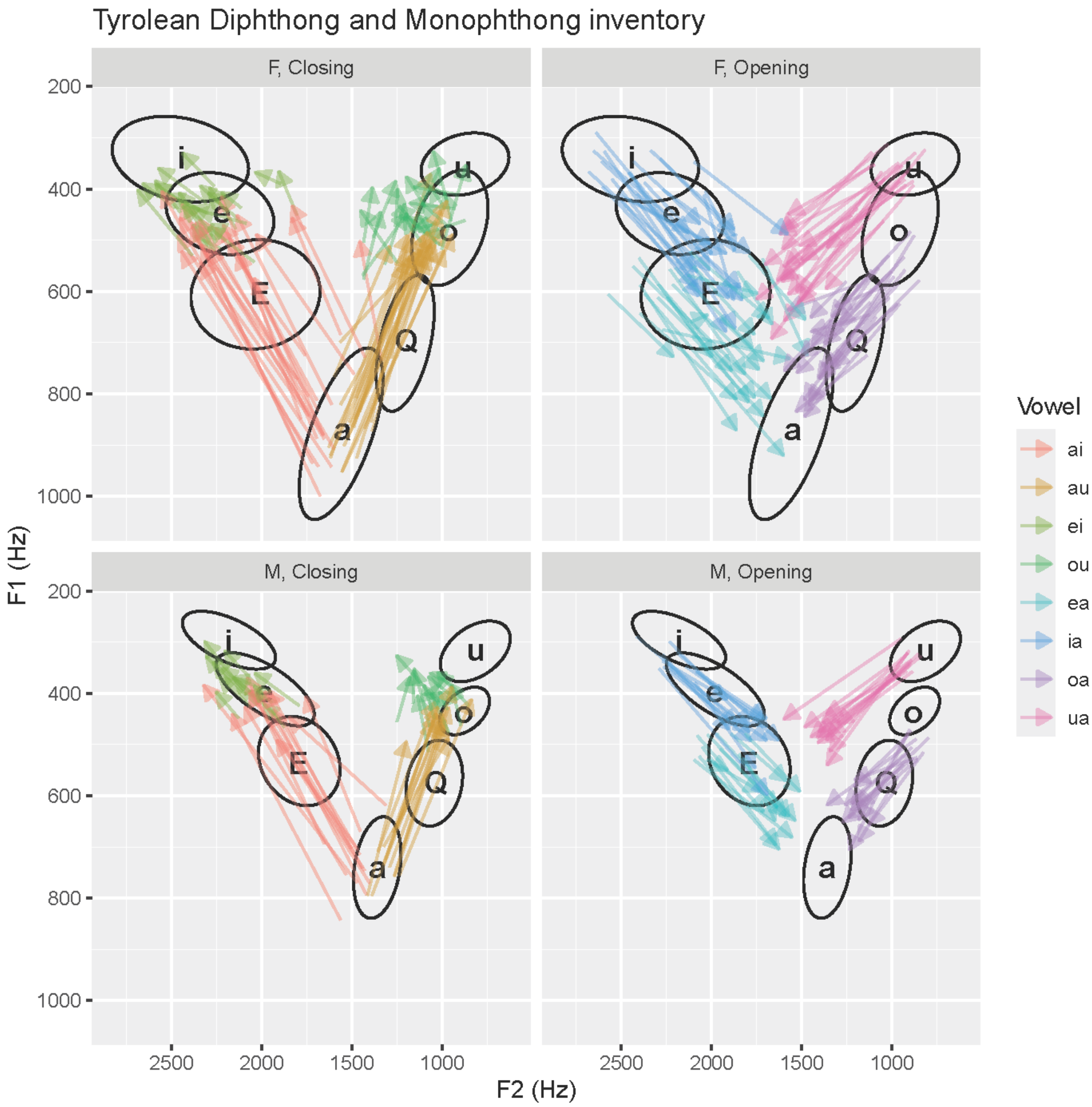


Figure 1. *Vowel inventory of Tyrolean by gender and diphthong type (closing/opening). The ellipses represent the monophthongs across speakers. Each arrow is a given speaker's mean diphthong onset/offset. F stands for female, M for male speakers.*

[1] Gay, T., Effect of Speaking Rate on Diphtong Formant Movements, *The Journal of the Acoustical Society of America*, 1968. 44(6): p. 1570-1573.

[2] Lubker, J. and T. gay, Anticipatory labial coarticulation: Experimental, biological and linguistic variables. *The Journal of the Acoustical Society of America*, 1982. 71(2): o. 437-448.

[3] Petersen, S., *Accounting for diphthongs: duration as contrast in vowel dispersion theory*. PhD Dissertation, Georgetown University, Washington D. C., USA, 2018.

[4] Xu, A, et al, Computational models for articulatory learning of English diphthongs: One dynamic target vs two static targets. *Proceedings of the 20^th^ International Conference of the Phonetic Sciences, Prague*, 2023: p. 453.

[5] Insam, M, *Der Lautstand des Burggrafenamtes von Meran*. 1936, Leipzig: Hirzel.

[6] Barreda, S., *Fast Track: fast (nearby) automatic formant-tracking using Praat*. Linguistics Vanguard, 2021. 7(1).

Poster Session 3a

1. Non-word repetition tasks: Should we consider children's articulation errors?

Benjamin P. Lange[1], Eugen Zaretsky[2], Christiane Hey[2]

[1] IU International University, Berlin, Germany, [2] Marburg University Hospital, Marburg, Germany,

Introduction: Non-word repetition tasks (NWRT) are widely used in language tests for the assessment of the phonological short-term memory (PSTM). The rationale behind the inclusion of such tasks in language tests consists in the associations between deficits in PSTM and various language(-related) impairments and disorders such as Developmental Language Disorder [1–5]. However, children's performance in NWRT does not necessarily reflect their PSTM capacity due to linguistic interferences from articulation and, more seldom, vocabulary skills [6]. It remains unclear why manuals of many language tests assessing PSTM (e.g., German "Sprachscreening für das Vorschulalter"/"Language Screening for the Preschool Age"/SSV [7] and "Kindersprachscreening"/Language Screening for Children"/KiSS.2 [8]) do not differentiate between articulation errors and PSTM errors in NWRT. In the current study, children's performance in two NWRT, those in SSV and KiSS.2, were analysed separately, with and without consideration of children's articulation skills in the German language, in order to quantify associations between NWRT performance and children's language-related medical issues.

Methods: A total of 1,772 four-year-old German children were tested with the language tests SSV and KiSS.2 in German daycare centers. SSV contains subtests on the repetition of non-words and sentences, KiSS.2 on speech comprehension, vocabulary, articulation, grammar as well as repetition of non-words and sentences. Correlations were calculated between children's performance in NWRT and some items on medical issues in KiSS.2 questionnaires for parents and daycare center teachers as well as with the dichotomized SSV and KiSS.2 results (pass/fail), first, under consideration of articulation errors (ART) and, second, without consideration of such errors (NART).

Results: Total scores of correct answers in ART were more closely associated, compared to NART, with questionnaire items such as "Has the child any language(-related) impairments or disorders?", "Did your child begin to speak his/her first language later than other children (the first word after the second birthday or later)?", "Are there any cases of dyslexia in the family?", "…of late language onset in the family (the first word after the second birthday or later)?", and with dichotomized KiSS.2 and SSV results. All associations were weak. There were no significant associations between NWRT and some other questionnaire items such as occurrence of language disorders in the family or children's permanent hearing disorders. Total scores of correct answers in the KiSS.2 articulation subtest correlated higher with ART than with NART. Total scores of correct answers in NART were significantly higher than those in ART.

Discussion: Since many four-year-old children still make errors in the articulation due to sigmatism and shetism, non-consideration of articulation errors influences total scores of correct answers in NWRT significantly and thus distorts the evaluation of children's PSTM. However, this distortion seems to be helpful in the identification of language-related medical issues, probably because many of such issues are more closely associated with the deficient articulation than with the deficient PSTM.

[1] Schwob, S., & Skoruppa, K. (2022). Detecting Developmental Language Disorder in monolingual and bilingual children: Comparison of language-specific and

crosslinguistic nonword repetition tasks in French and Portuguese. *Journal of Speech, Language, and Hearing Research, 65*(3), 1159−1165.
[2] Kiese-Himmel, C., & Nikisch, A. (2015). Diagnostische Genauigkeit einer AVWS-Testbatterie zur Klassifikation auffälliger Kinder [Diagnostic accuracy of a test set to classify children with Auditory Processing Disorders (APD)]. *Laryngorhinootologie, 94*(6), 373−377.
[3] Baird, G., Slonims, V., Simonoff, E., & Dworzynski, K. (2011). Impairment in non-word repetition: A marker for language impairment or reading impairment? *Developmental Medicine and Child Neurology, 53*(8), 711−716.
[4] Nadig, A., & Mulligan, A. (2017). Intact non-word repetition and similar error patterns in language-matched children with autism spectrum disorders: A pilot study. *Journal of Communication Disorders, 66*, 13−21.
[5] Alsaeed, S., Shalaan, S., & Alsaber, A. (2024). Non-word repetition and vocabulary in Kuwaiti Arabic-speaking children with Down syndrome and typically developing children. *Logopedics, Phoniatrics, Vocology, 24*, 1−11.
[6] Zaretsky, E., Lange, B. P., & Hey, C. (2023). The assessment of phonological short-term memory via “quasi-universal” vs. language-specific nonwords as a clinical marker for language impairments. *Studies in Applied Linguistics, 14*(2), 48−69.
[7] Grimm, H. (2017). *SSV. Sprachscreening für das Vorschulalter. Kurzform des SETK 3-5* [Language Screening for the Preschool Age. Short form of SETK 3-5] (2nd ed.). Goettingen.
[8] Holler-Zittlau, I., Euler, H. A., & Neumann, K. (2011). Kindersprachscreening (KiSS) – das hessische Verfahren zur Sprachstandserfassung [Kindersprachscreening (KiSS) – the Hessian language test]. *Sprachheilarbeit, 5*(6), 263–268.

2. A real-time analysis of aspirated /ptk/ in Yucatan Spanish

Allyson Gee, Jim Michnowicz
North Carolina State University

Yucatan Spanish (YS) is a regional variety in close contact with Yucatec Maya. YS differs from Central Mexican Spanish across a range of linguistic domains, and many regional variants are attributable to direct or indirect influence from Maya. Also, younger speakers are abandoning regional variants in favor of pan-Hispanic norms [1] At the same time, patterns based on age groups alone (apparent-time) are often ambiguous regarding the future development of these changes [2], and the application of real-time data collected at a later point in time can confirm or refute the patterns seen in apparent-time [3]. Recent studies have demonstrated how real-time data add to our understanding of ongoing language change in YS and have emphasized the need to extend real-time analyses to additional variables [4], [5]; [6].

The present study focuses on aspirated /ptk/ in YS, which stems from Maya contact [7]. Maya has both ejective and aspirated stops, and studies suggest a transfer of long-lag VOT to even some monolingual speakers [7]; [8]. The initial study compared three age groups in apparent-time (A = oldest; C = youngest), and found YS VOT to be a highly variable process, with intermediate VOT values between Maya and Central Mexican Spanish. Results by age group were mixed. While group A showed longer VOTs for /ptk/, age group was only significant for /t/, although additional significant differences were found between monolinguals and bilinguals. Author & Author (2013) suggests that future research should continue to examine YS /ptk/ to see how it develops alongside other regional variables undergoing standardization.

This study adds real-time data from 12 young speakers (Group D), collected 11 years after the data in (Author, 2013). While analysis of the real-time data is ongoing, preliminary results suggest that group D is producing lower VOT, more in-line with Central Mexican norms (Figure 1), and that this group is further consolidating the pattern of shorter VOT begun by groups B and C. Further details and conclusions will be presented, with special attention to differences based on language background.

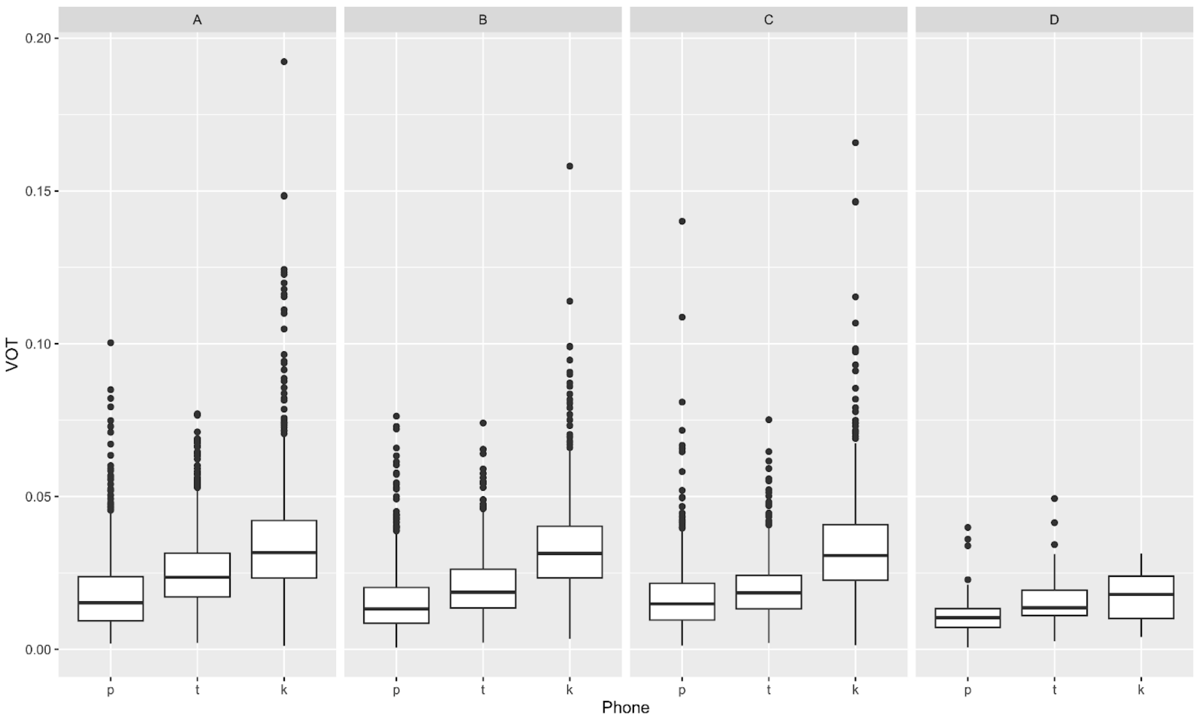


Figure 1. VOT values by age group. *Preliminary data for real-time Group D.

[1] Michnowicz, J. (2015) "Maya-Spanish contact in Yucatan, Mexico: Context and sociolinguistic implications". In S. Sessarego and M. González Rivera (Eds.). New Perspectives on Hispanic Contact Linguistics in the Americas. Madrid: Iberoamericana/Vervuert. 21-42.

[2] Michnowicz, J. (2008). Final nasal variation in Merida, Yucatan. Spanish in Context, 5(2), 278-303.

[3] Chambers, J. K. (2009). Sociolinguistic theory: Linguistic variation and its social significance. rev. ed. Oxford: Wiley-Blackwell.

[4] Michnowicz, J. (2021). Apparently real changes: Revisiting final (-m) in Yucatan Spanish. In M. Díaz-Campos (Ed.), Handbook of variationist approaches to Spanish (pp. 249-262). London: Routledge.

[5] Michnowicz, J. (2023). Real and apparent (time) changes in Yucatan Spanish: The case of /bdg/. In S. Fernández Cuenca, T. Judy, & L. Miller (Eds.), Innovative approaches to research in Hispanic linguistics: Regional, diachronic, and learner profile variation (pp. 130-151). Amsterdam: John Benjamins.

[6] Michnowicz, J., & Hyler, A. (2020). The changing accent of Yucatan Spanish. In R. Rao (Ed.), Spanish phonetics and phonology in contact: Studies from Africa, the Americas, and Spain (pp. 36-62). Amsterdam: John Benjamins.

[7] Author & Author (2013). Voiceless stop aspiration in Yucatan Spanish: A sociolinguistic analysis. *Spanish in Context*, 10(3), 410-437.

[8] Author et al. "Yucatan Spanish VOT: Bilingual phonological representations and sociophonetics". Presentation at the Linguistic Society of America (LSA) 2023 conference in Denver, CO. January, 2023.

3. Balearic *seseo*: understanding the social dynamics of the blurring of a phonemic distinction in a language-contact situation.

Andrés Enrique-Arias
Universitat de les Illes Balears (UIB)

This study looks at Spanish *seseo* (i.e. the non-distinction of /s/ and /θ/, as in standard Peninsular Spanish *cinco* 'five' [ˈθiŋko] pronounced [ˈsiŋko]) in the Spanish of Catalan-dominant speakers in the Balearic Islands. This non-canonical pronunciation is due to the lack of a /θ/ sound in Catalan; as a result, Catalan-dominant speakers with limited Spanish native input encounter difficulties in mastering the distinction of the two sibilants. In the Balearic islands *seseo* is largely confined to rural, elderly Catalan-dominant speakers with only primary education.

Although Balearic *seseo* has been mentioned in previous studies (Moll 1961: 470; Serrano Vázquez 1996-97a: 1019; 1996-97b: 376, Blas Arroyo 2007: 84, Radatz 2008: 120) to date no detailed description or empirical study exists. With the objective of describing the distribution and frequency of use of this phenomenon, and determining the factors that favor its appeareance, this study looks at data from 26 sociolinguistic interviews with Catalan-dominant elderly speakers with only primary education from Mallorca and Minorca. The sounds appearing in the contexts in which an interdental /θ/ would be expected in the canonical Peninsular Spanish pronunciation were extracted in the first and last 10 minutes of each interview. In total, some 2700 tokens were selected and classified as alveolar or interdental realizations after an auditory analysis made by the author.

The logistic regression analysis reveals that the degree of Catalan language dominance, formulated in terms of shorter periods of formal schooling, is the main predictor for the presence of [s] in contexts of [θ]. Another relevant factor is gender: in general terms, women produce less *seseo* than men, and, when comparing the two samples of each interview—opening minutes vs. final ones—women also tend to self-correct more and accommodate to the standard pronunciation of the interviewers. In contrast, men exhibit higher rates of *seseo* overall, with a slight increase of the [s] variant in the final part of the interviews. Another factor that favors the non-distinction of the two sibilants is the lexical frequency of the lexical item where the variant occurs, with higher rates of *seseo* in less frequent lexical items.

In conclusion, Balearic *seseo* operates like other features that are characteristic of unbalanced bilinguals. Two crucial triggers for deviation from the monolingual baseline in non-dominant speakers have been identified (Polinsky and Scontras 2019): the quantity and quality of the input from which the less dominant grammar is acquired, and economy of online resources when operating in a less dominant language. *Seseo* is thus an economizing device through the blurring of a phonemic distinction; as expected, this deviation of the monolingual baseline is more frequent in lower frequency lexical items, which are harder to learn and process. Also, in consonance with these principles, *seseo* is more prevalent in speakers with less exposure to monolingual varieties of Spanish (whether it be years of formal schooling or residence in areas with a greater density of Spanish speakers). Finally, from a sociophonetic point of view, the results point to a change from above, in which [s] represents a receding, stigmatized variant (Enrique-Arias 2022: 147-48) that is being supressed as speakers make an effort to adopt the prestigious [θ] variant associated with educated speakers.

Blas Arroyo, José Luis (2007): "Spanish and Catalan in the Balearic Islands". International journal of the sociology of language 184. 79-93.

Enrique-Arias, Andrés (2022) "El castellano en Mallorca 60 años después: un estudio en tiempo real". *Revista Internacional de Lingüística Iberoamericana*, vol. 20, nº 39, 137-163.

Moll, Francisco de B. (1961): "El castellano en Mallorca". En: *Studia Philologica. Homenaje ofrecido a Dámaso Alonso por sus amigos y discípulos con ocasión de su 60° aniversario*, Vol. 2, pp. 469-474. Madrid: Gredos.

Polinsky, Maria & Gregory Scontras (2020): "Understanding heritage languages". *Bilingualism: Language and Cognition*. 23 (1). 4–20.

Radatz, Hans-Ingo (2008): "Castellorquín. El castellano hablado por los mallorquines". En: Sinner, Carsten/Wesch, Andreas (eds.): *El castellano en las tierras de habla catalana*. Madrid / Frankfurt: Iberoamericana / Vervuert, pp. 113-132

Serrano Vázquez, María del Carmen (1996-97a): "El castellano en contacto con el catalán en una comunidad escolar bilingüe". En: *Anuario de Lingüística Hispánica*, 12, pp. 1013-1028.

Serrano Vázquez, María del Carmen (1996-97b): "Rasgos fonéticos de carácter interferencial en el castellano de una comunidad bilingüe". En: *Estudios de Lingüistica de la Universidad de Alicante*, 11, pp. 365-383.

4. An acoustic and articulatory key study of unvoiced coronal oral stops in Baniwa of Içana

Jenifer Vega Rodriguez[1,2], Thiago Costa Chacon[2], Rodrigo Alexandre[2], Nathalie Vallée[1], Christophe Savariaux[1]
[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab. [2]Brasilia University

Baniwa of Içana is a North Arawak language, spoken by approximately 7,000 people [1]. This key study analyzes the central Baniwa dialect spoken in the middle Içana region in Brazil. According to [2], Baniwa has two coronal oral stops: an apico-alveolar /t/, and a lamino-dental /t̪/. As this contrast is rare cross-linguistically [3], our aim is to describe their production through articulatory analyses using both palatography and linguography, and through acoustic measurements of the F_2 transitions and VOTs. This study is the initial phase of an ongoing investigation that involves more acoustic and articulatory measures such as loci, F_1 transitions, palatographic and ultrasound data, and acoustic-perceptual tests.

This pilot study involved a 25-year-old native speaker of the central Baniwa, born in Tunui Cachoeira, Içana, Brazil, who speaks Brazilian Portuguese as a second language. Palato-linguographic data were collected using 16 target words with /t/ and /t̪/ in stressed syllables, within #CV and VCV contexts, including all vowels /i e a o̞/ [1]. To avoid any superposition of articulatory traces on that of the target segment, the remaining syllables contained extrabuccal consonants. A palatographic mixture of activated carbon, olive oil, and chocolate was applied to the tongue and upper palate (respectively). Each word was repeated 4 times. Images were captured using a high-resolution smartphone and a mirror with a millimetre-scale ruler. Acoustic analysis involved embedding target words in a frame sentence, recorded four times. Praat [4] was used to segment the target plosive and the post-consonant vowel, and to measure F_2 every 5ms, from 10ms before the boundary between the plosive and the vowel, until 20 ms from the start of the vowel. Thus 7 F_2 measurement points were recorded for the target sequence. VOT duration was segmented into Transient (T), Friction (F), and Aspiration (A) phases [5]. Investigation was conducted using [6].

Palatographic and linguographic image analyses reveal two distinct articulations. For /t/, the palatographic data (Figures 1 and 2) show an apico-lamino-dental articulation, where the tongue blade and tip seem to be at the same height during the occlusion, resulting in a broader trace. In contrast, /t̪/ is produced with a lamino-pre-alveolar articulation, with the tip lower than the blade, resulting in a narrower trace. These patterns are consistent across all vowels and contexts (#CV; VCV), with /t/ exhibiting a more distributed occlusion compared to /t̪/. These findings differ from [2]'s descriptions.

Acoustic analyses revealed that in the #CV and VCV positions, F_{2mean}=1.6kHz and 1.8kHz at voicing onset, in the /t/ and /t̪/ contexts respectively (Figure 3). A less distributed constriction for /t̪/ suggests a larger anterior cavity than for /t/, resulting in a lower F_2. VOT analyses show significant differences only in VCV. The 3 phases of VOT show difference in VCV: A and F are longer for /t̪/.

This pilot study is the first phase of an ongoing investigation that will include detailed articulatory and acoustics-perceptual analyses of this speaker before proceeding to acquire additional data from other male and female speakers.

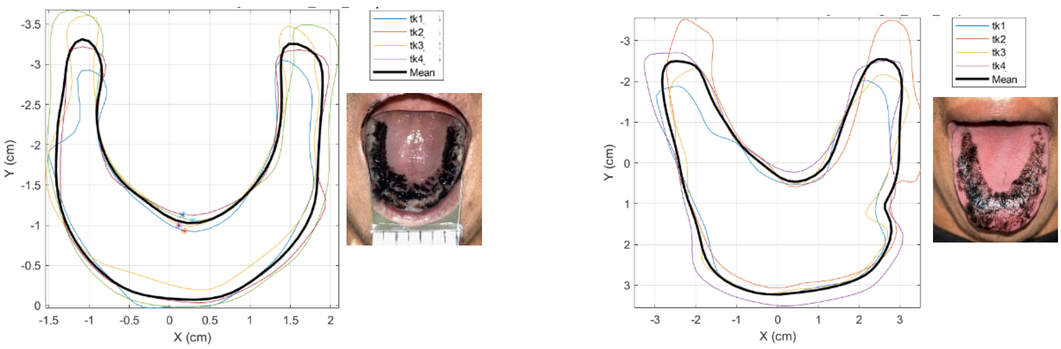


Figure 1. *Palatography(left) and linguography (right) of /t/ in the word /taapa/ 'nightfall'. Traces of 4 repetitions were calculated semi-automatically using an internal MATLAB script.*

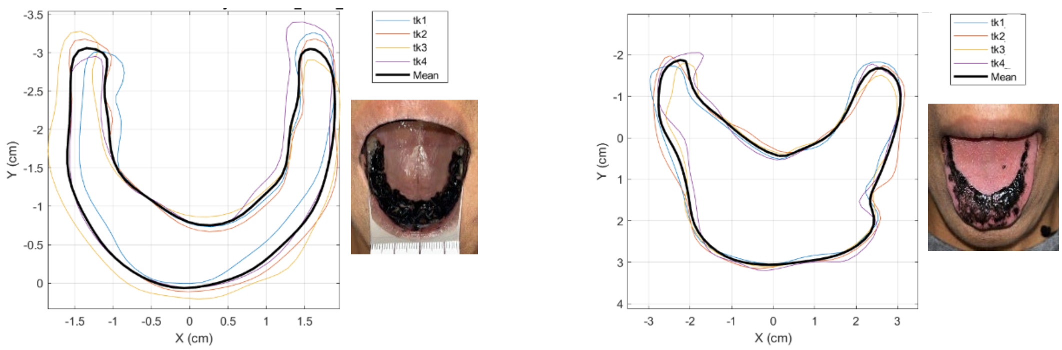


Figure 2. *Palatography(left) and linguography (right) of /t̪/ in the word /t̪amome/ 'light'. Traces of 4 repetitions were calculated semi-automatically using an internal MATLAB script*

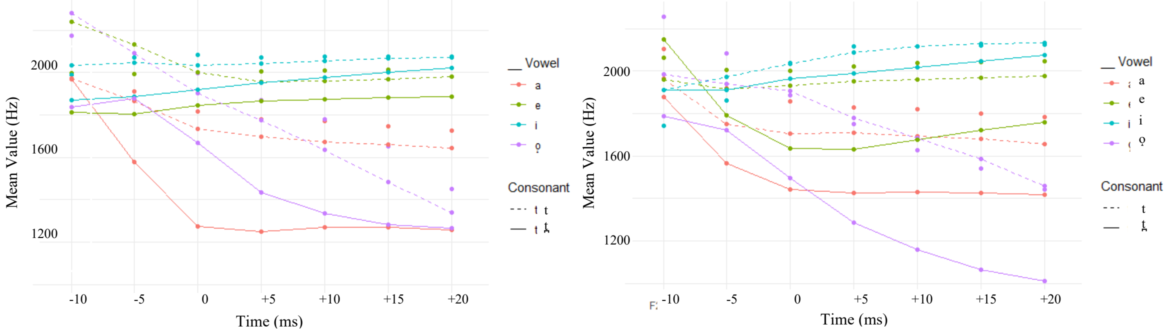


Figure 3. *Mean F_2 values of vowels by consonant type. Contexts VCV (left) and #CV (right).*

[1] D. M. Eberhard, G. F. Simons, and C. D. Fennig, *Ethnologue: Languages of the World*, 26th ed. Dallas, TX: SIL International, 2023.

[2] H. Ramirez, Uma gramática do Baniwa do Içana. São Paulo, Brazil: Universidade de São Paulo, 2001.

[3] P. Ladefoged, "The relationship between phonetics and phonology," in Phonological Theory: An Introduction, 1st ed., J. A. Goldsmith, Ed. Oxford: Blackwell, 1996, pp. 105–115.

[4] P. Boersma, "Praat, a system for doing phonetics by computer," Glot International, vol. 5, no. 9/10, pp. 341-345, 2001.

[5] K. N. Stevens, Acoustic Phonetics. Cambridge, MA: MIT Press, 2000.

[6] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2021.

5. Variability in the speech signal defines the rhythm of spoken language

Mireia Marimon[1], Alan Langus[2]
[1]*Center for Brain and Cognition, Department of Engineering, Universitat Pompeu Fabra, Barcelona, Spain,* [2]*Department of Linguistics, University of Potsdam, Potsdam, Germany*

Evidence for rhythm in spoken language comes primarily from experiments that present listeners with rhythmic speech stimuli. Here we argue that theories of speech rhythm must account for at least two types of variability. First, they must explain why listeners tend to rely more on syllabic regularities in the so-called syllable-timed languages like French and on prosodic regularities in so-called stress-timed languages like German and English [1,2]. Second, they must also show listeners can perceive rhythm in natural speech where the timing of speech units, including phonemes, syllables and stress intervals, depend on non-rhythmic structure of sentences [3].

One promising approach to explaining rhythm perception in natural languages are theories of neural entrainment. When listening to spoken language, the brain tracks temporal regularities at multiple timescales corresponding to basic speech units such as syllables (Theta band 4.5-7Hz) and speech prosody (Delta band, 1-3.5Hz) [4]. By tracking linguistic units across a range of narrow frequency bands could explain how phonemes, syllables, and prosodic constituents of varying length can be processed by the same neuronal populations. However, it remains unknown how neural entrainment is modulated by cross- and within-language variability. Because neural entrainment is thought of as a functional principle of auditory perception in general, it also remains unclear how it relates to rhythm perception.

Here we present a novel method that measures spontaneous sensorimotor synchronization between natural speech and listeners' changes in pupil size. German-speaking adults (N=30) listened to 18 sentences in French, Italian, Polish, Dutch (twice in random order) from Ramus et al. (1999) [5]. They were instructed to just listen to the stimuli and fixate on the screen. We hypothesized that if a language falls towards the syllable-timed end of the rhythmic continuum (Italian, French), listeners' pupils would synchronize to the speech amplitude envelope better in the Theta (4.5-7Hz) frequency band that corresponds to the durations of syllables in the speech signal. In contrast, if the language falls towards the stress-timed end of the continuum (Dutch, Polish), listeners' pupils should entrain better in the Delta (1-3.5Hz) frequency band, that corresponds to the durations of prosodic regularities.

We calculated the Phase-Locking Value (PLV) between the amplitude envelope and the pupillary response in the two frequency bands. We show that pupillary synchrony is better at syllable frequencies in theta in French and for prosodic frequencies in delta in Dutch (Figure 1). We further show that pupillary synchrony in Theta and Delta bands correlates with duration variability between successive peaks in the amplitude envelope in the respective frequency bands: When variability is small listeners track syllables and when variability is larger they track preferentially prosody (Figure 2). Our results show that cross-linguistic differences in speech rhythm perception - as evidenced by sesnori-motor synchronization between the speech and pupillary signal - emerge from duration variability within a language.

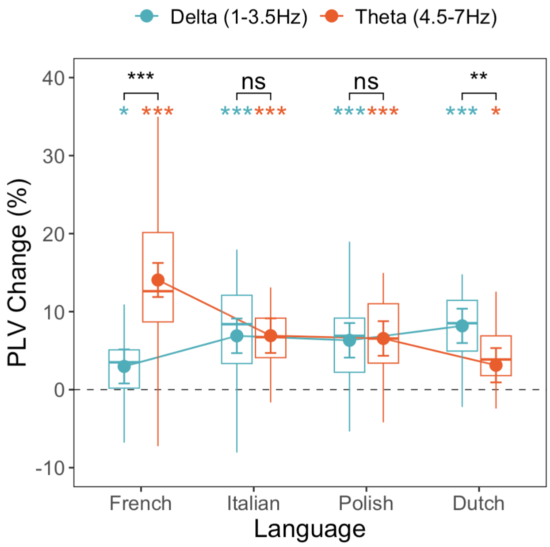


Figure 1. *Pupillary synchrony to French, Italian, Polish and Dutch sentences at Theta (red) and at Delta (blue) frequencies. Black significance levels correspond to comparisons between theta and deltabands and color-coded significance levels to comparisons against chance.*

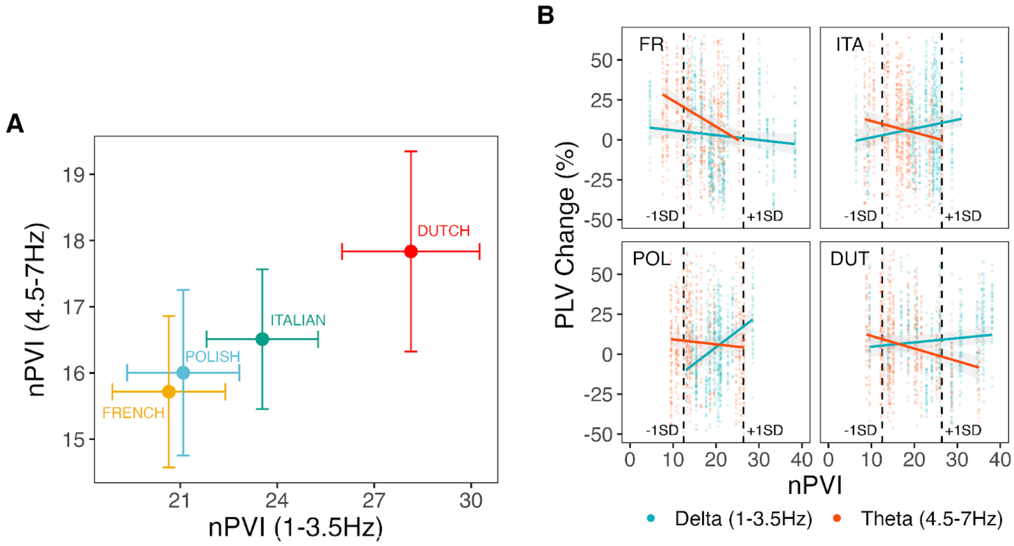


Figure 2. *Acoustic variability influences pupillary synchrony to natural speech. (A) Normalized Pairwise Variability Index (nPVI) of amplitude envelope peaks in Delta and Theta frequency bands. (B) nPVI predicts PLV between the amplitude envelope and the pupillary signal in Theta and Delta frequency bands.*

[1] Nazzi T., Iakimova I., Bertoncini J., Frédonie S., Alcantara C. (2006). Early segmentation of fluent speech by infants acquiring French: emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54, 283–299.

[2] Marimon, M., Langus, A. & Höhle, B. (2024). Prosody outweighs statistics in 6-month-old German-learning infants' speech segmentation. *Infancy*.

[3] Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control* .. Oxford University Press.

[4] Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511-517.

[5] Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265–292.

6. Can error-driven discriminative learning explain transitional probability learning during listening to running speech?

Jessie S. Nixon[1], Erdin Mujezinović[2], Ruben van de Vijver[2]
[1]*University of Bielefeld,* [2]*Heinrich Heine University of Düsseldorf*

In spoken language, speech consists of a more-or-less continuous stream of acoustic information, without breaks between words. This observation raises the question of how listeners are able to segment speech into words. In seminal work, Saffran and colleagues showed that adults[1] and infants[2] are able to use information about the statistical structure of syllable sequences to determine likely words. They presented participants with a set of six trisyllabic sequences ('words') combined into a running speech stream[1]. In a two-alternative forced-choice task following a short training with the speech stream, adults were presented with the trisyllabic words and 'nonwords', i.e. an 'illegal string' of the same syllables. Participants selected the words over the nonwords significantly more often than chance. Furthermore, accuracy among the six words was greater for the words with a higher word-internal transitional probability, compared to those with a lower transitional probability.

This work took the language acquisition world by storm, as it provided a means by which infants might solve the complex problem of word segmentation by learning from the input. Often termed 'transitional probability learning', this paradigm is still used today[3] and has become influential in many areas of cognitive science. Yet few studies have examined the underlying *learning mechanisms*. It was recently shown that motor sequence learning is better predicted by *error-driven learning* than either n-gram frequency or transitional probability[4]. In the present study, we investigate whether error-driven learning also predicts sequence learning during listening to a stream of spoken syllables.

Here we present simulations which will be used to predict participant response accuracy in our experiment, based on and using the same stimuli as [1]. Instead of a forced-choice task at the end of the experiment, participants were probed for their responses at pseudorandom intervals throughout the experiment: they were asked to type the next syllable in the sequence. We simulated the experiment using the Rescorla-Wagner equations. Syllable bigrams (cues) were used to predict upcoming single syllables (outcomes) in a moving window across the experiment. On each trial, the model adjusts the degree to which the current syllable sequence (cues) predicts all possible upcoming syllables. Weights increase to the syllable that actually occurs and decrease to those that do not, according to the Rescorla-Wagner equations.

The error-driven learning simulation output predicts dynamic learning of the sequences over the experiment (Figure 1). This is in contrast to Saffran and colleagues' analysis using a single static value, transitional probability. The model learns predictive cues (e.g. 'pa', 'patu') and, importantly, also *unlearns* cues that predict the absence of the outcome, e.g. 'bi' (Figure 1, right). For the outcome 'bu', which occurs in multiple words and word positions, a different pattern emerges, reflecting its different predictability (Figure 2). Furthermore, the models show differential effects due to cue competition, even among cues for which the transitional probability is 1. These simulations will be tested against by-trial participant responses in our experiment. We have collected data for the first of twenty participants.

This work contributes to the growing literature investigating the learning mechanisms and possible contribution of prediction error underlying language learning. In addition to the initial motivation in [1,2] to understand word segmentation, the learning mechanisms involved in sequence learning are relevant to multiple questions in phonetics and phonology. For example, error-driven learning of speech stream has previously been used to model infant phonetic learning[5] and is also relevant for the learning of phonotactics [compare 6]. With this work, we aim to gain further insight into the dynamic, trial-by-trial learning processes involved in auditory serial pattern learning.

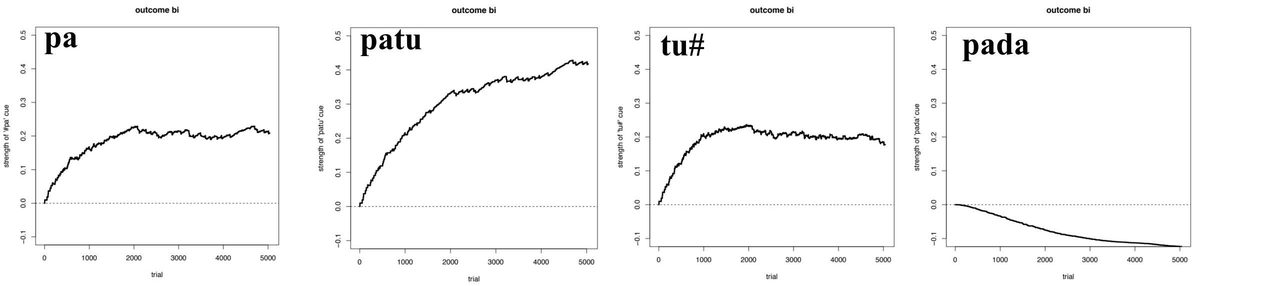


Figure 1. *Simulation, illustration 1: selection of cue weights to* ***outcome 'bi'****, which occurs only in 'patubi'. Left to right, cues '#pa', 'patu' and 'tu#' are learned as predictive cues for 'bi'. The hash symbol is used to encode temporal order: '#pa' is the first cue in the string; 'tu#' is the last. Importantly, as shown in the rightmost panel, the model also 'unlearns' cues that predict the absence of the outcome. Here, 'pada' develops negative weights to 'bi'.*

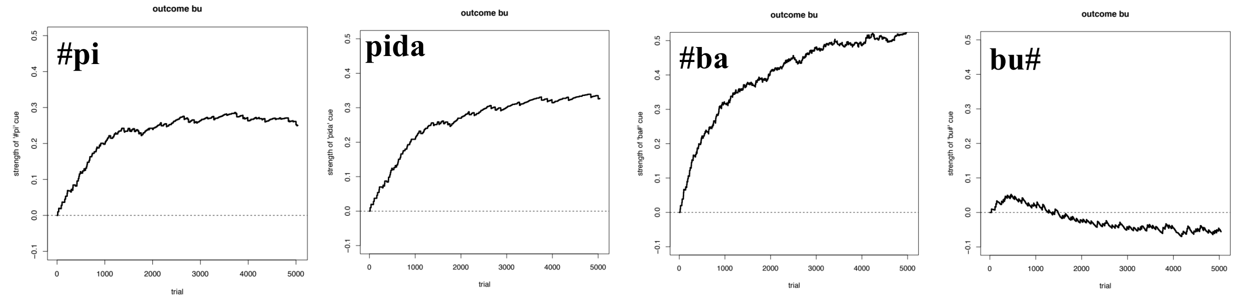


Figure 2. *Simulation, illustration 2: selection of cue weights to* ***outcome 'bu'****, which occurs in four different words, 'pidabu', 'tutibu', 'bupada' and 'babupu'. In the first and second panels, the cues '#pi' and 'pida' develop strong weights to 'bu'; these syllables occur in 'pidabu' and the result is comparable to the first and second panels for 'bi' in Figure 1. However, an even stronger cue for 'bu' is '#ba' (from the word, 'babupu'). Right: because the syllable 'bu' occurs in multiple words, it can be directly preceded by another 'bu' as a cue. Even though this occurs throughout the experiment and even though the cue 'bu#' initially increases weight as a predictor of 'bu', this pattern changes, such that cue 'bu#' eventually develops* negative *weights to the outcome 'bu'. This curious reversal is due to cue competition and the effects of previous learning on current learning.*

[1] Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. Journal of memory and language, 35(4), 606-621.

[2] Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274(5294), 1926-1928.

[3] Saffran, J. R. (2020). Statistical language learning in infancy. Child development perspectives, 14(1), 49-54.

[4] Tomaschek, F., Ramscar, M., & Nixon, J. S. (2024). The Keys to the Future? An Examination of Statistical Versus Discriminative Accounts of Serial Pattern Learning. Cognitive Science, 48(2), e13404.

[5] Nixon, J. S., & Tomaschek, F. (2021). Prediction and error in early infant speech learning: A speech acquisition model. Cognition, 212, 104697.

[6] Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. Linguistic inquiry, 39(3), 379-440.

7. Phonological contrast perception and production in bilingual speech: Intra- and inter-speaker dynamics

Caihong Weng[1, 2],Alexander Martin[2], Ioana Chitoran[1]
[1] Université Paris Cité, Paris, France, [2] University of Groningen, Groningen, Netherlands

Background: Gaining insight into individual variation in speech perception and production is of growing importance [1, 2]. The present study aims to explore this variation in a bilingual population. We examine how bilingual speakers of L1 Quanzhou Southern Min (QSM) and L2 Mandarin produce and perceive the Mandarin contrast /s/~/ʂ/. We specifically explore the variation encountered in the relation between perception and production in the same speakers. The study addresses a key question: are individual speakers who produce a stronger contrast also those who can better perceive the contrast?

Method: 60 target QSM-Mandarin bilinguals from Quanzhou, China, as well as 15 control L1 Mandarin speakers completed both a sentence reading production task and an ABX discrimination task. In sentence reading, all materials were CVCV Mandarin real words. For the ABX task, we compared perception of a control contrast [m]~[p] (present in both the L1 and L2 and taken as baseline performance) to the target contrast [s]~[ʂ] in two vowel contexts ([a] and [u]). Stimuli were CVCV nonsense words, with a high level tone.

Results: *Production*: 4 spectral moments along with the intensity of the fricative were extracted using time-averaging [3]. Utilizing the Random Forest classification method, Center of Gravity (CoG) emerged as the most important acoustic feature for distinguishing between the Mandarin [s] and [ʂ] fricatives based on the control group's contrast production. We employed mixed-effects models to investigate the effects of Target and Vowel, as well as their interaction on CoG values, while accounting for individual variability with by-participant random intercepts. The results reveal that both Target ($\chi^2(1) = 47.8$, $p < 0.001$) and Vowel ($\chi^2(1) = 86.5$, $p < 0.001$) independently affect the CoG values, while their interaction ($\chi^2(1) = 3.8$, $p > 0.05$) did not significantly influence the CoG values for the bilingual QSM speakers. We speculated that the lack of the interaction significance is due to substantial individual variability within this group. To further elucidate this variability, individual productions were evaluated through a perceptual identification task conducted by 10 L1 Mandarin listeners. Bilingual speakers were categorized into two subgroups: 16 distinctive / 44 merged for [a] and 11 distinctive / 49 merged for [u] (See fig.1), showing that a majority of the bilinguals failed to make a reliable contrast distinction in either vowel context. *Perception*: We used a logistic mixed-effect model to compare discrimination accuracy of the target contrast in each vowel context to the control contrast. Our full model included fixed effects for Contrast (control, target-[a], target-[u]) and random intercepts for the factor Participant. We found significant effects for Contrast ([ma]~[pa] vs. [sa]~[ʂa]: $z = -28$, $p < 0.001$; [ma]~[pa] vs. [su]~[ʂu]: $z = -24$, $p < 0.001$), indicating that participants had a significantly harder time discriminating the target fricatives than the control contrast (See fig.2). *Production-perception link:* We ran two linear regressions, using performance in the ABX task on the target contrast in each vowel context to predict performance on the control contrast. From these regressions, we extracted the residuals, with lower values indicating worse discrimination of the target [s]~[ʂ] contrast relative to the control contrast [4]. For each vowel context, participants' CoG difference score in production (ΔM = (M (alveolar) − M (retroflex)) was then predicted using these residuals, the speaker type (distinctive vs. merged) and the interaction of residuals and type (See fig.3). In both contexts, the "distinctive" group shows an upward trend in the correlation between ABX residuals and CoG contrast differences, while the "merged" group remains relatively flat. However, despite this visual divergence, the statistical analysis did not confirm a significant interaction effect between ABX residuals and speaker type in either context (both $p > 0.05$).

Discussion: This study found individual variability in the relationship between production and perception of the Mandarin /s/~/ʂ/ contrast, suggesting that while some bilingual QSM-Mandarin speakers maintain distinct perception-production mappings, others, particularly those in the "merged" group, show no significant link between these domains. We will discuss the potential influence of extra-linguistic factors on the perception-production link, suggesting avenues for future work on this relationship.

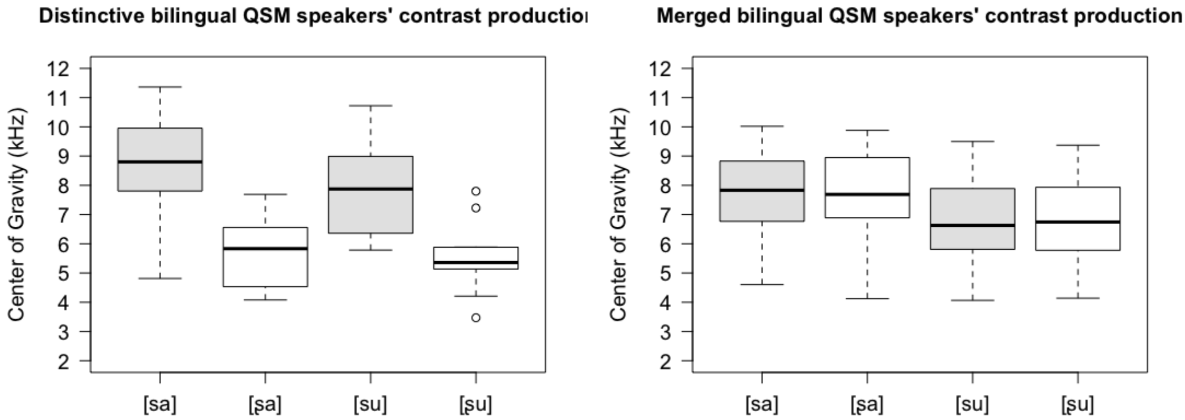


Figure 1: *Comparison of CoG values for [s] and [ʂ] in both contexts between "distinctive" (left) and "merged" (right) bilingual QSM speakers*

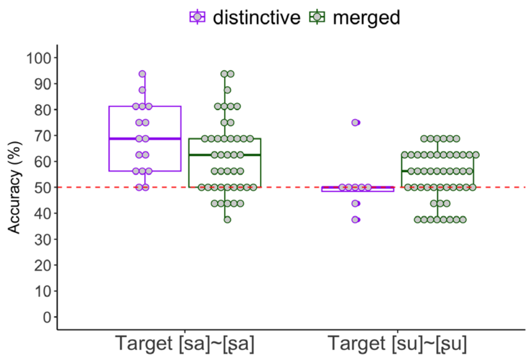


Figure 2: *ABX discrimination accuracy for control and target contrast between "distinctive" and "merged" bilingual QSM speakers*

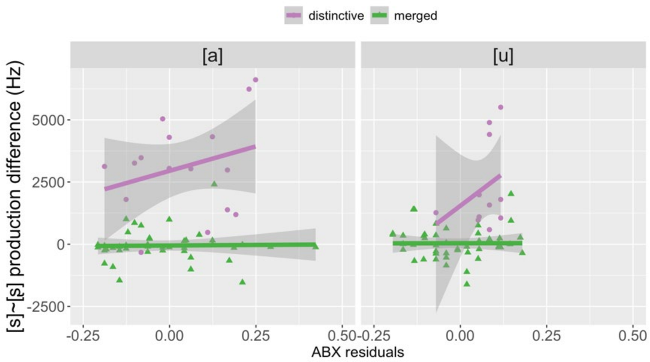


Figure 3: *Regression of contrast production difference by participant as a function of the residuals extracted from the ABX task between "distinctive" and "merged" bilingual QSM speakers.*

[1] Flege JE (1993) Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America* 93: 1589–1608.

[2] Yu, A. C. L. (2019) On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. In: *Journal of Laboratory Phonology* 10(1): 2, pp. 1–29.

[3] Shadle, C. H. (2023). "Alternatives to moments for characterizing fricatives: Reconsidering Forrest et al.(1988)," *The Journal of the Acoustical Society of America* 153(2), 1412–1426.

[4] Alexander Martin et al (2022). "Marginal contrast in loanword phonology: Production and perception". In: *Journal of Laboratory Phonology* 13.1.

8. Lenition, fortition, and word-recognition in Mawng and Iwaidja

Rikke Bundgaard-Nielsen[1, 2], Robert Mailhammer[1], Brett Baker[2], Yizhou Wang[1], Mark Harvey[3]
[1]Western Sydney University, [2]University of Melbourne, [3]University of Newcastle, Australia

Most models of continuous parsing, like Shortlist B 0, argue that speech recognition begins with the first incoming phoneme. This first segment and the following segments activate possible lexical competitors in the listener's vocabulary 0; parsing relies on both a top-down process, progressively eliminating competitors that do not match the presented material, and a bottom-up process, providing further information. Most models assume that the incoming material will be delivered in '*a form suitable for accessing stored lexical entities'* 0, and (unpredictable) variability in the initial segments of words thus poses a challenge to consider. Here, we tested speaker tolerance for word-initial phonological mutations in two Australian languages: Iwaidja 0 and Mawng 0. Both languages have grammatically conditioned initial mutations in verb and noun roots, resulting from lenition and fortition 0; see (1) and (2). These alternations are historical and lexicalised. Mawng also has synchronic external sandhi processes affecting initial segments at word boundaries: /ɡ/ is lenited to [ɣ ~ ɰ ~ w] following vowels, liquids, and glides 0, as in (3a, c), whereas only /ɡ/ is found following nasals and stops (3b, d). Iwaidja is not described with synchronic lenition of this kind, e.g., /ɡa/ in Mawng, but not in Iwaidja, might be realised as [ɡa ~ɣa ~ ɰa ~ wa] (see example (3)).

In a 2AFC experiment, we asked 11 speakers each of Mawng and Iwaidja (all bilingual in the other language as is standard in this speech community) to indicate preference for one of two sequentially presented utterances, differing only in the first segment of the target noun (15 nouns; 60 trials). Stop-initial nouns were presented in canonical form versus a lenited but non-canonical continuant-initial form e.g. *baŋga* 'forked stick' vs. **waŋga*, while continuant-initial nouns were presented in canonical form and in an unattested stop-initial ('hardened') form, e.g. *wamba* 'shark' vs **bamba*. We also included two control items: a minimal pair of nouns differing in the initial segment (/b/ vs /w/), where both nouns were licit. Stimuli were presented in two frames ('*I see one X*' and '*I can see X*'): in one frame, the target followed vowel a word ending in /a/; in the other, it followed a word ending in a stop (/b/ in Mawng; /d/ in Iwaidja). We presented the experiment in two counterbalanced versions: a 'natural' and a 'slow speech' version, generated by inserting of 500 ms of silence before the target noun (to investigate whether this affected the likelihood of lenition).

The results are presented in Figure 1. We fitted a series of GLMMs (binomial link) to the data, and the best fit model indicates that both groups prefer canonical forms over unattested 'hardened' forms (e.g. *wamba* 'shark' vs **bamba;* $p < .001$) and over lenited forms (e.g. *baŋga* 'forked stick' vs. **waŋga*; $p = .001$) though Mawng speakers have a lower preference for canonical forms in the lenited trials ($p = .007$) than in the hardening trials ($p = .359$). There was no effect of 'conditioning frame' or 'speech rate'. We fitted an additional GLMM for a confirmatory analysis of a Place of Articulation effect on the acceptability of lenited forms (Figure 2). This confirmed that Mawng speakers are particularly tolerant of lenition of the velar stop (significant intercept $p < .001$; significant interaction between 'Mawng' and /ɡ/, $p = .013$).

The results are consistent with earlier reports that speakers of Mawng are tolerant of lenition of word initial /ɡ/ 0, but—in contrast to previous reports—also suggest that lenition in Mawng is not phonetically conditioned ('conditioning frame' and 'speech rate' *ns*). The results also demonstrate that speakers reject word-forms with initial lenition and fortition if the mutated forms result in collapse of phonological contrasts (/b/ → /w/, /j/ → /ɟ/, or vice versa). The higher acceptability of the lenition from [ɡ] → [ɣ ~ ɰ ~ w] *in Mawng only* likely constitutes a case of free variation, with [ɡ] as the typical realisation. Together, these results suggest that word recognition in Mawng and Iwaidja indeed requires input in '*a form suitable',* and that speakers' rejection of deviant forms reflects fine-grained language-specific phonological and phonetic constraints (bilingual speakers only accept free variation of [g ~ ɣ ~ ɰ ~ w] in Mawng), as well as the shape of the lexicon (including historical and lexicalized mutations). These results highlight that models of continuous parsing and word-recognition must include considerations of language specific phonetic distributions as well as phonological status, phonotactic distributions and frequencies, and the characteristics of the lexicon.

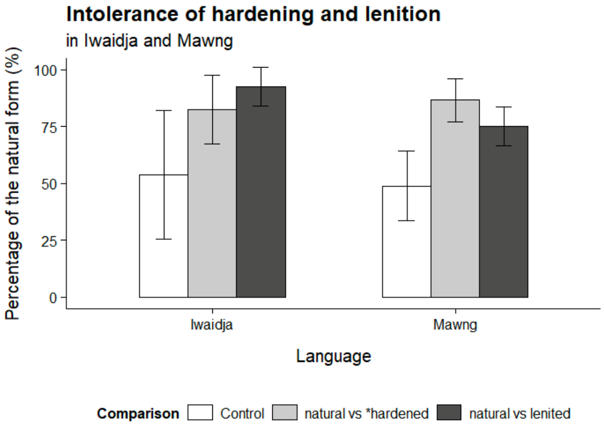


Figure 1. *Mean preference results.*

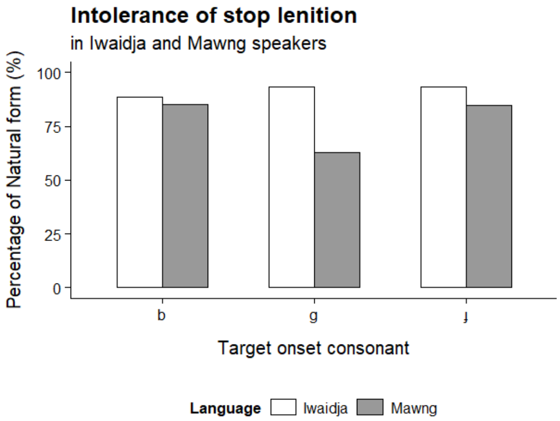


Figure 2. *POA effects.*

Initial mutation in Iwaidja ([7]), adapted transcription, alternating root-initial segments in bold.

(1) a. ŋa-**w**ani-ŋan — b. Ø-**b**ani-ŋan
1SG-sit-PST — 3SG-sit-PST
'I was sitting' — 'He/She/It was sitting'.

(2) a. a-**j**ama-ŋ — b. Ø-**ɟ**ama-ŋ
3PL-work-PRES — 3SG-work-PRES
'They are working' — 'He/she is working'

Word-initial velar stop lenition ([5]), adapted transcription

(3) a. ge — [g ~ ɰ]apala
it.goes — boat
b. gaɾgbin — [g]apala /*ɰabala
big — boat

c. mada — [g ~ w]ubuɲ
VEG.the — canoe
d. marig — [g]ubuɲ/*wubuɲ
NEG — canoe

[1] Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357
[2] Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10(1), 26–63.
[3] Cutler, A. (2008). The 34th Sir Frederick Bartlett Lecture: The abstract representations in speech processing. Quarterly Journal of Experimental Psychology, 61(11), 1601-1619. https://doi.org/10.1080/13803390802218542

[4] Pym, N., & Larrimore, B. (1979). *Papers on Iwaidja phonology and grammar*. (Work Papers of SIL-AAB, Series A, Vol. 2.) Darwin: Summer Institute of Linguistics Australian Aborigines Branch.
[5] Capell, A., & Hinch, H. E. (1970). *Maung grammar: Texts and vocabulary* (Vol. 98). Walter de Gruyter.
[6] Evans, N. (2000). Iwaidjan, a very un-Australian language family. *Linguistic Typology*, *4*(2), 91-142.
[7] Mailhammer, R. & Harvey, M. (2018). A reconstruction of the Proto-Iwaidjan phoneme system. *Australian Journal of Linguistics, 38*(3), 329-359. https://doi.org/10.1080/07268602.2018.1470455

9. Acoustic and Phonetic Adaptations in Sfyria Whistled Speech: An Analysis of Vowel-Consonant Transitions

Sophia Katsiouris[1], Michela Russo[1]
[1]CNRS SFL U. Paris 8/UJM Lyon 3

Sfyria, an endangered whistled language from the Greek village of Antia, transposes spoken Greek into whistled signals for communication over long distances. This study examines vowel and consonant adaptation in Sfyria, focusing on vowel-consonant transitions and consonantal loci. Utilizing acoustic and phonetic methods, naturally recorded Sfyria sequences were analyzed in Praat, segmenting VCV (vowel-consonant-vowel) patterns. This enabled detailed examination of whistled vowel-consonant dynamics and particularly of the 1st harmonic structures (H1 or whistled F0, produced in the front oral part of the vocal tract), with results highlighting H1 modulations as key to phonetic clarity in whistled speech (they determine the perceived pitch of the whistled melodic line Figure 1, p. 2).

Introduction: Whistled languages like Sfyria, Silbo (Canary Islands), and Kuş Dili (Turkey) convert spoken syllables into whistled signals, simplifying phonetic complexity while preserving essential information for distance communication. In Sfyria, spoken Greek syllables are encoded as pitch-based signals, enabling communication across Antia's rugged terrain. This study investigates how Sfyria adapts Greek vowels and consonants, focusing on consonantal loci and vowel-consonant transitions essential for phonetic distinctiveness in whistled speech.

Methodology: The study applies acoustic and phonetic segmentation techniques to map vowel and consonant features in whistled speech. Using Praat, Sfyria utterances were segmented into VCV sequences, with TextGrid annotations precisely marking vowel-consonant boundaries (see Figure 1, p. 2). Analysis focused on whistled F0 (or H1) properties, critical for encoding phonetic cues in whistled speech.

Key Findings:

1. **Vowel Transposition**: Formant contours of spoken vowels are retained in the whistled form with the influence of F2 being a key factor guiding the transposition, suggesting a systematic mechanism. This includes the upward modulation transitions shown in Figure 1. Another example is that high-frequency whistled vowels like /i/ are the result of the transformation of the spoken F2-F3 proximity, reflecting the consistent encoding of vowel sounds. This transposition aligns with patterns observed in other whistled languages, where F2 contours guide vowel representation, ensuring phonological stability despite acoustic reduction.
2. **Consonantal Loci**: Consonants in Sfyria exhibit modulations of H1 that indicate place and manner of articulation, particularly for nasals and dentals. Consonantal loci are marked by H1 shifts influenced by surrounding vowels, encoding articulation points essential for consonant distinction in whistled speech.
3. **Simplified Encoding**: Sfyria compresses phonetic complexity, with vowels preserving formant-like structures, while consonants are encoded through H1 harmonic adjustments. This streamlined encoding preserves linguistic distinctions over distances, showing Sfyria's adaptation of Greek phonetics for efficient long-distance communication.

Conclusion: Sfyria's phonetic system reflects a refined adaptation of Greek phonological features, employing harmonic H1 modulations to convey vowel and consonant information across distances. By relying on systematic harmonic cues, Sfyria achieves intelligibility within a limited acoustic range, offering insights into how whistled languages adapt spoken elements for clear transmission over challenging environments.

(1) Example of TextGrid Annotation for VCV Sequences in Sfyria

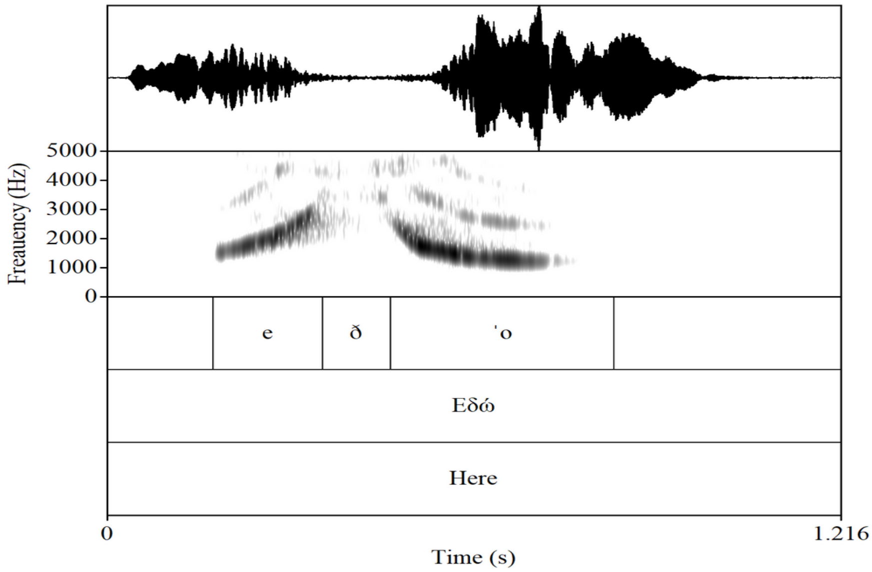


Figure 1. *Waveforms and spectrogram of [eˈðɔ] = here (εδώ), whistled by a whistler. Spectrogram and TextGrid annotations for VCV sequences in Sfyria, illustrating F2 transposition in vowels and H1 markers for consonants*

[1] Breit, F., van t' Veer, M., Botma, B. & van Oostendorp, M. (eds.) (2023) Primitives of Phonological Structure. Oxford University Press. xiv+379 pp.

[2] Charalambakis, Christopher (1994). A case of whistled speech from Greece.Themes in Greek linguistics: Papers from the 1st International Conference on Greek Linguistics. Amsterdam & Philadelphia: Benjamins. 389-396.

[3] Jones, M. C., & Ogilvie, S. (2013). *Keeping Languages Alive : Documentation, Pedagogy and Revitalization* (M. Kouneli, J. Meyer & A. Nevins, eds.; 198-211). Cambridge University Press. https://doi.org/10.1017/CBO9781139245890

[4] Meyer J. (2015). *Whistled Languages. A Worldwide Inquiry on Human Whistled Speech*. Springer Berlin, Heidelberg.

[5] Rialland, A. (2005). Phonological and Phonetic Aspects of Whistled Languages. *Phonology*, *22*(2), 237–271. http://www.jstor.org/stable/4615531.

[6] Varoucha Magdalini. (s. d.). *Grèce Hebdo—Grèce : Le village qui siffle sur l'île d'Eubée*. Grèce Hebdo. Open Access https://www.grecehebdo.gr/loisirs/traditions/2626-gr%C3%A8ce-le-village-qui-siffle-sur-l%E2%80%99ile-d%E2%80%99eub%C3%A9e

[7] Xiromeritis, N. & H. C. Spyridis (1988-89). A whistling language of the village Antias in the Greek island of Evia. *Glossologia* 7-8. 219-224.

10. Phonetic documentation of coronal continuants in Akuzipik

Giulia M. Soldati[1], Harim Kwon[2], Sylvia L.R. Schreiner[1]
[1]George Mason University, [2]Seoul National University

Background: Akuzipik (St. Lawrence Island Yupik, ISO 639-3: ess) is an under-documented, endangered Indigenous language spoken natively by 800-900 people mainly on St. Lawrence Island, Alaska, USA [1]. Previous impressionistic descriptions of Akuzipik propose a phonemic inventory consisting of 31-32 consonants and 4-7 vowels [2, 3]. Most subsequent studies on the language have focused on its morphosyntactic properties [4, 5], but a recent acoustic analysis of the Akuzipik vowel system confirmed 7 distinct phonetic vowel qualities [6]. Akuzipik consonants, however, remain to be phonetically investigated.

Goal: The current work is a descriptive study of the acoustic and articulatory properties of coronal continuants in Akuzipik. It consists of a production experiment designed to investigate the sounds represented by the graphemes ⟨l ll r rr s z y⟩ in intervocalic environments. We use graphemes because no consensus has been achieved as to the identity of the phonemes corresponding to each grapheme. The study targets coronal continuants because their articulatory properties in the language are particularly unclear. For instance, earlier studies describe each sound in the pairs ⟨l⟩/⟨ll⟩ and ⟨r⟩/⟨rr⟩ as voiced/voiceless counterparts, but more recently it has been suggested that they may also differ in other aspects [7].

Methods: Two adult native speakers of Akuzipik participated in this production experiment: one male in his 30s (“M”) and one female in her 40s (“F”). As is the case with most, if not all, current Akuzipik speakers, M and F were bilingual in English, which they acquired around ages 6-7. The stimuli consisted of a list of 71 inflected Akuzipik nouns of 2-5 syllables. In each word, one of the seven target consonants occurred in the onset of a stressed syllable in an intervocalic environment. Audio recordings and ultrasound images of the tongue body movements at the midsagittal plane were collected simultaneously, and the entire recording session took approximately 60 minutes per speaker. Spectrographic analysis was used to measure duration and acoustic correlates for each consonant. For the articulatory analysis, we examined the constriction location, the shape of the tongue body, and the relative location of the anterior parts of the tongue (tongue blade and tip) for each of the target consonants.

Results and discussion: The acoustic analysis confirmed seven distinct coronal continuants in the investigated environments. As expected, the graphemes ⟨s z y⟩ correspond to the segments [s z j]. However, in contrast with early descriptions, the pairs ⟨l⟩/⟨ll⟩ and ⟨r⟩/⟨rr⟩ differed not only in voicing, but also in manner of articulation: while ⟨l r⟩ are voiced approximants, ⟨ll rr⟩ are voiceless fricatives (Fig. 1). Furthermore, while there was variation in duration across speakers and items, it did not indicate a gemination contrast for either grapheme pair. Articulatory analyses found overlapping tongue configurations for the two sounds in the pair ⟨l⟩/⟨ll⟩, indicating the same place of articulation. However, while there was within-speaker consistency, between-speaker variation was observed (Fig. 2): each participant had a different tongue configuration for the pair ⟨l⟩/⟨ll⟩, which could be due to a number of sociolinguistic factors (e.g., age, place of residence, dominant language). Moreover, for both participants, different tongue configurations were observed for each sound in the pair ⟨r⟩/⟨rr⟩: the tongue body was visibly higher for ⟨rr⟩, resulting in greater constriction. Notably, although ⟨r⟩/⟨rr⟩ were previously described consistently as retroflex, neither participant produced them as retroflex in this experiment. Further research will include more participants and extend to more varied phonological contexts as well as to other sounds in the language. This work will contribute to a deeper understanding of the phonetic properties of Akuzipik consonants, providing a valuable resource for the community members and researchers working toward the revitalization of this endangered language.

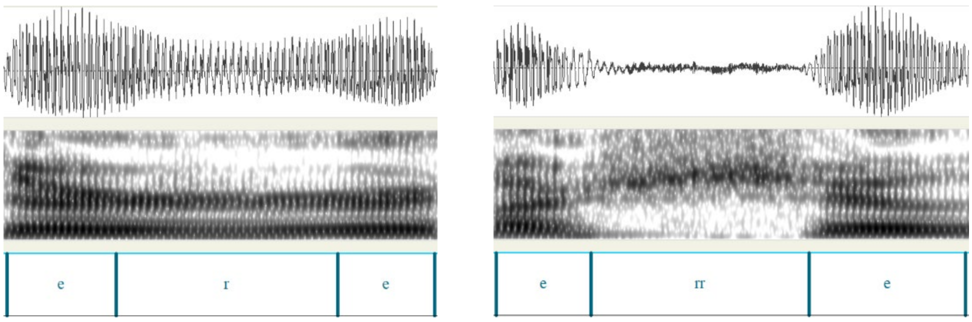


Figure 12. *Waveform and spectrogram representations of the female speaker's production of* ⟨ere⟩ *(left) and* ⟨erre⟩ *(right).*

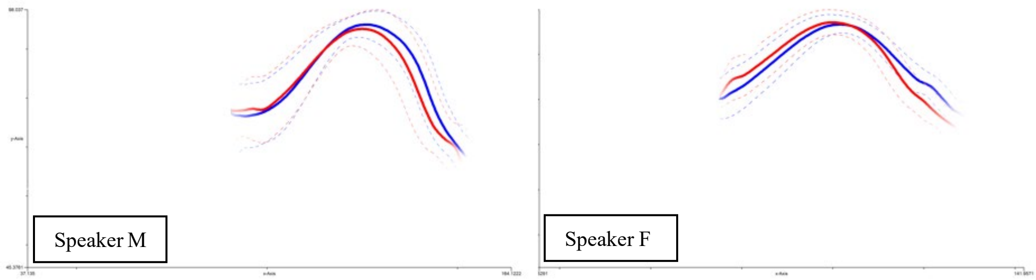


Figure 13. *Mean (solid lines) and standard deviation (dashed lines) tongue configurations for* ⟨l⟩ *(red) and* ⟨ll⟩ *(blue); tongue tip to the right.*

[1] Schwartz, L., Schreiner, S.L.R., & Chen, E. 2019. Community-focused language documentation in support of language education and revitalization for St. Lawrence Island Yupik. *Études/Inuit/Studies, 43*(1/2), 291-312.

[2] Menovshchikov, G.A. 1962. *A grammar of the Asiatic Eskimo language* [Work in translation].

[3] Krauss, M.E. 1975. St. Lawrence Island Eskimo phonology and orthography. *Linguistics, 152*, 39-72.

[4] de Reuse, W.J. 1994. *Siberian Yupik Eskimo: The language and its contacts with Chukchi* (Studies in Indigenous languages of the Americas). Salt Lake City: University of Utah Press.

[5] Nagai, K. 2004. *A Morphological Study of St. Lawrence Island Yupik: Three Topics on Referentiality*. Doctoral dissertation, Kyoto University.

[6] Hunt, B., Kwon, H., & Schreiner, S.L.R. 2020. An acoustic analysis of St. Lawrence Island Yupik vowels. *The Journal of the Acoustical Society of America, 148*, 2471-2471.

[7] Schreiner, S.L.R., Schwartz, L., Hunt, B., & Chen, E. 2020. Multidirectional leveraging for computational morphology and language documentation and revitalization. *Language Documentation & Conservation, 14,* 69-86.

11. The cue of pitch variability in speech segmentation: a pilot study with Mandarin Chinese listeners

Shu-chen Ou
National Sun Yat-sen University

The present study examines how pitch variability influences speech segmentation using an artificial language learning (ALL) experiment. In listening to continuous speech, listeners must identify discrete units such as words using various distributional cues, such as transitional probability (TPs) [4]. Other possible cues include pitch variability, which can be defined as the degree of pitch change in spoken language and can be quantitatively measured across syllables, words, or phrases (e.g., pairwise variability index (PVI)). Previous research suggests that low pitch variability often indicates continuity or a single unit while high pitch variability signals the possible start of a new phrase or word [1, 2 ,3]. Thus, pitch variability likely serves as a cue for speech segmentation. This study uses an ALL experiment with Mandarin Chinese speakers as an initial pilot to explore this potential role of pitch variability in speech segmentation.

The artificial language included six "words," each comprising a meaningless trisyllabic sequence (e.g., [banume]). Each syllable within these words was recorded separately and manipulated to have the same duration but a flat F0 contour. TPs between syllables served as the sole cues for word definition. Additionally, F0 patterns with either "low variability (LV)" or "high variability (HV)" were superimposed onto the words, resulting in two experimental conditions (Figure 1). In the LV condition, the mean F0 of each word, denoted by $F0_w$, followed a normal distribution with mean = 130 Hz and standard deviation (SD) = 15 Hz whereas the F0 values of the word-internal syllables followed a normal distribution with mean = $F0_w$ and SD = 5 Hz. The design of the HV condition was the same except the SD of word-level F0 and that of word-internal syllables were set to 5 Hz and 15 Hz, respectively. The experiment was conducted using E-Prime 3.0. Participants initially heard 150 repetitions of each of the six words, which were pseudo-randomly concatenated without pauses. They subsequently took a two-alternative forced-choice test, with each trial presenting one actual word from the artificial language and one part-word, asking participants to identify the word belonging to the language. All test stimuli had a flat F0 at 130 Hz. Thirty participants were assigned to each condition (mean age =23.4).

Figure 2 shows the response accuracy of the forced-choice test by condition. A mixed-effects logistic regression model was fitted to the binary responses (correct or incorrect), including Condition as the fixed factor and log-transformed response time as a covariate. The random effects included a random intercept for the words of the artificial language. The results indicated that responses in the LV condition (mean = 63.80%) was significantly more accurate than those in the HV condition (mean = 49.63%) ($\beta = 0.62$, SE = 0.09, $z = 6.86$, $p < 0.001$).

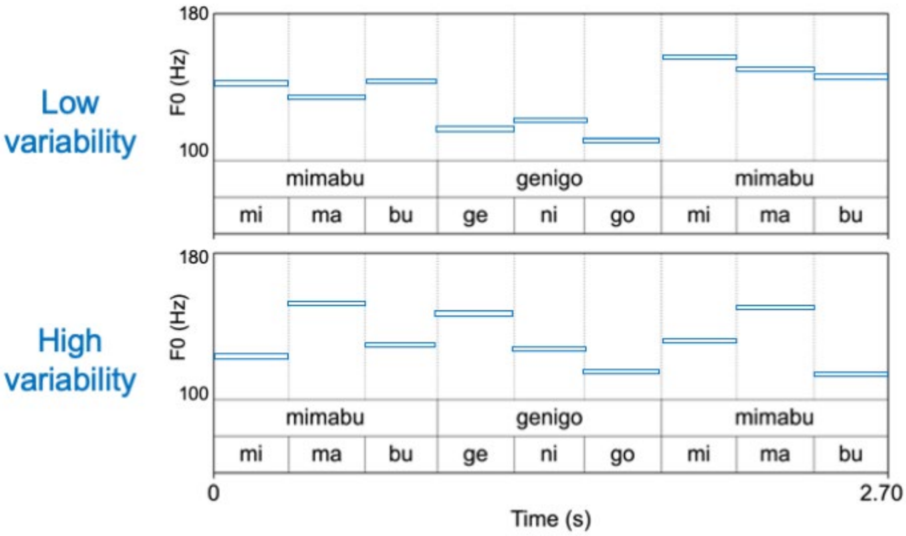


Figure 1. *F0 contours of a sample speech stream under the low and high variability conditions.*

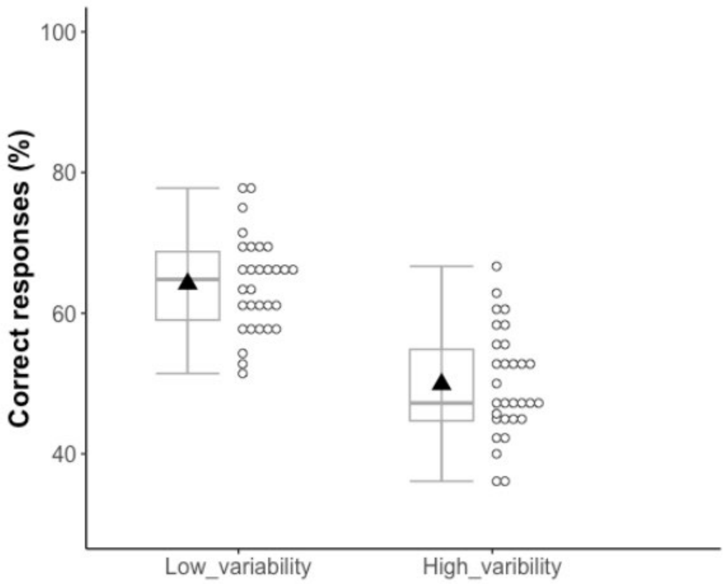


Figure 2. *Percentages of correct responses of individual participants (empty circles) and the means of low variability-cued and high variability-cued conditions (filled triangles).*

The findings support the hypothesis that listeners segment the artificial language more effectively with low pitch variability, which enhances segmentation, whereas high pitch variability hinders this process. However, it remains unclear whether the preference for low pitch variability is specific to language experience or if reflects a more general cognitive mechanism. One possible explanation is that the low pitch variability resembles the phonetic characteristics of Mandarin Chinese tones; specifically, pitch variability within a word tends to be lower than that between words [1]. Alternatively, low pitch variability may serve as a domain-general cue, similar to TPs. To investigate this further, future research will include studies with speakers of various languages.

[1] Chen, A.-H. 2024. Pitch variability in spontaneous speech production and its connection to usage-based grammar. *Journal of Phonetics 106* (101342), 1-25.
[2] Lin, P. M. S. 2013. The prosody of formulaic expression in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics 18*(4), 561-588.
[3] Pan, S., & Hirschberg, J. 2000. Modelling local context of pitch accent prediction. In *Proceedings of the 38th annual meeting of the Association for Computational Linguistics* (pp. 233-240). Hong Kong: Association for Computational Linguistics.
[4] Saffran, J. R., Newport, E. L., & Aslin, R. N. 1996. Word Segmentation: The role of distributional cues. *Journal of Memory and Language 35*(4), 606-621.

12. Nasality overlap in French rhotics

Andrés Felipe Lara[1] and Jenifer Vega Rodriguez[2,3]
[1]*LPP, CNRS, U. Sorbonne Nouvelle,* [2]*GIPSA lab, CNRS, U. Grenoble Alpes,* [3]*Brasilia Univ.*

Nasalization occurs when nasal airflow from lowering the velum in nasal consonants overlaps with adjacent segments, influenced by perceptual and linguistic factors [1]. Anticipatory nasalization involves greater velopharyngeal opening (VPO) than carryover nasalization, with peak VPO varying by context and reflecting nasal airflow differences [2]. Fricatives resist nasalization due to maintaining both friction and airflow, which is biomechanically and perceptually difficult, while French liquids, particularly rhotics, show more anticipatory nasalization [4]. French rhotics face unique challenges as combining nasal airflow with fricative friction is complex. Despite extensive research on French nasalization [4, 5], studies on rhotics remain sparse [3]. French nasal coarticulation is restricted, with vowels before nasals typically remaining oral, and nasalization arising from physiological timing rather than deliberate articulation [1, 5]. While nasalization is more prominent in vowels, rhotics may display subtle nasalization patterns requiring further study. This study hypothesizes that rhotic nasalization is unintentional [3] and shaped by phonetic contexts, following the nasalization continuum proposed by [2]: *oral sounds < contextually nasal vowels < nasal consonants < nasal vowels < pauses*.

Three native French speakers (1 male, 2 female) from Paris participated in this study. Two corpora were designed words: (1) three repetitions of nasal vowels /ɑ̃, ɛ̃, ɔ̃/ and oral vowels /a, i, ɔ/ in PṼ and PVP context respectively to determine a nasalization threshold, and (2) 16 words repeated three times with /ʁ/ in different phonetic contexts (#ʁV, Vʁ#, ʁVN, VʁN, ʁṼ, NVʁṼ) where V and Ṽ include /i, a, ɔ, ɑ̃, ɛ̃, ɔ̃/, and N is /m/. Three channels were recorded synchronously using an EGG D800 (Laryngograph Ltd.): nasal airflow (NAF), oral airflow (OAF), and acoustic (WAV). Praat, Matlab, and Python were used for data segmentation, processing, and statistical analyses.

Nasal rate (*NR*) was calculated as *(NR =NAF/(NAF+OAF)* to establish nasality thresholds for oral vowels ([i, o, a], *NR* = 19.5%) and nasal vowels ([ɛ̃, ɔ̃, ɑ̃], *NR* = 31.4%) within the same context (/p_p/). Figure 1 shows that higher vowels in both nasal and oral categories exhibited higher *NR* values. The nasality threshold was set at 20%, consistent with the French threshold established by [6]. Figure 2 illustrates a nasalization continuum for [ʁ], with the lowest nasalization observed in oral contexts (#ʁV, ʁV#) at 14% ± 9.4%, followed by nasal non-adjacent contexts (ʁVN) at 18% ± 8.1%. Nasal-adjacent contexts exhibited progressively higher nasalization: ʁṼ (20.3%), VʁN (27%), and NVʁṼ (30%). A significant difference in *NR* across contexts was found for [ʁ] ($p \approx 2.53 \times 10^{-13}$). Figure 3 compares *NR* values for [ʁ] in the ʁṼ context, showing vowel-specific effects: [ɔ̃] had higher nasalization than [ɑ̃] or [ɛ̃]. This contradicts [5] and aligns with findings by [7], who reported a weak correlation between nasal airflow (NAF) and tongue height, suggesting the nasalization continuum may not fully capture all contributing factors. Figure 4 plots individual productions across contexts. Speaker M1 demonstrated clear distinctions between #ʁV, Vʁ#, ʁVN, and ʁṼ, NVʁṼ, with the most variability in ʁṼ. Speakers F1 and F2 separated oral and nasal contexts, with greater F1 variability in #ʁV and ʁV#, and greater F2 variability in #ʁV and VʁN. No significant differences were observed between speakers ($p = 0.22$). Interestingly, the observed *NR* continuum differed from predictions [2], with ʁṼ showing lower nasalization than VʁN. The revised continuum for rhotics is as follows: #ʁV < ʁV# < ʁVN < ʁṼ < VʁN < NVʁṼ. Preliminary results suggest greater anticipatory nasal overlap near nasal consonants but are based on a small speaker sample. Further research is needed to confirm coarticulatory trends and explore potential language-specific nasalization patterns

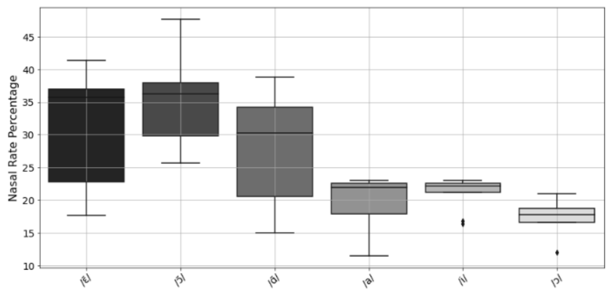


Figure 1. *Nasal rate (%) for oral* /a, i, ɔ/ *and nasal vowels* /ɑ̃, ɛ̃, ɔ̃/ in Pṽ and PvP contexts.

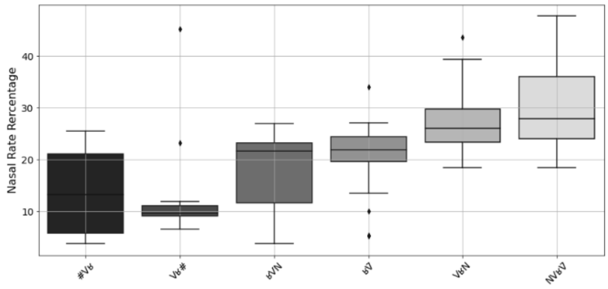


Figure 2. *Nasal rate (%) for [ʁ] in different phonetic contexts (all vowels combined)*

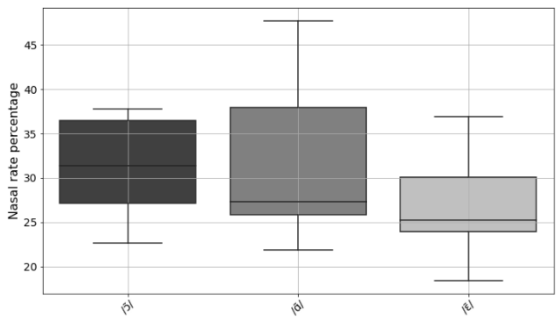


Figure 3. *Nasal rate (%) for [ʁ] when preceding three vocalic contexts /ɑ̃, ɛ̃, ɔ̃/.*

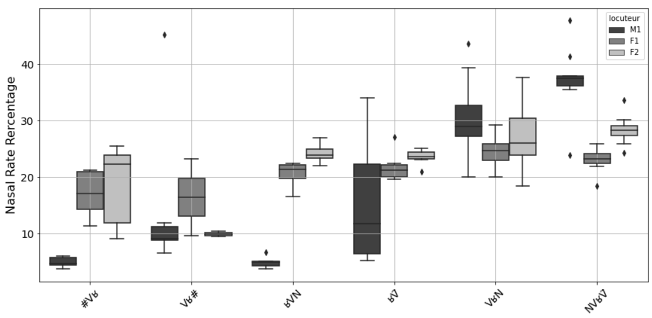


Figure 4. Speaker productions of *[ʁ]across different contexts. M= male and F = female*

[1] H. Clumeck, "Patterns of soft palate movements in six languages," Journal of Phonetics, vol. 4, pp. 337-351, 1976.

[2] G. De Boer, J. Islam, C. Purnomo, L. Wu, et B. Gick, « Revisiting the nasal continuum hypothesis: A study of French nasals in continuous speech », Journal of Phonetics, vol. 98, p. 101244, mai 2023, doi: 10.1016/j.wocn.2023.101244

[3] M. Solé, "Aerodynamic characteristics of trills and phonological patterning," *J. Phonetics*, vol. 30, pp. 655-688, 2002.

[4] V. Delvaux, D. Demolin, B. Harmegnies, and A. Soquet, "The aerodynamics of nasalization in French," *J. Phonetics*, vol. 36, pp. 578-606, 2008.

[5] A.P. Benguerel, "Nasal airflow patterns and velar coarticulation in French," in *Speech Communication Seminar*, Stockholm, 1974, pp. 105-112.

[6] A. Elmerich, A. Amelot, S. Maeda, Y. Laprie, J. F. Papon, and L. Crevier-Buchman, "F1 and F2 measurements for French oral vowel with a new pneumotachograph mask," in ISSP 2020-12th International Seminar on Speech Production, December 2020.

[7] R. K. Shosted, *The aeroacoustics of nasalized fricatives*, PhD dissertation, University of California, Berkeley, 2006.

13. Uhmming and uhing across languages: Evidence from Italian learners of German

Simona Sbranna, Simon Wehrle
University of Cologne

Filled pauses (FP) such as “uh” or “uhm” are often considered a byproduct of speech planning, and have usually been judged and perceived negatively, especially in public speech [1,2]. This is why FP tend to receive little attention, or are even stigmatised, in second language (L2) teaching contexts although native speakers use them in spontaneous speech across languages to fulfil various functions beyond speech planning, such as mitigating opinions, managing turn-taking, and signalling the status of discourse entities [3]. However, languages differ in the distribution and phonetic realisation of FPs [4,5,6] and deviations from native norms can distract native listeners from the message, leading them to focus on the learner’s speech style instead [7]. Having established the importance of FPs in non-native speech, we investigated FPs of the form “uh(m)” produced by Italian learners of German in their L1 and L2, which were for this first analysis compared to a parallel corpus of L1 German speech employing the same methods and elicitation context [8].

We analysed 20 Map Task dialogues [9] involving 10 pairs of learners speaking in their L1 Italian and L2 German. All relevant tokens including a final nasal were annotated as <ehm>, and all tokens without a nasal were annotated as <eeh>. We measured the rate, duration, type (ehm vs eeh), and intonational realisation of all 840 tokens in the current data set. For the prosodic analysis, we manually corrected and smoothed all tokens, extracted pitch values at 10% and 90% of token duration, and calculated the difference between these values in semitones (ST; with a reference value of 1 Hz; cf. [10,11, 9].

We found a strong effect of proficiency for the rate of FPs, with L2 German speech (mean = 9.5) containing three times more FPs compared to L1 Italian (mean = 3.26) and L1 German data (mean = 3.63); see Fig. 1. Italian and German L1s also differ in FP type, with Italians using comparatively fewer “ehms” (35%) than Germans (60% as reported in [11]), possibly due to phonotactic differences between the two languages. In their L2, learners produced more “ehms” than in their L1 (41%), somewhat resembling the target group. Duration was very similar in Italian L1 and German L2, averaging around 600 ms—longer than the average in the German L1 corpus (~450 ms). Finally, in terms of intonation, there was no significant difference between learners’ L1 and L2, but across L1s. German L1 reportedly has 70% level contours and only 22% falling contours [11], whereas the opposite is true for Italian learners (Italian L1: 68% falls and 29% level; German L2: 71% falls and 28% level). Notably, rises were found only in 2 of 10 Italian dyads, but in 7 of 10 L2 German dyads.

To conclude, the use of FP by Italian learners of German reflects lower proficiency in L2 and transfer from L1. In line with the literature, the effect of proficiency is evident in the rate of FP, which is much higher in L2 than in L1, likely due to the greater cognitive load of speaking a foreign language, while it is fairly similar across L1s. In contrast, FP type and intonation differ across L1s (as in [5, 6]) and show a transfer of the learners’ native style into L2. Indeed, learners show little change in FP production across their L1 and L2, aside from a slight increase in nasalised FP (partial adaptation was also found by [6]). Whether this reflects progress toward an ideal target is debatable with the current amount of data, but merits investigation on a larger scale. Worth noting is that learners also present more rise contours, which may indicate uncertainty. Overall, these results support the inclusion of FP in L2 curricula to raise awareness of crosslinguistic differences and help learners sound more spontaneous in oral interactions.

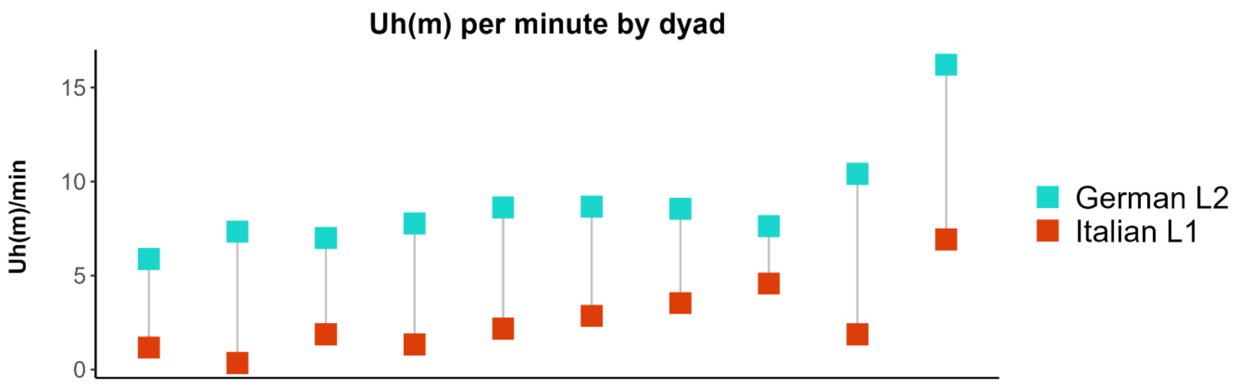


Figure 1. *Rate of « uh(m) » by dyad and language group. Dyads on the x-axis, ordered by grand mean; German L2 in blue, Italian L1 in red.*

[1] Niebuhr, O., & Fischer, K. (2019, September). Do not hesitate! – Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. In *INTERSPEECH 2019* (pp. 544-548). International Speech Communication Association (ISCA).

[2] Fox Tree, J.E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse processes*, *34*(1), 37-55.

[3] Rose, R.L. (2008). *Filled pauses in language teaching: Why and how* (Doctoral dissertation, Waseda University).

[4] Clark, H.H., & Tree, J.E.F. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73-111.

[5] Lo, J.J.H. (2020) Between *Äh(m)* and *Euh(m)*: The Distribution and Realization of Filled Pauses in the Speech of German-French Simultaneous Bilinguals. *Lang Speech*; 63(4):746-768.

[6] de Boer, MM, Heeren, W.F.L. (2020) Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English. *J Acoust Soc Am.*;148(6):3612.

[7] Fayer, J. and Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37:313–326.

[8] Wehrle, S.; Grice, M. & Vogeley, K. (2023). Filled Pauses Produced by Autistic Adults Differ in Prosodic Realisation, but not Rate or Lexical Type. *Journal of Autism and Developmental Disorders.*

[9] Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., ... & Weinert, R. (1991). The HCRC map task corpus. *Language and speech*, *34*(4), 351-366.

[10] Sbranna S., Wehrle S. & Grice M. (2022). The use of Backchannels and other Very Short Utterances by Italian Learners of German. In *The position of the speaker in interaction: attitudes, intentions, and emotions in verbal communication*, edited by Riccardo Orrico and Loredana Schettino, Studi AISV 9:149-169. Milan: Officinaventuno.

[11] Wehrle, S. (2021). *A Multi-Dimensional Analysis of Conversation and Intonation in Autism Spectrum Disorder.* PhD Thesis, University of Cologne (Germany).

14. Effects of Sexuality on the Phonation of Female Spanish Speakers

Jesus Duarte
University of California, Los Angeles

The study of queer speech has largely focused on English-speaking contexts, with a wealth of research exploring how sexual identity influences phonetic and acoustic characteristics. However, much less is known about how these patterns manifest in other languages, particularly Spanish. This lack of research leaves a critical gap in our understanding of how linguistic markers of sexual identity may vary across linguistic and cultural contexts. Research on English-speaking lesbian women has identified creakiness as a recurring feature, potentially linked to stereotypes of lower pitch and perceived masculinity in lesbian speech. While similar stereotypes and linguistic associations may exist in Spanish-speaking communities, the degree to which these patterns hold and how they are influenced by cultural norms remains unclear. Furthermore, questions about the role of identity performativity—how openly one expresses their sexual orientation—add another layer of complexity.

This paper addresses these gaps by investigating the phonetic features of creakiness in Spanish-speaking lesbian women. It aims to determine whether these speakers exhibit patterns similar to their English-speaking counterparts and to explore the influence of being openly versus closeted on voice quality. The present results analyzed speech data collected from four college students in Michoacán, Mexico, including two lesbian (one openly lesbian and one closeted) and two heterosexual women. Speech samples were obtained through semi-spontaneous description tasks, sociophonetic interviews, and sociodemographic surveys, providing a range of linguistic and formality contexts. Acoustic analysis focused on measures of creakiness, including Open Quotient (OQ) and Degree of Spectral Tilt (DST).

The results revealed significant differences in voice quality between the groups. The openly lesbian participant exhibited the highest levels of creakiness, as indicated by lower OQ and DST values, consistent with patterns observed in English-speaking lesbian women. In contrast, the closeted lesbian participant displayed speech patterns more similar to the heterosexual participants and even showed greater breathiness in some instances. These findings suggest that openness about one's sexual orientation influences phonation, with closeted individuals potentially adopting voice qualities more aligned with societal norms for femininity.Heterosexual participants showed no significant differences between themselves in either OQ or DST. These results highlight that voice quality, particularly creakiness, serves as a potential marker of sexual identity among Spanish-speaking lesbian women, but that identity performance and social pressures significantly shape these linguistic expressions.

These findings contribute to the understanding of how sexual identity is encoded and perceived in speech, and bridges a crucial gap in the literature by providing insights into how sexual identity is expressed and perceived through speech in non-English languages. They underscore the influence of cultural norms and individual identity disclosure on linguistic behavior. This work not only highlights the parallels between queer speech patterns in English and Spanish but also points to the complexities of performativity and social pressures on speech production. Data continues to be analyzed to further add to these results.

Barron-Lutzross, A. (2015). The Production and Perception of a Lesbian Speech Style. *UC Berkeley PhonLab Annual Report*, 11. http://dx.doi.org/10.5070/P77f6332bh

Barron-Lutzross, A. (2018). *Speech stereotypes of female sexuality* (Order No. 10821814). Available from Linguistics and Language Behavior Abstracts (LLBA). (2101574066).

Camp, M. (2010). *Japanese lesbian speech: Sexuality, gender identity, and language* (Order No. AAI3366693). Available from Linguistics and Language Behavior Abstracts (LLBA). (85706585; 201006376).

Crisosto A., J., Pérez M., H., & Dagnino R., F. (2015). La influencia de la orientación sexual de hombres y mujeres en las características acústicas de las vocales del español. Un estudio vocálico comparativo. *Revista Chilena de Fonoaudiología, 14*, 55-67. doi:10.5354/0719-4692.2015.37614

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech, 50*(1), 125-142. doi:http://dx.doi.org/10.1177/00238309070500010601

Van Borsel, J., Vandaele, J., & Corthals, P. (2013). Pitch and pitch variation in lesbian women. *Journal of Voice, 27*(5), 656.

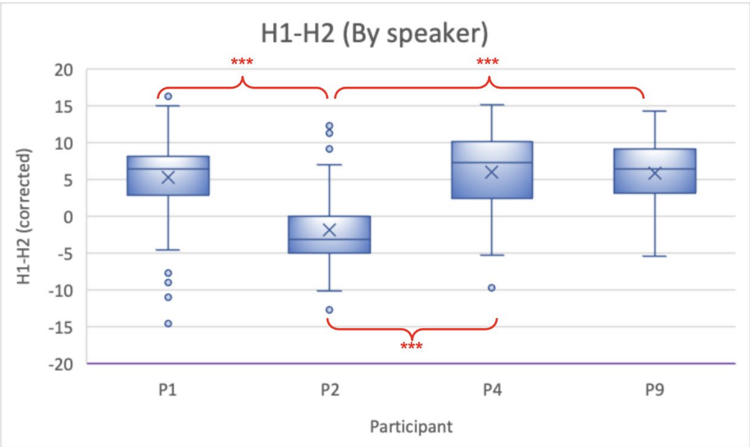


Figure 3. Open Quotient values by participant. P1: not openly gay, P2: Openly gay, P3 & P4: straight

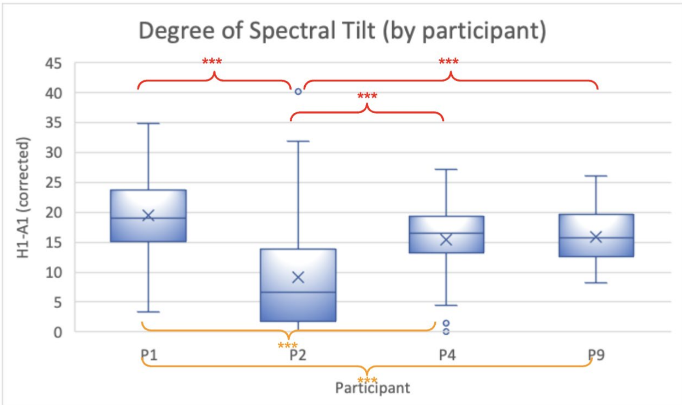


Figure 5. Degree of Spectral Tilt values by participant (P1: not openly gay, P2: Openly gay, P3 & P4: straight)

15. From acoustic cue weighting to perceptual learning of tonal categories in Mandarin distorted speech

Pàn Zhāng
University of Cambridge, Cambridge, UK, Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián

Although speech is highly variable, listeners usually automatically and subconsciously adapt to overcome any challenges this poses, resulting in veridical perception.[1, 2] Four types of processes have been shown to support such "adaptive" perception: (1) perceiving inputs from a hierarchy of acoustic cues[3]; (2) categorising acoustic cues into pre-existing mental representations[4]; (3) recalibrating and learning degraded speech with contexts[5]; and (4) generalising the adaptation to acoustically similar speech[6]. Previous research has examined these processes in, for instance, consonant or syllable perception[3-6], but it is unknown to what extent adaptive perception works the same way in prosody perception.

This study aims to address the gap by examining adaptive speech perception with lexical tone. In tonal languages, lexical tones, acoustically marked by the fundamental frequency ($F0$), are used to express semantic and syntactic functions just as segmental features do, which makes their veridical perception vital for successful communication[7]. It is therefore important to establish that to what extent the same adaptive perception processes support the perception of variable tonal information.

We carried out two behavioural experiments investigating (1) the role of duration as a secondary acoustic cue to tone when the primary cue $F0$ has been removed; and (2) perceptual learning and generalisation effects when the primary cue $F0$ is ambiguous between two contrasting tones. In experiment 1, using forced-choice identification of vocoded tones by 30 native Mandarin listeners, a t-test showed that duration successfully cued the identity of a tone in the absence of $F0$ information (Figure 1). Experiment 2 (see Table 1 for design) used acoustic continua of lexical tone contrasts (Figure 2a) in two forced-choice identification tasks in pre-training, training, and post-training phases. Pre-training established the phonetic categorisation baseline which performance post-training was compared against, thus providing a measure of perceptual learning and generalisation effects generated during the training phase (lexically biased contexts were used to facilitate phonetic recalibration and perceptual learning of ambiguous steps in the continua). Mixed-effects logistic regression confirmed that the representational category (Tone 1, here) was significantly expanded post-training rather than shifted, as expected (Figure 2b).[8] The results also showed generalisation effects across novel words within but not across speakers, even with trained words (Figure 3). This suggests that effective generalisation requires abstractionist acoustic relevance and similarity[9]. In conclusion, lexical tone perception was found to share similar adaptive mechanisms observed previously for segments and syllables, favouring abstractionist over exemplar models.

	Monosyllabic words	**Speakers**
Condition 1: same words, same speaker	bā-bá, chū-chú, hē-hé, mī-mí, qū-qú	Male speaker
Condition 2: different words, same speaker	bī-bí, chā-chá, hū-hú, mō-mó, qī-qí	Male speaker
Condition 3: same words, different speaker	bā-bá, chū-chú, hē-hé, mī-mí, qū-qú	Female speaker

Table 1. *Experimental design*

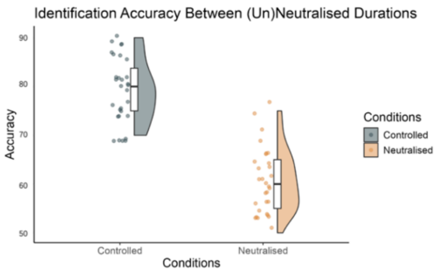


Figure 1. *Identification accuracy (%) for tone stimuli with (controlled) and without (neutralised) durational cues* (Paired samples $t = 7.87$, df = 28, $p < .001$).

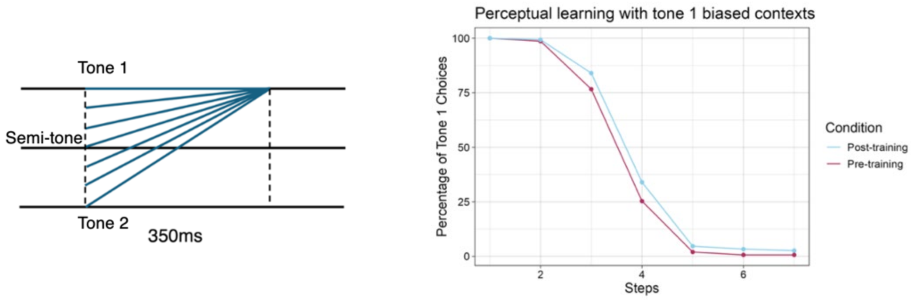


Figure 2. *(a) Schematic of the 7-step acoustic continuum from tone 1 to 2, and (b) Significant phonetic recalibration post-training, showing perceptual learning*

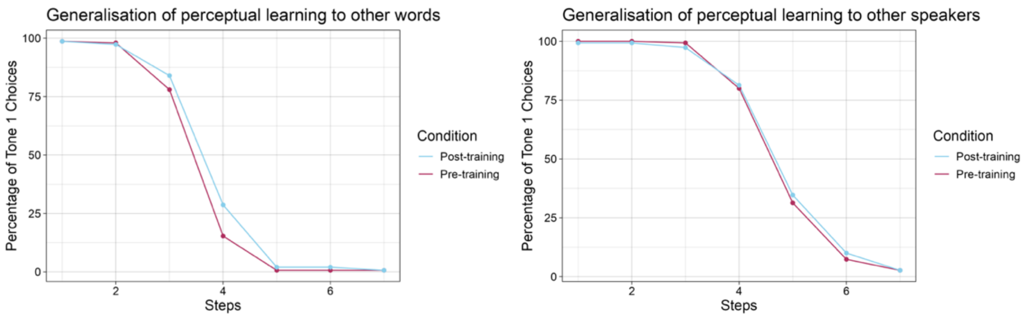


Figure 3. *(a) Significant generalisation to novel words within speaker, and (b) no generalisation across speakers*

[1] Bradlow, A. R., & Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707-729.

[2] Clarke, C. M., & Garrett, M. F. 2004. Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647-3658.

[3] Idemaru, K., & Holt, L. L. 2011. Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1939.

[4] Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358.

[5] Norris, D., McQueen, J. M., & Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204-238.

[6] Kraljic, T., & Samuel, A. G. 2006. Generalisation in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13(2), 262-268.

[7] Chao, Y. R. 1965. *A Grammar of Spoken Chinese*.

[8] Melguy, Y. V., & Johnson, K. 2022. Perceptual adaptation to a novel accent: Phonetic category expansion or category shift? *The Journal of the Acoustical Society of America*, 152(4), 2090-2104.
[9] Mitterer, H., Scharenborg, O., & McQueen, J. M. 2013. Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2), 356-361.

16. The perception of intonational emphasis: an individual differences approach

Chen Zhou[1], Jason Bishop[1,2], Mei-Ying Ki[1]
[1]*CUNY, the Graduate Center;* [2]*CUNY, the College of Staten Island*

Recent research on prosody perception and processing highlights the importance of listener-based properties, including aspects of cognition related to constructs such as "pragmatic skill" [1]. Individual differences in pragmatic skill (relevant to predicting the extent to which a listener engages in pragmatic inference, integrates discourse context, and is attentive to speaker intentions) have been implicated in listeners' sensitivity to the subtle gradient phonetic cues that might mark focus types [2] or distinguish intonational pitch accent categories [3]. Some researchers have suggested (e.g., [4]) that such gradient phonetic cues encode differences in emphasis, a paralinguistic aspect of speech, and thus said individual differences related to pragmatic skill may actually reflect individual differences in attention to paralinguistic uses of prosody. The present study aimed to explore this issue further by revisiting a long-standing but still poorly understood issue – the extent to which intonational emphasis is cued gradiently versus categorically. In their classic study using a categorical perception paradigm, [5] observed a surprisingly S-shaped identification curve for emphatic/non-emphatic decisions in response to continua for f0 peak values (see Figure 1). Given that assigning paralinguistic meaning to f0 peaks in this way requires the interpretation of speaker intentions (rather than being based on grammatical categories as such), we hypothesized that the sigmoid function that [5] found at the group level would be less reliable for individual listeners with lower levels of pragmatic skill (who, again, should be less driven to assign intentions to cues). To test this, and to explore cross-listener variation in categorical vs. gradient behavior more generally, we carried out a categorical perception experiment like [5]'s, with a focus on individual variation.

Stimuli for the current experiment were created similarly to the stimuli in [5]. Recordings were made of a male native American English speaker producing twelve trisyllabic names with either initial (e.g., *Emily*) or penultimate (*Armando*) stress, bearing with L+H* pitch accents. Using PSOLA resynthesis in PRAAT [6], the f0 peak of each production was then manipulated in 11 steps from 105 Hz to 215 Hz in 11 Hz-increments. This produced a total of 132 stimuli (11 steps x 12 names), that listeners heard in pseudorandomized order. Based on work like [2] and [3], listeners also completed three measures believed to estimate pragmatic skill. The first was the Empathy Quotient (EQ), a self-report questionnaire [7]. The second and third were tests of emotion recognition: the Reading the Mind in the Eyes Test (EYES) [8] and the Bell-Lysaker Emotion Recognition Test (BLERT) [9]. Listeners were 158 American English speakers (90 female, 68 male) who participated in both an identification and discrimination task in the manner of the standard categorical perception paradigm, although we limit our discussion to primarily the results of the identification task.

As a group, listeners in this study produced a fairly sigmoid identification function, suggesting relatively categorical-like behavior, much like the smaller group of listeners in [5]'s study exhibited. To assess cross-listener variation, however, we fitted regression lines to each individual listener's identification curve (calculated across items for each listener) so that slopes could be calculated for each. These slopes served as an indication of how gradient versus categorical a listeners' treatment of the continua was: steeper slopes indicate more sigmoid functions (i.e., more categorical-like perception), shallower slopes more linear ones (i.e., more gradient perception). Results showed considerable variation across listeners in terms of their treatment of the f0 continua, with two measures of pragmatic skill (EYES and BLERT) having the predicted effect: listeners with lower levels of pragmatic skill were associated with more gradient/linear identification functions (see Figure 2); listeners with higher levels of pragmatic skill more categorical/sigmoid ones. However, despite the direction of the effect of pragmatic skill being in the one predicted, a large proportion of variation was left accounted for; listener gender effects were also apparent. Implications of these findings will be discussed.

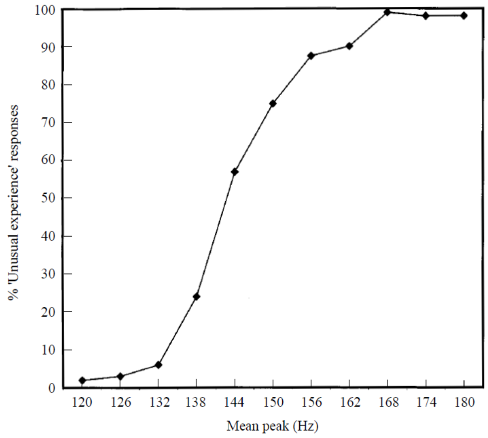


Figure 1: *Identification function for listeners in [5]'s categorical perception experiment.*

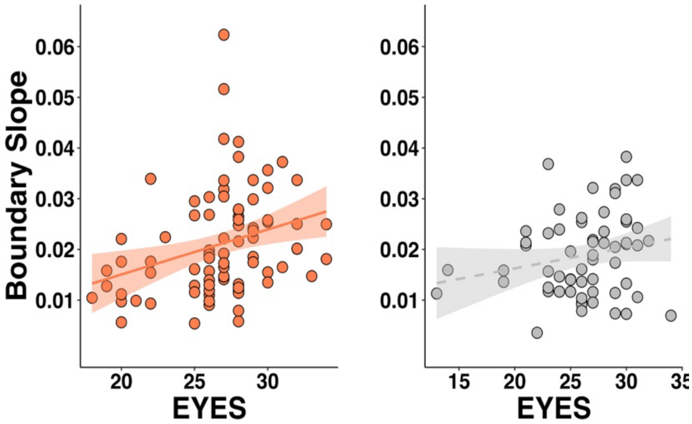


Figure 2: *Boundary slope as a function of EYES scores (left: female listeners, right: male). Higher slopes indicate more sigmoid response functions; higher EYES scores indicate higher levels of pragmatic skill.*

[1] Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology*, 4, 708.

[2] Bishop, J. (2017). Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience*, 32(2), 236–253.

[3] Arvaniti, A., Gryllia, S., Zhang, C., & Marcoux, K. P. (2022). Disentangling emphasis from pragmatic contrastivity in the English H*∼ L+ H* contrast. *Proc. Speech Prosody 2022*, 837-841.

[4] Gussenhoven, C. (2015). Does phonological prominence exist? *Lingue e Linguaggio*, 14(1), 7–24.

[5] Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, 25(3), 313–342.

[6] Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer [Computer program]. Version 6.1. 15.

[7] Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34, 163–175.

[8] Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(2), 241–251.

[9] Bell, M., Bryson, G., & Lysaker, P. (1997). Positive and negative affect recognition in schizophrenia: A comparison with substance abuse and normal control subjects. *Psychiatry Research*, 73(1–2), 73–82.

17. Final devoicing across contexts: the case of two varieties of Ossetic

Yury Makarov[1, 2], Mark Gibson[3]
[1]U. of Cambridge; [2]V.V. Vinogradov Russian Language Institute, RAS; [3]Universidad de Navarra

Introduction. Iron and Digor are varieties of Ossetic, a minority Iranian language spoken in the Caucasus by approximately 400,000 and 100,000 people, respectively. Although these varieties are claimed not to be mutually intelligible, they exhibit significant similarities in basic vocabulary and phoneme inventories. To the best of our knowledge, the phonetic aspects of voicing in Ossetic have not been studied since the mid-20th century [1]. This study investigates final devoicing in Iron and Digor, where previous reports have been inconclusive[3], and the types of production cues that are used to discriminate the final voice contrast in consonants when voicing is neutralised, a topic not previously explored for Ossetic.

Data. Our data come from 11 (9 female, 2 male) speakers of Iron and 5 (3F, 2M) speakers of Digor from Vladikavkaz, Russia; the fieldwork was conducted in 2023. We recorded 10 and 6 (near-)minimal pairs for Iron and Digor, respectively (see Appendix). Each word was first repeated three times with pauses in between pronunciations, then twice in phrase-final and in phrase-medial prevocalic positions.

Focus and methods. In this preliminary study, we examined several acoustic features cross-linguistically associated with [± voice], namely, segment voicing (primary cue), preceding vowel duration, closure and aspiration duration (for plosives), and frication duration (fricatives) [2], [3]. To account for potential differences in speech tempo, we utilised relativised voicing ($\frac{voicing\ duration\ (s)}{closure/frication\ (s)}$) and vowel duration ($\frac{vowel\ (s)}{word\ (s)}$). We also tested the effects of context, namely, isolated (i), phrase-final (f), and phrase-medial prevocalic (m), and manner (plosive/fricative) on vowel duration and the degree of devoicing.

Results. The results are summarised in Table 1. Both Iron and Digor Ossetic exhibit final devoicing; however, in Iron, [+ voice] obstruents preserve voicing in up to 40% of their closure and up to 35% of their frication. Closure duration varies insignificantly (~10 ms) in both varieties, while frication is up to 63 ms longer in Iron [– voice] fricatives. Final aspiration distinguishes between [± voice] plosives by ~43 ms in i&f and ~22 ms in m in Iron, while in Digor, significant differences are observed only in i (~67 ms). Overall, vowel duration is slightly longer before [+ voice] obstruents, though the differences do not exceed 5 per cent.

Discussion. It was shown that vocal fold vibration is generally absent in [+ voice] obstruents in final position, which means final devoicing is present in Ossetic. At the same time, differences in segment voicing, aspiration and frication duration imply that the neutralisation may be incomplete, and further perceptual studies are required. The two varieties seem to utilise the same production cues, although the deltas associated with different voicing categories can have different magnitude (cf. segment voicing and aspiration duration).

Our findings suggest that Ossetic does not use some cues for voicing typical of other Iranian languages (most importantly, closure and preceding vowel duration, cf. Shughni [2] and Persian [4]). This may be explained by the language's unique geographical position, isolated from other Iranian languages and surrounded by non-Indo-European languages of the Caucasus. Moreover, rather wide use of plosive geminates in various morphophonemic processes can be the reason why the contrast based on closure duration is reserved for lexical oppositions not related to [± voice].

Finally, our results showcase the importance of considering stimuli in different contexts rather than in just one (usually in isolation) as is the case in some studies on voicing in final

[3] Cf. Sokolova's '[final plosives] are either always or optionally either fully or half-voiceless.' [1, p. 39]

position. When word- and phrase-final positions coincide, fortition effects associated with final position are particularly strong; phrase-medial prevocalic position lessens these effects while the obstruent technically remains in (word-)final position.

	[+ voice]	*[– voice]*
Segment voicing	**Iron**: <40% of closure and <35% of frication is voiced **Digor**: ~15% of closure is voiced; no manner effect *No effect of context.*	**Iron**: <10% is voiced **Digor**: ~15% of closure is voiced; no manner effect *No effect of context.*
Closure duration (means)	**Iron**: 120 ms **Digor**: 142 ms *No effect of context.*	**Iron**: 130 ms **Digor**: 136 ms *No effect of context.*
Aspiration duration (means)	**Iron**: 50 ms (i and f), 3 ms (m) **Digor**: 118 ms (i), 9 ms (m), and 84 ms (f)	**Iron**: 93 ms (i and f), 25 ms (m) **Digor**: 157 ms (i), 16 ms (m), and 78 ms (f)
Frication duration (means)	**Iron**: 215 ms (i), 126 ms (m), and 197 ms (f) **Digor**: not enough data	**Iron**: 282 ms (i), 164 ms (m), and 260 ms (f) **Digor**: not enough data
Vowel duration	**Iron**: ~34% (i), ~38% (m), and 36% (f) of the word; 39% before fricatives, 33% before plosives **Digor**: ~27% of the word; no manner effect	**Iron**: ~29% (i), ~35% (m), and 31% (f) of the word; 36% before fricatives, 29% before plosives **Digor**: ~23% of the word; no manner effect

Table 1. *Acoustic parameters cross-linguistically associated with voicing tested for Iron and Digor Ossetic*.

Appendix. (Near-)minimal pairs used in the study (written in IPA)

Iron

1) ard ‘oath’ — art ‘fire; bonfire’
2) arv ‘sky’ — arf ‘depth’
3) aʒ ‘year’ / rɜʒ ‘grow!’ — aʃ ‘height; size’ / baʃ ‘broth’
4) fad ‘foot’ — fat ‘arrow’
5) n(w)aʒ ‘drink!’ — naʃ ‘pumpkin’
6) sɜv ‘hit!’ — sɜf ‘wound’
7) səd ‘motion’ — sət ‘glory; respect’
8) wad ‘walking, running; breeze’ — wat ‘room; bed’
9) ward ‘falling (about precipitation)’ — wart ‘shield’

Digor

1) boz ‘gratitude’ — bos ‘lace’
2) fad ‘foot’ — fat ‘arrow’
3) idʑag ‘full’ — murtak ‘sugar’
4) lub ‘sound of falling’ — gəpː ‘jump (n)’
5) wad ‘running’ — wat ‘room; bed’
6) χərez ‘porous’ — χəres ‘willow’

[1] V. S. Sokolova, *Ocherki po fonetike iranskikh jazykov [Outlines of the phonetics of Iranian languages]*. Moscow, Leningrad: Izdatel’stvo Akademii Nauk SSSR, 1953.

[2] Y. Makarov, ‘Aspiration length and intensity as potential acoustic cues to [voice] in final position: evidence from Shughni’, in *Proceedings of the 20th International Congress of Phonetic Sciences*, R. Skarnitzl and J. Volín, Eds., Prague: GUARANT International spol. s r.o., 2023, pp. 2298–2302. [Online]. Available: https://guarant.cz/icphs2023/582.pdf

[3] G. K. Iverson and J. C. Salmons, ‘Final Devoicing and Final Laryngeal Neutralization’, in *The Blackwell Companion to Phonology*, M. van Oostendorp, C. J. Ewen, E. Hume, and K.

Rice, Eds., Oxford, UK: John Wiley & Sons, Ltd, 2011, pp. 1–22. doi: 10.1002/9781444335262.wbctp0069.
[4] Z. Tofigh and V. Abolhasanizadeh, 'Phonetic Neutralization: The Case of Persian Final Devoicing', *Athens Journal of Philology*, vol. 2, no. 2, pp. 123–134, 2015.

18. Perception of Serbian Restricted Tone by Native and Non-native Listeners: Evidence from an Online Free Classification Task

Dušan Nikolić
University of Calgary

The typological continuum of lexical prosody consists of stress, tone, restricted tone, and no-lexical-prosody languages [1]. Native speakers of languages without lexical prosodic contrasts can often exhibit perceptual insensitivity to lexical prosodic contrasts in other languages [2, 3, 4]. While the previous research on the perception of lexical prosody has focused mainly on a two-way contrast between tone and stress, what has been to a large extent neglected is the perception of restricted tone [5].

In the present study, we examined the perception of restricted tone by native and non-native listeners. We specifically explored whether Serbian and French speakers could distinguish between Serbian restricted tone. Serbian is a language that has four restricted tone types which combine short and long vowels (S and L) and rising and falling pitch (R and F). The restricted tones in Serbian are thus descriptively categorized as short-falling (SF), short-rising (SR), long-falling (LF), and long-rising (LR). Unlike Serbian, French has no prosodic categories that distinguish between lexical items. Therefore, based on the previous research [3, 4], our hypothesis was that the French speakers would fail to distinguish between Serbian restricted tone categories as French has no word-level prosody, and thus their speakers would not be able to encode restricted tone in their phonological representations.

In this study, we recruited 41 native speakers of Serbian and 63 native speakers of French. Both groups of participants completed an online free classification task in which the listeners freely grouped recordings based on their similarity. They listened to 2 groups of 2 nonsense words, [rejav] and [kavan], both of which were patterned according to Serbian phonotactic rules. Each word was produced in 4 Serbian restricted tone types by one male and one female speaker. 16 nonsense words were produced in total (2x4x2). In the free classification task interface, the listeners could click on distinct icons, spaced around a circle, each of which played a unique recording of a nonsense word production. The listeners' task was to place the icons representing each of these words next to each other in the circle, based on similarity (Fig.1). The more similar the nonsense words were, the closer they should be placed.

We compared the classification of the two groups by producing the perceptual similarity matrices which were submitted to multi-dimensional scaling (MDS) (Fig.2) and clustering analyses [6]. The results showed that the majority of Serbian listeners did not consistently hear the difference between short-rising and short-falling restricted tones. Yet, they still managed to create four categories of restricted tone (as the norm implies). The French listeners confused long-falling restricted tones with long-rising and short-falling/rising tones. Since French listeners' prosodic system does not have pitch targets that map acoustic properties to a native restricted tone category or to lexical items, French listeners were perceiving the similarity between the restricted tone categories based on the categories that already exist in their language – AP and IP. The results revealed that non-native listeners were not necessarily perceptually insensitive to lexical prosodic contrasts. Instead, they encoded restricted tone categories through their native language prosodic system, which confirms that language experience shapes the perceptual space of lexical prosodic categories.

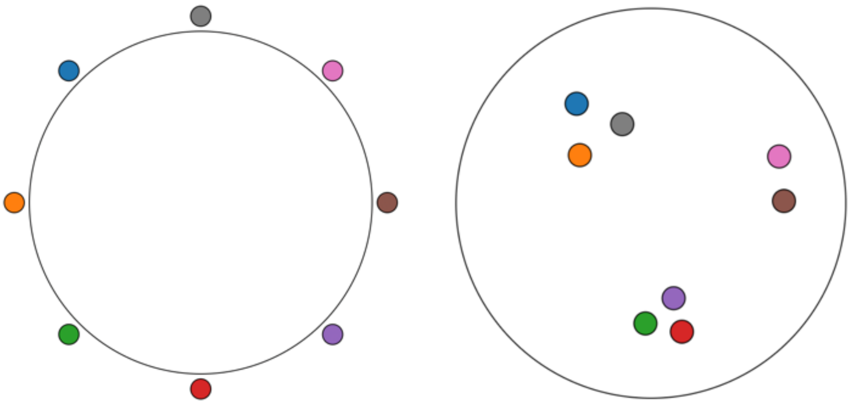


Figure 1. *Each circle represents one recording. The figure on the left shows the circles before they were classified, while the figure on the right shows an example of the circles after their classification.*

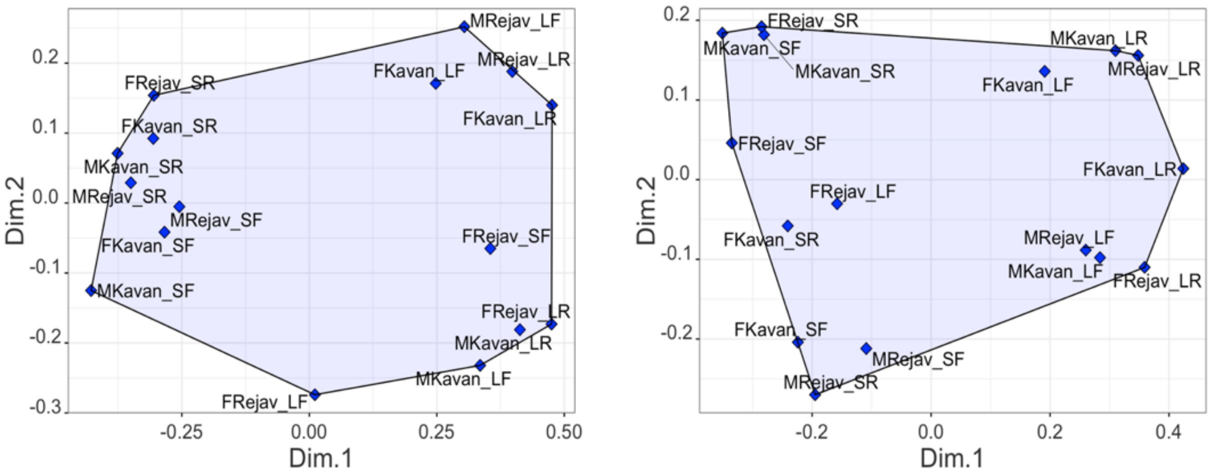


Figure 2. *The MDS output of Serbian listeners' (left) and French listeners' (right) classification of Serbian restricted tones ("M" and "F" before the items denote the gender of the speaker).*

[1] Hyman, L. M. 2006. Word-prosodic typology. Phonology 23(2), 225-257.

[2] Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. 2001. A robust method to study stress "deafness". The Journal of the Acoustical Society of America 110(3), 1606-1618.

[3] Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. 2008. Persistent stress 'deafness': The case of French learners of Spanish. Cognition 106(2), 682-706.

[4] Rahmani, H., Rietveld, T., & Gussenhoven, C. 2015. Stress "deafness" reveals absence of lexical marking of stress or tone in the adult grammar. PloS one 10(12), e0143968.

[5] Baese-Berk, M. M., McLaughlin, D. J., & McGowan, K. B. (2020). Perception of non-native speech. Language and Linguistics Compass 14(7), e12375.

[6] Clopper, C. G. (2008). Auditory free classification: Methods and analysis. Behavior Research Methods 40(2), 575-581.

19. Phonetic effects of morphosyntactic variation for English -s suffixes

Colleen E. Fitzgerald, Jason A. Whitfield, Rebecca Ebert
Bowling Green State University

Recent investigations into English -s suffixes have indicated that plural, possessive, and third person singular morphemes manifest with distinct phonetic realizations. However, the literature shows discrepancies regarding the acoustic noise duration and fricative centre of gravity (CoG) of these bound morphemes compared to non-morphemic -s. Some studies report longer durations for morphemic uses [1, 2, 3, 4, 5, 6] whereas others find non-morphemic -s to be longer [7, 8] or report no significant difference [9]. Few studies included phoneme category analyses, with mixed outcomes regarding whether [s] or [z] had longer durations [1, 2, 7]. Plag, Tomaschek, and colleagues used corpus data (i.e., The Buckeye Corpus) and indicated a need for a replication of their work with experimental data [7, 8]. The present study addressed this need using a controlled reading task to investigate how sentence position, morpheme type, and phoneme category interact to influence spectral properties.

Nine monolingual, native speakers of North American English (*M* age = 27 years) read aloud 80 target allomorphs embedded in 43 sentences designed to balance coda complexity, word and sentence stress, verb transitivity, sentence length, and adjacent phonemes. Each intersection of morpheme type (i.e., plural, possessive, third person singular, non-morphemic), phoneme category (i.e., [s] or [z]), and sentence position (i.e., medial or final) appeared five times in the stimuli (see Fig. 1). Acoustic noise duration and CoG were derived from audio files using a Praat script [10]. Linear mixed models were constructed for duration and for CoG.

The duration model revealed a significant interaction between sentence position, morpheme type, and phoneme category ($F = 5.838$, $p < .001$, $\omega_p^2 = .02$) with additional two-way effects for each pairing. All morpheme types were longer in the final position (see Table 1). The phoneme category [s] was longer than [z] for all morpheme types and sentence positions. Non-morphemic duration generally did not differ from morphemic uses, with the notable exception of a significantly shorter duration of plural [z] (Est = 12.44 ms; SE = 3.59) and shorter duration of third person singular [z] (Est = 15.19 ms; SE = 3.59) when compared to non-morphemic [z] (Est. = 66.20 ms; SE = 6.44) in medial sentence positions. The CoG analysis indicated a small overall effect of sentence position, with final-position fricatives showing lower spectral frequencies ($F = 62.208$, $p < .001$, $\omega_p^2 = .08$). No significant CoG differences emerged across morphemic and non-morphemic fricatives.

This study's experimental design incorporated a range of phonetic and morphosyntactic conditions to replicate duration findings from corpus studies and extend CoG findings. The previous work that had reported a longer duration for morphemic uses of -s had limited the inclusion of morpheme types and sentence positions [2, 3, 5]. In contrast, more inclusive studies such as Plag et al. (2017), Tomaschek et al. (2019), or Macey (2019) reported that non-morphemic instances were longer or not different in line with the results we reported here. Thus, our findings that non-morphemic and morphemic durations generally do not differ except for medial [z] align with prior studies that employed similarly robust methodologies with a full suite of morphemes [7, 8, 9]. We extended the previous finding of a sentence-final lengthening effect for plural and third person singular [4, 7] by confirming that -s related allomorphs are longer in the sentence-final positions irrespective of morpheme type or phoneme category. Prior work had reported an overall effect of morpheme type on CoG without between-morpheme comparisons [6] or sentence-balanced tokens [1], but the present study which included these considerations did not replicate this finding, reporting instead only an effect of sentence position. Future directions include examining the speech productions of multilingual speakers to determine how cross-linguistic phonetic and phonological systems may influence the acoustic realization of morphosyntactic features.

1. The truck's annoying engine stops.
    a. Possessive [s], medial position, coda /k/
    b. Third person singular [s], final position, intransitive, coda /p/
2. She grabs a handful of chips.

a. Third person singular [z], medial position, transitive, coda /b/

b. Plural [s], final position, coda /p/

3. The box up there has mugs.

a. Non-morphemic [s], medial position, coda /k/

b. Plural [z], final position, coda /g/

Figure 1. Excerpt from Stimulus Sentences

Table 1. Mean (SD) for Duration (ms) of Allomorphs

	Phoneme Category			
	Sentence-Medial		Sentence-Final	
Morpheme Type	[s]	[z]	[s]	[z]
Non-morphemic	67.5 (19.3)	67.8 (30.3)	148.6 (41.8)	119.0 (30.6)
Morphemes				
Plural	67.1 (14.3)	53.7 (23.3)	142.6 (43.1)	126.0 (36.9)
Possessive	72.0 (17.5)	59.3 (20.2)	142.5 (43.1)	125.3 (39.7)
Third Person Singular	65.4 (13.0)	51.0 (10.9)	144.0 (39.6)	124.1 (36.6)
Morpheme Mean of Means	68.2 (14.9)	54.7 (18.1)	143.0 (41.9)	125.1 (36.7)

[1] Fitzgerald, C.E., Ebert, R.L., & Whitfield, J. (2019, June). *Acoustic Differences among English Allomorphs in a Children's Book Reading Task.* Poster presentation at the Phonetics and Phonology in Europe Convention, Lecce, Italy.

[2] Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, *33*(1), 32-49.

[3] Schwarzlose, R., & Bradlow, A. R. (2001). What happens to segment durations at the end of a word?. *The Journal of the Acoustical Society of America*, *109*(5), 2292-2292.

[4] Song, J.Y., Demuth, K., Evans, K., & Shattuck-Hufnagel, S. (2013). Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *The Journal of the Acoustical Society of America*, *133*(5), 2931-2946.

[5] Walsh, T., & Parker, F. (1983). The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics*, *11*(2), 201-206

[6] Zimmermann, J. (2016). Morphological status and acoustic realization: Is there a difference between Bra[d] Pitt and a grille[d] cheese omelet, or between Kate Mo[s] and killer robot[s]? In Johannes Wahle, Marisa Köllner, Harald R. Baayen, Gerhard Jäger & Tineke Baayen-Oudshoorn (Eds.), Proceedings of 6th Conference on Quantitative Investigations in Theoretical Linguistics (QITL-6). Tübingen: Universität Tübingen. http://hdl.handle.net/10900/68949.

[7] Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, *53*(1), 181-216.

[8] Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2019). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics*, 1-39.

[9] Macey, N. M. (2019). S morpheme durations in English: An experimental approach (Doctoral dissertation, Yale University).
[10] Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [HP]. Version 6.0.19 retrieved 13 June 2016 from http://www.praat.org/.

20. Exploring phonological rule application in code-switching among Spanish-English bilinguals

Hironori Katsuda[1], Gemma Repiso-Puidgelliura[2], Kie Zuraw[3], Jinyoung Jo[3]
[1]University of Kansas, [2]Universitat Autònoma de Barcelona, [3]University of California, Los Angeles

Introduction: Bilinguals maintain vocabulary from both languages. However, it remains largely unknown how a phonological rule from one language might interact with a word from the other. In a code-mixed utterance, the structural description of a phonological rule could be partially provided by a word in the other language. Does the rule still apply in such cases? Olson [1] examined how Spanish-English bilinguals apply Spanish s-voicing (/s/ becomes [z] before a voiced consonant) and spirantization (a voiced stop becomes a fricative after a continuant) when a language switch occurs between target and environment. Olson found that bilinguals could apply s-voicing when its environment came from English (*escuchas noises* "hear noises"), though with less voicing than in all-Spanish contexts. In contrast, spirantization was much less frequent when the environment was from English (*flee guerras* "flee wars"), with no significant difference from all-English contexts. Olson speculated that phonological rules can apply as long as the target is from the rule's language and attributed reduced spirantization to hyperarticulation, which strengthens segments following a language boundary. This study aims to expand on Olson's work by (i) including two additional English rules–tapping (/t, d/ becomes a tap between vowels) and the rhythm rule (stress shifts to a preceding syllable)–alongside the two Spanish rules Olson studied, (ii) adding a baseline condition where rules are not expected to apply, and (iii) using measures of language dominance and code-switching. This abstract presents a preliminary analysis of the two Spanish rules based on data from 31 out of 49 participants for spirantization and 39 out of the same 49 participants for s-voicing. We plan to analyze data from the remaining participants as well as for tapping.

Methodology: Participants were Spanish-English bilingual undergraduates. We created sentences containing four quadruplets of stimuli (Sp#Sp: [e.g., *hombre gana*], Sp#Eng: [e.g., *hombre gossips*], Eng#Eng: [e.g., *boy gossips*], Eng#Sp: [e.g., *boy gana*]) for each rule, in both an experimental condition and a baseline condition (environment that does not trigger rule application). We used the Bilingual Language Profile [2] to measure relative proficiency, and Bilingual Code-Switching Profile [3] to assess code-switching behavior and attitudes. To assess rule application, we measured CV ratio (intensity ratio of the target consonant and following vowel) for spirantization and Percent Voiced (percentage of voiced portion in /s/) for s-voicing [1].

Results: Our results show greater rates of spirantization (Figure 1a) with a Spanish target (Eng#Sp, Sp#Sp) than with an English target (Eng#Eng, Sp#Eng). Thus, differently from Olson's [1], our speakers also spirantize in the Eng#Sp condition, providing stronger evidence that rules can apply as long as the target is from the rule's language. Unexpectedly, s-voicing results (Figure 1b) show negligible voicing across conditions. We suspect this may be due to dialectal differences between Olson's participants and ours. Finally, our preliminary results show a correlation between spirantization and code-switching profile. Specifically, CV ratios decrease–meaning there is less spirantization–as participants report higher rates of code-switching experience (Figure 2). In the code-switching conditions, this may suggest that greater code-switching experience helps participants separate the two languages more, but this does not explain the same trend in the all-English and all-Spanish condition. We will present data from all participants, including results for tapping, along with the corrections between rule application and language dominance/code-switching profiles.

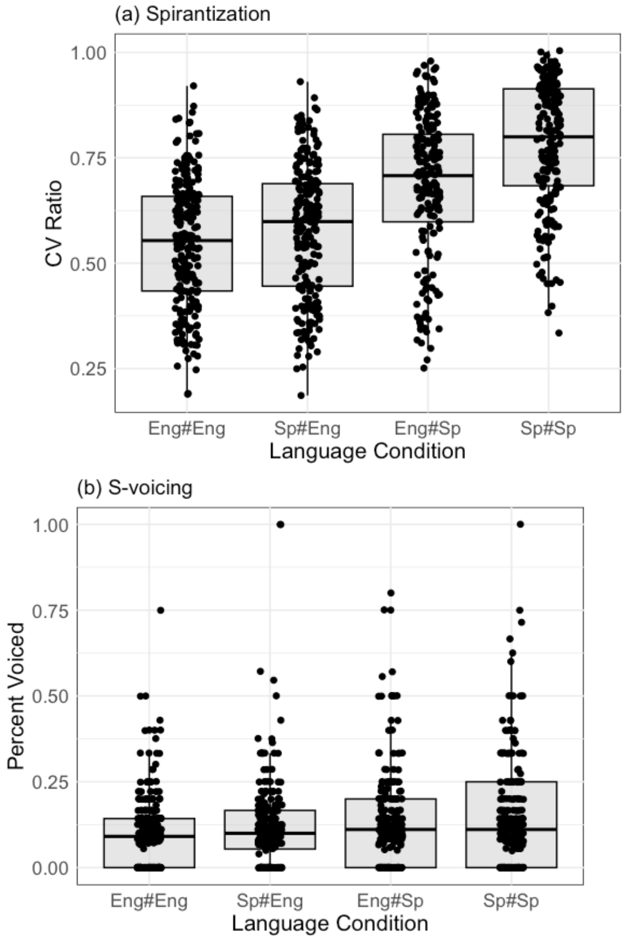


Figure 1. *Results of spirantization (a) and s-voicing (b). CV ratios and Percent Voiced are plotted by repetition and language condition, using only data from the experimental condition.*

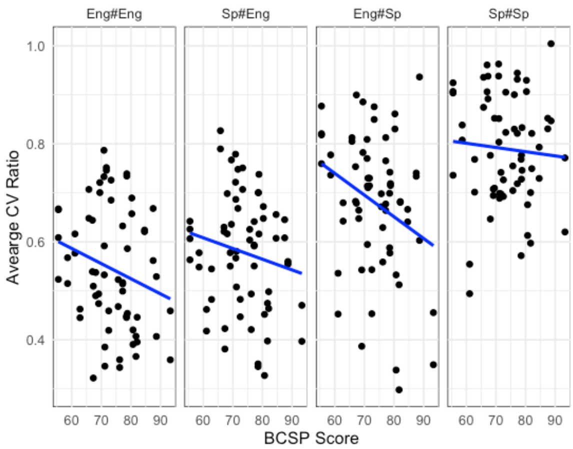


Figure 2. *Correlation between CV ratios (averaged by participant and language condition) and Bilingual Code-Switching Profile scores, based only on the experimental condition.*

[1] Olson, D. J. 2019. Phonological processes across word and language boundaries: Evidence from code-switching. *Journal of Phonetics*, *77*, 100937.

[2] Birdsong, D., Gertken, L.M., & Amengual, M. 2012. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin.

[3] Olson, D. J. (2022). The Bilingual Code-Switching Profile (BCSP): Assessing the reliability and validity of the BCSP questionnaire. *Linguistic Approaches to Bilingualism*, 1–34. https://doi.org/10.1075/lab.21039.ols

Poster session 3b

1. Perception of emotions across Portuguese varieties

Diana Santos, Marisa Cruz
University of Lisbon

It is already known that emotions are conveyed through pitch contour, tone, intensity, and rhythmic properties of speech ([1]), and that emotional perception in speech, while universal, is influenced by linguistic and cultural factors ([2-4]). However, as far as we know, emotional perception across language varieties is not explored yet. Portuguese is a good case study as Brazilian (BP) and European (EP) varieties are similar from a rhythmic perspective – mixed pattern ([5, 6]) – but different from an intonational point of view – high tonal density in BP, while EP is more tonally sparse ([7, 5]). This study aims to investigate whether these prosodic differences influence the perception and recognition of emotions in BP and EP or, in other words, whether speakers' linguistic backgrounds may shape their emotional perception.

A pilot perceptual task was conducted with 16 participants (8 BP and 8 EP native speakers), who were exposed to segmentally masked (i.e., low-pass filtered at 400Hz) sentences produced by actresses from both varieties expressing 4 different emotions – happiness, sadness, anger, and neutral. These productions were acoustically analysed in terms of F0 height and contour, intensity, and speech rate to ensure that stimuli used in the perceptual task are similar within variety and emotion. An identification task was run with a training and a test block to assess participants' accuracy and reaction time in identifying emotions through an open response options list (neutral, happiness, sadness, anger, other). A total of 48 responses per participant was analysed.

Chi-square tests were run to observe whether participants' accuracy (i.e., the correctness of their responses in the identification of emotions) is related to the (i) participants' native variety (BP and EP), (ii) the variety being perceived (native vs. non-native), and (iii) the emotion under perception (neutral, happiness, sadness and anger). For the inspection of reaction times (RTs), a generalized linear mixed model (GLMM) was run with (i) participants' native variety, (ii) variety perceived, (iii) emotion under perception, and (iv) correctness of responses as fixed effects. In order to see whether RTs are influenced by each emotion, and depending on the variety perceived, interactions were also included in the model as fixed effects, namely: (i) correctness*emotion under perception, (ii) correctness*emotion under perception*native variety, and (iii) correctness*emotion under perception*variety perceived.

Our results for the accuracy show that participants' correct responses are significantly related to their native variety (χ^2 (1)=36.58, p<.001), and to the emotion under perception (χ^2 (3)=39.86, p<.001). The variety being perceived (native vs. non-native) does not play a relevant role (p>.05). Indeed, in Fig. 1, we may observe that BP participants are better than the EP ones in identifying emotions, independently of the variety being perceived. Additionally, among emotions, neutral and sadness globally trigger higher correctness rates (except for EP participants perceiving native neutral sentences), whereas happiness and anger are globally more difficult to be identified (except for BP participants perceiving non-native anger – 60%). Considering RTs, the GLMM revealed a significant effect of correctness (F (1, 685)=22.11, p<.001), of the emotion under perception (F (3, 685)=17.39, p<.05) and of the interaction correctness*emotion under perception*variety perceived (F(7, 685)=2.64, p<.05). Namely, correct responses are significantly faster (M=1516ms) than incorrect responses (M=2396ms), and this is influenced by the emotion under perception and the variety perceived. More precisely, sadness, produced either by BP and EP speakers, triggers the shortest RTs (M_{BP}=950ms, M_{EP}=1034ms), whereas anger produced by BP speakers triggers the longest RTs (M=3810ms). These results seem to suggest a higher sensitivity of Brazilian Portuguese speakers in perceiving emotions in speech, thus confirming our hypothesis that speakers' linguistic brackgrounds may shape their emotional perception, even within a given language.

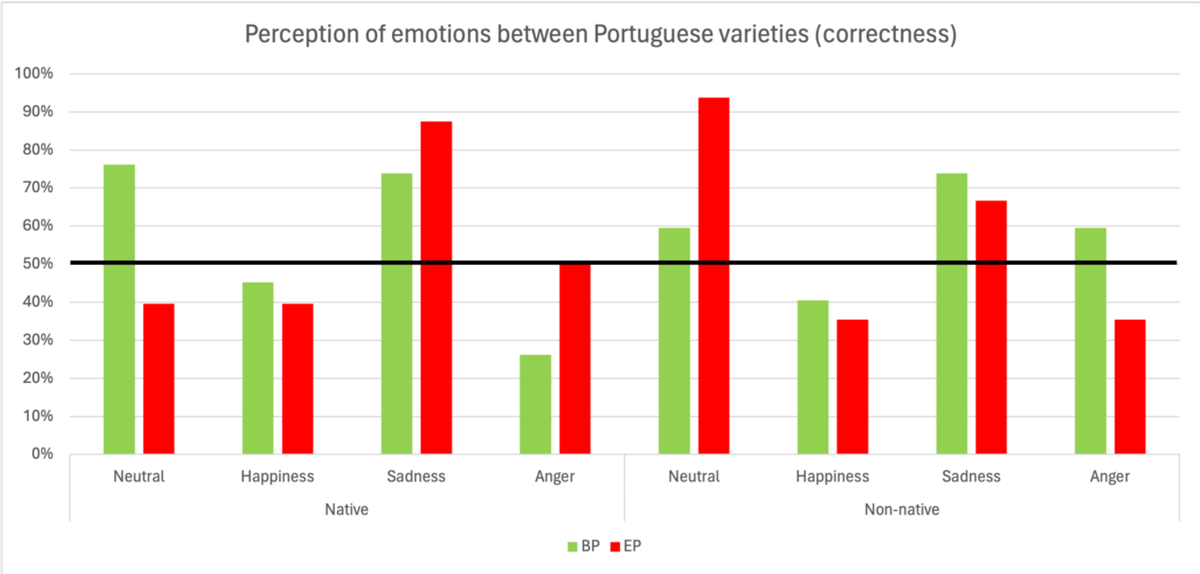


Figure 1. *Proportion of correct responses given by BP and EP participants (green and red bars, respectively) per emotion and variety perceived (native vs. non-native). The black solid line represents the at-chance level.*

[1] Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

[2] Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169–200. https://doi.org/10.1080/02699939208411068

[3] Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior, 33*(2), 107–120. https://doi.org/10.1007/s10919-008-0065-7

[4] Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication, 50*(6), 519–530. https://doi.org/10.1016/j.specom.2008.03.006

[5] Frota, S., & Vigário, M. (2000). Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. In R. V. Castro & P. Barbosa (Eds.), *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística*, Vol.1 (pp. 533-555). APL.

[6] Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus, 13*, 247-273. doi: 10.1515/prbs.2001.005

[7] Frota, S., Cruz, M., Fernandes-Svartman, F., Collischonn, G., Fonseca, A., Serra, C., Oliveira, P., & Vigário, M. (2015). Intonational variation in Portuguese: European and Brazilian varieties. In S. Frota & P. Prieto (Eds), *Intonation in Romance* (pp. 235-283). Oxford University Press.

2. The source and significance of consonant-intrinsic pitch: New evidence from Danish

Rasmus Puggaard-Rode[1], James Kirby[1], Nicolai Pharao[2]
[1]*Institute for Phonetics and Speech Processing, LMU Munich,* [2]*University of Copenhagen*

It is well-established that laryngeal contrasts in consonants can exert a localized influence on pitch. These effects, which [1] dubs CF0, have been documented in many languages with various kinds of laryngeal contrasts. There is a substantial literature on CF0, but several open questions remain about the underlying mechanisms. For example, it remains unclear if CF0 is actively controlled, and e.g. serves to enhance a phonological [voice] feature [2,3], or if it is a by-product of gestures aimed at inhibiting phonation [4], and in the latter case, what those gestures might look like. It also remains an open question how exactly CF0 interacts with other more and less localized sources of prosodic modification. In this study, we present data from two different varieties of Danish which shed new light on these open questions.

Danish contrasts a series of aspirated stops [p^h t^h k^h] with a series of voiceless unaspirated stops [p t k] which are actively devoiced due to light glottal spreading, achieved through tensing of the posterior cricoarytenoids [5,6]. The physiology of the contrast has been studied extensively using e.g. electromyography [6]. If both series of stops trigger CF0 effects, this would suggest that CF0 is indeed a result of devoicing gestures specifically, and that at least in Danish, it is not used to signal an underlying phonological contrast (*contra* [3]). The existing physiological literature can further help us potentially deduce *which* devoicing gestures are responsible for the CF0 effects. Previous studies suggest that both series of stops do indeed trigger CF0 effects [7,8], although these studies have certain methodological limitations. Here, we test whether the previous results replicate with more speakers and an experimental design with somewhat higher ecological validity. Danish is furthermore a good test case for investigating how competing demands on the larynx interact, since the language has a voice quality contrast (known as *stød*), which is, among other things, cued by pitch differences immediately after an initial consonant; the physiology of the voice quality contrast has also been investigated extensively [9]. Stød may either enhance CF0 effects due to synergistic co-activation of gestures which raise pitch [10], reduce CF0 effects if other demands on the larynx restrict the range of F0 modulations.

For this study, we recorded 29 speakers of Danish coming from two different dialect areas (Greater Copenhagen or Central Jutland). The speakers read aloud 132 alternative question sentences. Items of interest were crossed to allow for testing interactions between onset category (nasal, unaspirated, aspirated), place of articulation, vowel height, and stød. We calculated pitch trajectories over the items of interest using a cross-validation method based on comparisons of estimates from Praat [11] and REAPER [12]. The resulting pitch trajectories are analyzed using generalized additive mixed models, fitted using the mgcv library in R [13,14].

Figure 1 shows model-predicted trajectories. We find that, across all conditions, both series of stops increase the pitch of the following vowel relative to nasals (which we consider a neutral baseline, following e.g. [15]), and aspirated stops raise pitch more than unaspirated stops. This result supports the analysis of CF0 as essentially a by-product of gestures that inhibit phonation [16] and not as an enhancement of an underlying [voice] feature [2,3]. However, it also challenges previous accounts of CF0 as specifically a result of cricothyroid tensing [17], since there is no evidence of cricothyroid tensing in the Danish unaspirated stops [6]. We also find that the temporal extent of CF0 is reduced in the stød context, presumably since stød places competing demands on the pitch trajectory which limit the range that CF0 can have without endangering other contrasts. This suggests that strengthening of CF0 in high pitch environments, which has been observed in previous studies (e.g. [10]), does not simply depend on F0 level, but rather depends on the precise nature of the gestures and contrast involved.

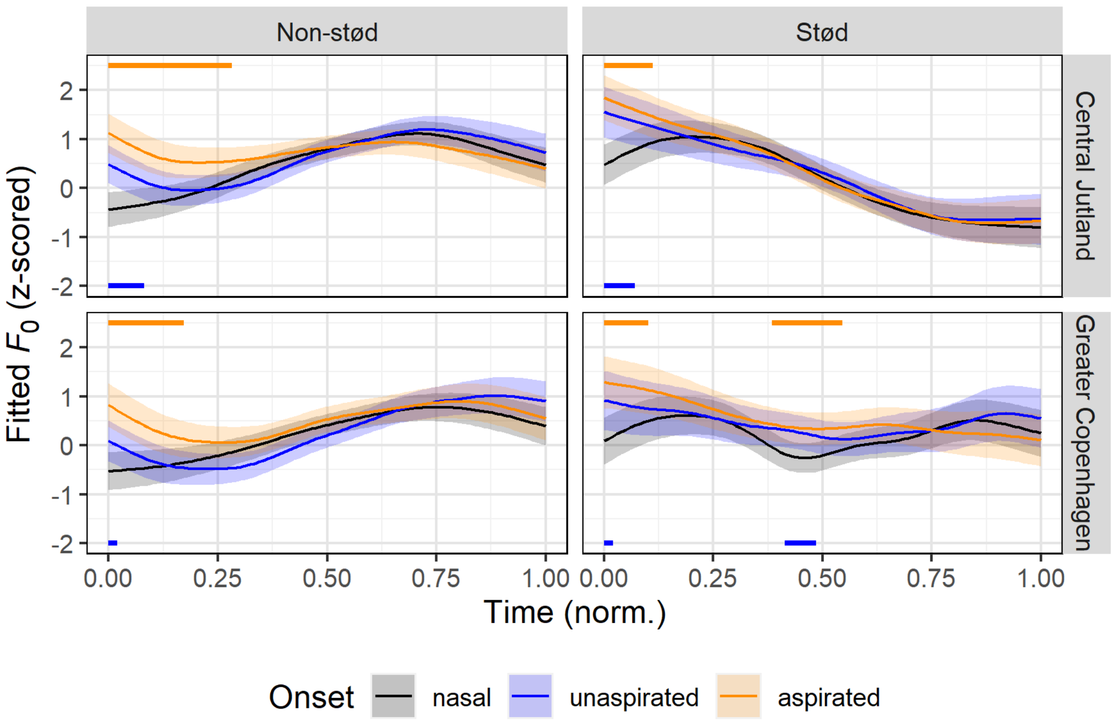


Figure 1. *Model predicted pitch trajectories with 95% confidence intervals, faceted by phonemic voice quality and colored by onset category. Straight lines at the top and bottom of each facet indicate parts of trajectories that are significantly above the nasal baseline.*

[1] Kingston, J. 2007. Segmental influences on F0. Automatic or controlled? In Gussenhoven, C. & Riad, T. (Eds.), *Tones and tunes. Volume 2: Typological studies in word and sentence prosody*, 171–210. Berlin & New York: Mouton de Gruyter.

[2] Keating, P.A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60(2), 286–319.

[3] Kingston, J. & R.L. Diehl. 1994. Phonetic knowledge. *Language* 70(3), 419–454.

[4] Goldsmith, L.M. & C.P. Browman. 1986. Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics* 14(2), 339–342.

[5] Fischer-Jørgensen, E. & H. Hirose. 1974. A preliminary electromyographic study of labial and laryngeal muscles in Danish stop consonant production. *Status Report on Speech Research* 39/40, 231–253.

[6] Hutters, B. 1985. Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica* 42(1), 1–24.

[7] Jeel, V. 1975. An investigation of the fundamental frequency of vowels after various consonants, in particular stop consonants. *ARIPUC* 9, 191–211.

[8] Petersen, N.R. 1983. The effect of consonant type on fundamental frequency and larynx height in Danish. *ARIPUC* 17, 55–86.

[9] Fischer-Jørgensen, E. 1989. Phonetic analysis of the stød in Standard Danish. *Phonetica* 46(1/2/3), 1–59.

[10] Chen, Y. 2011. How does phonology guide phonetics in segment-f0 interaction? *Journal of Phonetics* 39(4), 612–625.

[11] Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Institute of Phonetic Sciences, Universiteit van Amsterdam, Proceedings* 17, 97–110.

[12] Talkin, D. 2015. *REAPER. Robust Epoch and Pitch EstimatoR*. Computer program.

[13] Wood, S.N. 2017. *Generalized additive models. An introduction with R*. 2nd ed. Boca Raton: CRC Press.
[14] R Core Team. 2023. R. A language and environment for statistical computing, v4.3.2. Computer program.
[15] Hanson, H.M. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal of the Acoustical Society of America* 125(1), 425–441.
[16] Kirby, J.P. & D.R. Ladd. 2016. Effects of obstruent voicing on vowel F0. Evidence from "true voicing" languages. *Journal of the Acoustical Society of America* 140(4), 2400–2411.
[17] Löfqvist, A., T. Baer, N.S. McGarr & R.S. Story. 1989. The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85(3), 1314–1321.

3. Lexical stress and vowel length in South Tyrolean German

Aleese Block[1], Marianne Pouplier[1], Alessandro Vietti[2]
[1]Ludwig-Maximilians University of Munich, [2]Free University of Bozen-Bolzano

In the current study, we aim to describe the interaction between lexical stress and contrastive vowel length in the variety of German spoken in South Tyrol in Northern Italy. In Standard German, tenseness/laxness and vowel length are inseparably connected; long vowels are tense and short vowels are lax [1]. At the same time, duration is a primary correlate of lexical stress in German with intensity as a secondary correlate. In unstressed position, the vowel duration difference is thus neutralized while the qualitative contrast persists [2]. This way, the vowel contrast can be maintained in the face of stress induced durational variation [2] Interestingly, in Bavarian German dialects, the phonological vowel length contrast is not qualitative; the contrast is carried by a complementary relationship between vowel and consonant duration [3]. Here we investigate the interaction of phonological vowel length and lexical stress in Tyrolean German to understand how cues such as vowel duration, vowel quality, and intensity are used.

We collected data from 35 (22F, 13M) native speakers of the Meran dialect of German completing a read speech task. We recorded fourteen monophthongs, long and short /a, ɒ, e, ɛ, o, i, u/ in both lexically stressed and unstressed position. We measured intensity, duration (vowel and postvocalic consonant), and the first two formants using the Praat plug-in *FastTrack* [5]; midpoint values were derived from the average of the middle 10% of the vowel.

First, we analyzed the acoustic correlates of phonological vowel length. Vowel duration was predictive of vowel length; long vowels were on average 42% longer than short vowels. Consonants after short vowels were also 21% longer than those after long vowels. Using a Pillaiscore based analysis, we found that the distributions of long and short vowels in the F1/F2 space were not distinct (Pillai < 0.3). This confirms that vowel length is purely durational in this dialect. We also analyzed whether duration and intensity were predictive of lexical stress in our data. We found that stressed vowels were 20% longer and more intense than unstressed vowels, in line with previous work on German stress [6-7]. However, stress didn’t change the overall duration of postvocalic consonants. Next, we examined how vowel length and lexical stress interacted. For vowel duration, we found a significant interaction between Length and Stress, and this interaction was further explored with Tukey’s post hoc pairwise comparisons. While the overall duration of long and short vowels was longer in unstressed syllables, ratio of long-to-short vowel durations remained the same; the ratio of long-to-short postvocalic consonants was also consistent across stress. This shows that while stress influences the overall duration for vowels, the durational contrast between long and short vowels is maintained. We did not find an interaction between vowel length and stress for intensity, suggesting intensity is only correlated with stress. We did not expect nor find and effect of stress on how much long and short vowel formant distributions overlapped via Pillai-score analysis.

We will discuss two main aspects of our findings. First, we will explore the differences between vowel length in the South Tyrolean variety and Standard German, highlighting the importance of dialect-level variation in our understanding of language systems. We will also discuss the relationship between the acoustic correlates of vowel length and lexical stress and expand upon the implications of these findings for our understanding of segmental and suprasegmental contrast.

[1] Kohler, Klaus J. (1990). "Segmental Duration in German: Syllables, Sentences, and Rhythmic Structure." *Journal of Phonetics*, 18(2), 131–151.
[2] Jessen, M. (1993). Stress conditions on vowel quality and quantity in German. *Working Papers of the Cornell Phonetics Laboratory*, *8*, 1-27.

[3] Kleber, F. (2020). Complementary length in vowel–consonant sequences: Acoustic and perceptual evidence for a sound change in progress in Bavarian German. *Journal of the International Phonetic Association*, *50*(1), 1-22.
[4] Insam, M., Der Lautstand des Burggrafenamtes von Meran. 1936, Leipzig: Hirzel.
[5] Barreda, S. (2021). Fast Track: fast (nearly) automatic formant-tracking using Praat. Linguistics Vanguard, 7(1).
[6] Lunden, A., Campbell, J., Hutchens, M., & Kalivoda, N. (2017). Vowel-length contrasts and phonetic cues to stress: An investigation of their relation. *Phonology*, *34*(3), 565-580.
[7] Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, *3*(1).

4. On the link between auditory acuity and production variability in continuous speech

Floris Cos, Lieke van Maastricht, Esther Janse
Centre for Language Studies, Radboud University, Nijmegen

Successful speech production requires the felicitous and consistent realization of the language's individual segments. One prominent way of monitoring the quality of the speech signal is through auditory feedback. Indeed, the mechanisms underlying speech production and perception have been shown to be tightly linked. This is demonstrated, for example, by online perturbation studies, where speakers rapidly adapt their phoneme production to compensate for a manipulation in the feedback signal [1]. Phoneme production has therefore been argued to be (at least partly) based in auditory terms [2]. Moreover, those speakers who are better able to discriminate between different phonemes, and are better at perceiving minute variations in their speech production, can more finely tune their motor programs. This results in phonemes that are produced with relatively little variability. Such a correlation between auditory acuity and (reduced) production variability has regularly been observed (e.g., [3, 4]).

However, up until now, this link has been investigated exclusively for the production of carefully pronounced (non-)words in isolation. Yet speech is naturally uttered in longer, continuous sentences, which are subject to coarticulation and reduction effects. If production norms are indeed tuned on the basis of auditory feedback, this must be the case for naturalistic speech production as well. The present research is the first to provide a tentative insight into the link between speech production and perception based on continuous speech.

In the current study, 30 Dutch participants read short Dutch texts based on news items. The texts were modified to elicit the production of 149 vowels in stressed syllables. Following [4], these vowels were spread across the categories /ɑ, o, ɛ/ and /ɪ/. Speech recordings were automatically annotated using the *Forced Alignment 2* algorithm [5]. F1 and F2 frequencies were estimated and converted to the Bark scale. A vowel's production variability was expressed as the area of an ellipse, with participant-specific *SD*s of F1 and F2 as its axes (Figure 1) [4]. To determine participants' auditory acuity, we used a discrimination task modelled after [3, 4]. Replicating [4], this resulted in a participant-specific Discrimination Score for each of two vowel continua (/ɑ-o/ and /ɛ-ɪ/).

Based on the current sample of continuous speech, no reliable relationship was observed between Discrimination Scores and variability Ellipse Area (Figure 2). Both in production and perception, participants exhibited considerable individual differences that did not appear to be systematically related (Figure 1). Exploratory analyses on the level of individual productions, such as the Euclidean distance of vowel observations to the centre of the corresponding vowel cluster, did not show the expected relations either.

We have no reason to assume that the relationship between auditory acuity and production variability does not exist. Nevertheless, the current results imply that the explanatory power of Discrimination Score on vowel production variability is small ($F(1, 49.97) = 0.006$) compared to coarticulatory effects, as shown by the effect sizes of the preceding ($F(6, 99.82) = 2.840$) and following consonant ($F(5, 98.49) = 2.605$) on vowel formants. Other factors, such as vowel duration, centralization, and position in the phonological phrase are also likely to contribute to variability of vowel production in continuous speech, further obscuring the relationship of interest. As such, we would like to present these results as a call for more research into the connection between auditory acuity and phoneme production, specifically using continuous and spontaneous speech. Given the suggested influence of coarticulation and reduction effects on the production of phonemes, employing more naturalistic data in phonetic experiments will complement our knowledge of the feedback mechanisms involved in speech production. In turn, this will allow us to better appreciate previously demonstrated effects in a more ecologically valid context.

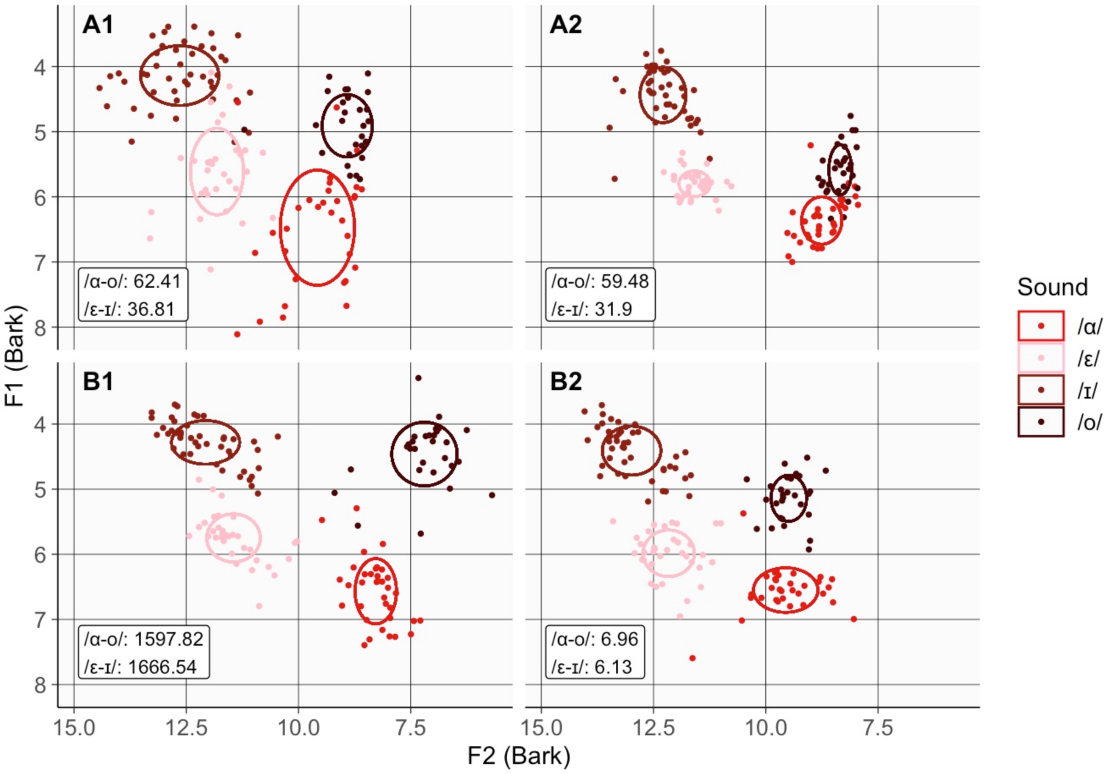


Figure 1. *Vowel production data for four participants. Dots represent individual observations; ellipses indicate production variability. Panels A1 and A2 exemplify that participants with similar Discrimination Scores (white labels) may exhibit differences in vowel production variability. Contrastively, panels B1 and B2 show that participants may have similar vowel production variability despite large Discrimination Score differences.*

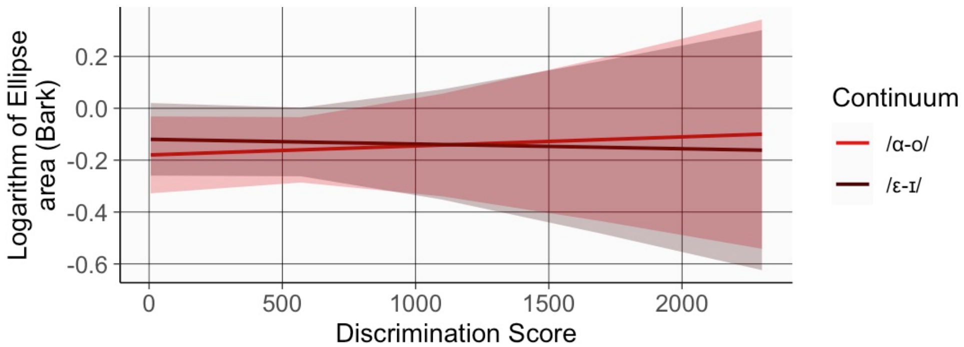


Figure 2. *The relationship between Discrimination Score as a perceptual measure, and Ellipse area as a measure for production variability, for each of the two vowel continua.*

[1] Houde, J.F., Jordan, M.I. 1998. Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213-1216.

[2] Tourville, J.A., & Guenther, F.H. 2011. The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952-981.

[3] Perkell, J.S., Lane, H., Ghosh, S., Matthies, M.L., Tiede, M., Guenther, F.H., & Ménard, L. 2008. Mechanisms of vowel production: auditory goals and speaker acuity. *Proceedings of the Eighth International Seminar on Speech Production* (Strasbourg, France), 29-32.

[4] Franken, M.K., Acheson, D.J., McQueen, J.M., Eisner, F., Hagoort, P. 2017. Individual variability as a window on production-perception interactions in speech motor control. *The Journal of the Acoustical Society of America*, *142*(4), 2007-2018.
[5] Ten Bosch, L. n.d. *Forced Alignment 2*. Computer program.

5. Speaking rate effects on stop consonant coarticulatory resistance in initial and final positions in French

Mohammad Abuoudeh
Department of Language and Linguistics, Al-Hussein Bin Talal University, Ma'an - Jordan

Speech production can be influenced by many extrinsic factors such as speaking rate. We can produce a slow speaking rate and a fast speaking rate according to the situation. The articulators adapt their movements to manage these temporal differences (Lindblom, 1990). This coarticulatory adaptation may modify vowel and consonant qualities as well as their coarticulatory relationship (Agwuele et al., 2008; Lindblom, 1963). Therefore, the coarticulatory resistance (CR) of the consonants could be influenced by these temporal variations. As a matter of fact, the CR is considered as the capacity of a segment to stop the coarticulation (Bladon and Al-Bamerni, 1976; Recasens, 1985). It could be measured by calculating the standard deviation of the beginning and the middle of the formants (primarily the F2onset and F2mid) for one consonant in many vocalic contexts. A high standard deviation indicates a low CR, whereas a low standard deviation means a high CR. The degree of coarticulation of a consonant, and hence the CR, is also impacted by the syllabic position (ex. initial and final). The consonants produced in the final position are more coarticulated (less resistant) than when they are produced in the initial position (Krull, 1988; Lacerda, 1985; Sussman et al., 1997). This study aims to examine the impact of speaking rate on the CR of French stop consonants in initial and final positions and on consonant and vowel qualities. The implications of the temporal changes on consonant classification will also be evaluated.

Ten native speakers of French (5 females and 5 males) produced C1VC and CVC1 real words embedded in a carrier sentence "il a dit... huit fois" in slow, normal, and fast speaking rates. The studied consonants (C1) were /p,t,k,b,d,g/ produced with four vowels /i,a,u,y/. Each word was repeated 7 times by each speaker. The frequency of the beginning (Fonset), the middle (Fmid), and the end (Foffset) of the first three formants were measured in addition to the vowel duration (from the beginning to the end of formant transitions). The data from males (M) and females (F) were treated separately to avoid gender influence on formant frequency analysis.

Linear mixed models (LMM) indicate that the duration in the 3 speaking rates are significantly different for M ($F_{(2,4)}=10.1$, $p<.05$) and F ($F_{(2,4)}=32.98$, $p<.01$). The visual evaluation of the vertical dispersion of $F2_{onset}$, $F2_{offset}$, and $F2_{mid}$ in function of time (as in Figure 1) indicates that temporal variation influence depends on vowels, consonants and consonant position. The standard deviation was calculated to estimate the CR for consonants in the 3 speaking rates (Table 1). Results indicate that, in most cases, the CR of consonants increases (less dispersion) when the speaking rate is accelerated. In most cases, the consonants produced in the final position seem less resistant than in the initial position. In the other hand, LMM indicate that the $F2_{mid}$ is significantly influenced by speaking rate in C1VC and CV1C conditions for both groups ($F2_{midCV}$: $F_{(2,2518)}=54.89$, $p<.001$, $F2_{midVC}$: $F_{(2,2520)}=36.34$, $p<.001$ for M, and $F2_{midCV}$: $F_{(2,2509)}=29$, $p<.001$, $F2_{midVC}$: $F_{(2,2518)}=14.26$, $p<.001$ for F). In addition, the frequencies of $F2_{onset}$ are also impacted by the speaking rate for both groups ($F_{(2,2518)}= 11.98$, $p<.001$ for M, and $F_{(2,2509)}=6.46$, $p<.001$ for F). However, the frequencies of the consonant in the final position ($F2_{offset}$) do not seem to be significantly influenced by time variations for both groups ($F_{(2,2520)}= 1.49$, $p=.225$ for M, and $F_{(2,2518)}=2.37$, $p=.093$ for F). In sum, the results in the present study indicate that time variations influence consonant and vowel qualities besides their coarticulatory relationship in different degrees. The impact of temporal variations is less important when consonants are produced in the final position. These assumptions were confirmed by linear discriminant analysis (LDA) that showed that the consonant classification accuracy is influenced by the speaking rate as well as by the consonant position.

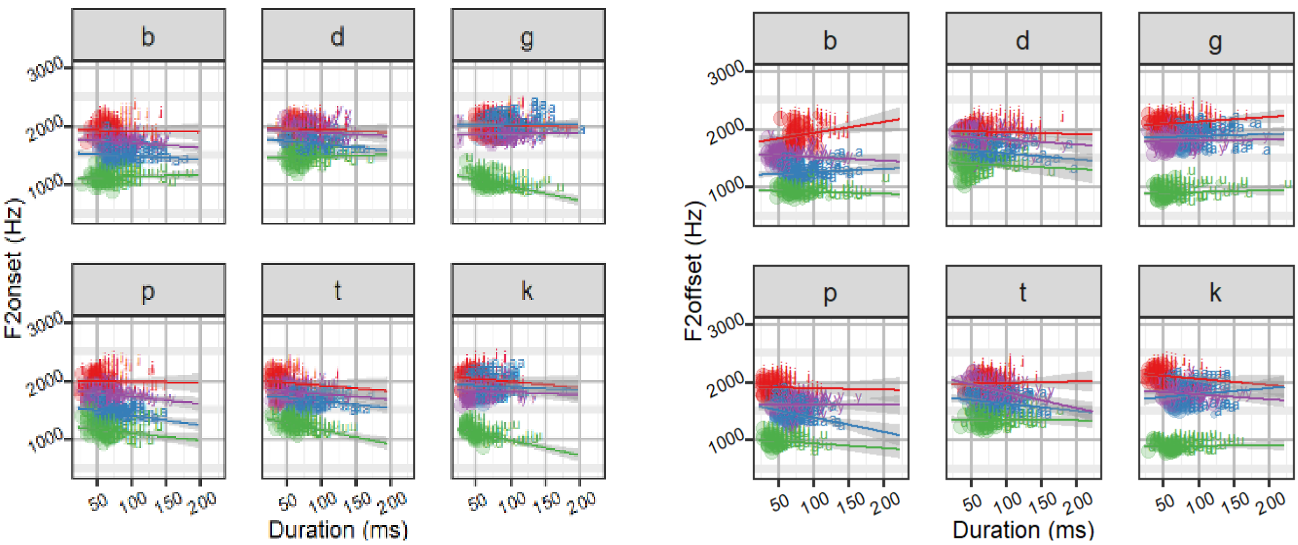


Figure 14. *Values of $F2_{onset}$ C1VC (left) and $F2_{offset}$ CVC1 (right) of the studied consonants in the context of the 4 vowels in function duration produced by male speakers.*

		M				**F**			
C	Rate	$F2_{onset}$	$F2_{offset}$	$F2_{midCV}$	$F2_{midVC}$	$F2_{onset}$	$F2_{offset}$	$F2_{midCV}$	$F2_{midVC}$
b	Slow	317	400	428	470	432	394	500	589
	Normal	315	393	424	436	449	436	547	550
	Fast	320	370	417	401	436	378	547	478
d	Slow	220	264	357	389	229	270	457	485
	Normal	229	271	323	338	242	239	399	440
	Fast	229	257	300	295	223	268	356	409
g	Slow	468	491	390	452	595	541	528	597
	Normal	441	470	374	442	571	622	479	603
	Fast	399	464	326	438	582	611	456	580
p	Slow	359	371	368	414	443	393	468	527
	Normal	356	359	334	397	423	364	435	479
	Fast	330	345	315	376	409	398	443	463
t	Slow	330	281	352	365	440	325	451	442
	Normal	308	282	327	334	392	334	407	406
	Fast	272	277	264	281	353	279	339	365
k	Slow	420	468	368	448	457	558	487	564
	Normal	395	479	330	452	494	579	436	600
	Fast	361	469	302	428	476	615	419	573

Table 1. *Standard deviation in the 3 speaking rates of the $F2_{onset}$ and $F2_{mid}$ when the target consonants are produced in initial position (C1VC) and of the $F2_{offset}$ and $F2_{mid}$ when they are realized in final position (CVC1) for M and F speakers.*

Agwuele, A., Sussman, H.M., Lindblom, B., 2008. The effect of speaking rate on consonant vowel coarticulation. Phonitica 65, 194–209.

Bladon, R.A.W., Al-Bamerni, A., 1976. Coarticulation resistance in English /l/. J. Phon. 4, 137–150.

Krull, D., 1988. Acoustic properties as predictors of perceptual responses: A study of Swedish voiced stops.

Lacerda, F., 1985. Descriptive acoustic studies for the synthesis of spoken Swedish. Phon. Exp. Res. Inst. Linguist. Univ. Stockh.-PERILUS 4, 51–65.
Lindblom, B., 1990. Explaining phonetic variation : A sketch of H&H Theory, in: Hardcastle, W.J., Marchal, A. (Eds.), Speech Production and Speech Modelling. Kluwer Academic Publishers, pp. 403–439.
Lindblom, B., 1963. Spectrographic Study of Vowel Reduction. J. Acoust. Soc. Am. 35, 1773–1781. https://doi.org/10.1121/1.1918816
Recasens, D., 1985. Coarticulatory patterns and degrees of coarticulatory resistance in Catalan VCV sequences. Lang. Speech 28, 97–114.
Sussman, H.M., Bessell, N., Dalston, E., Majors, T., 1997. An investigation of stop place of articulation as a function of syllable position: A locus equation perspective. J. Acoust. Soc. Am. 101, 2826–2838.

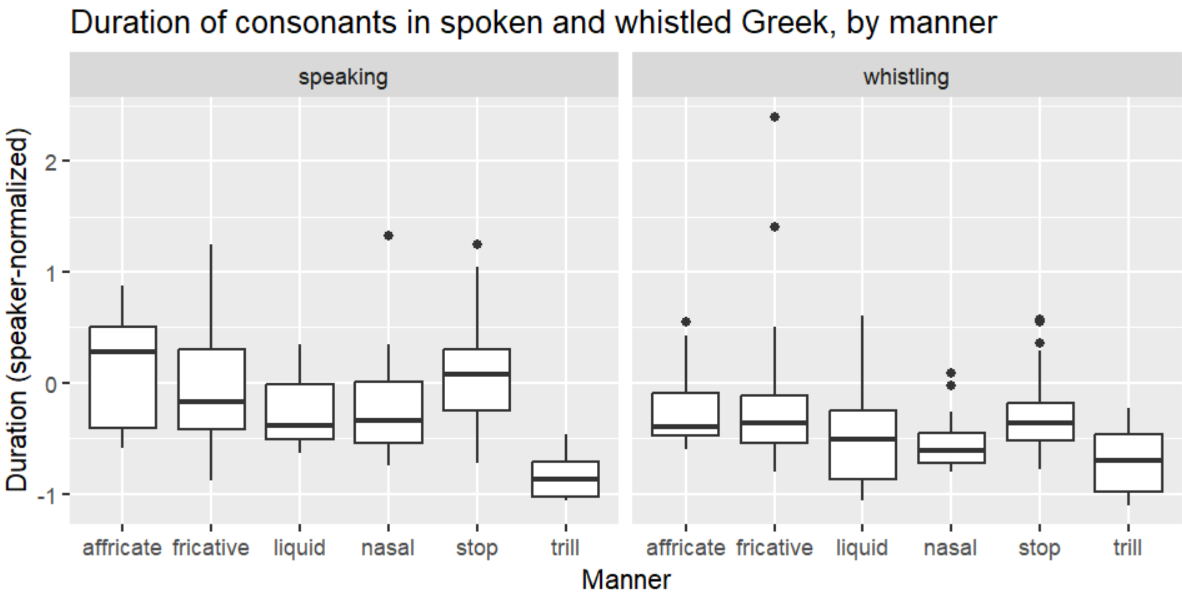
6. Consonantal distinctions in whistled Greek

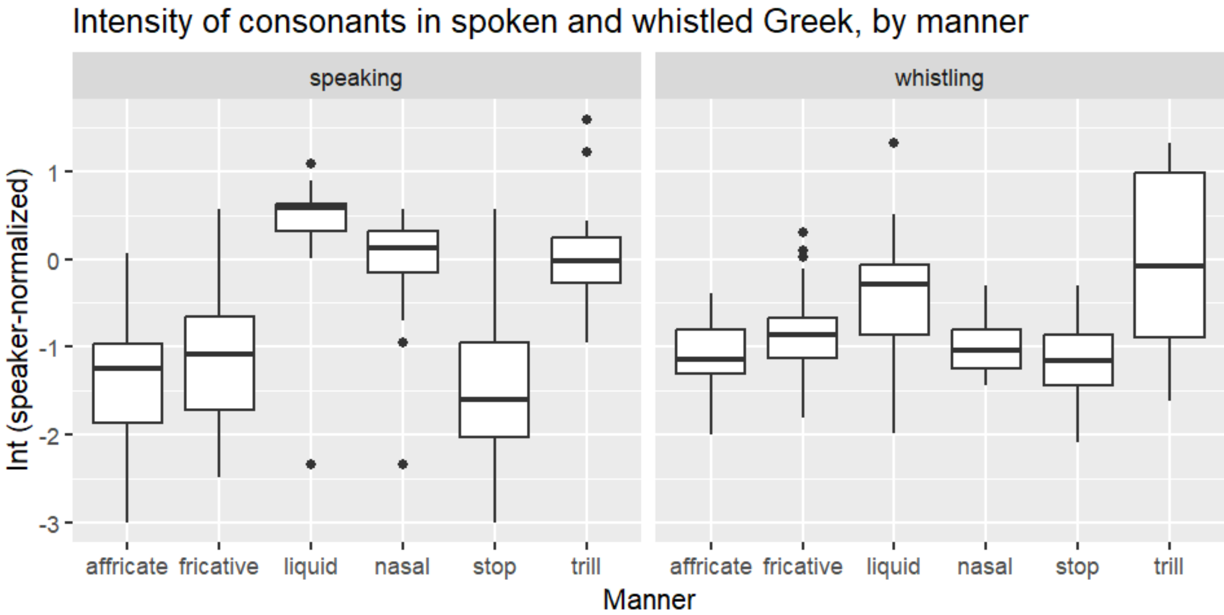
André Batchelder-Schwab, Vasileios Michos
Boston University

Introduction The village of Antias in Euboea maintains a whistled register of Greek named *Sfyria*. Linguistic work has documented some sociolinguistic [3,4] and phonological [4,7] properties of the whistled language. Vowels in Sfyria are stratified by pitch, with /i/ at the highest tone and /o/ at the lowest [4,7]. Consonants are modulated as pauses and pitch variations between vowels; consonants are collapsed into at least seven categories, though prior studies did not test the entire consonant inventory of Greek, specifically voiced stops [4]. This paper uses novel experimental data to expand prior phonological descriptions of whistled Greek [4].

Method The experiment uses the frame sentence *ipe* [X] *kathara* [1] with minimal pairs tested utterance-internally. Tokens were controlled to fit a $C_1V_1C_2\acute{V}_2$ template with final stress and /a/ vowel quality, with no nonce words. The study treated C_2 as the experimental variable, testing all obstruents /b d g p t k f θ s x v ð z ɣ ts dz/, nasals /m n/, and liquids /r l/ with 1-2 tokens per phoneme shown three times each through the duration of the slides (3-6 instances per obstruent per participant). Three whistlers (out of four current living whistlers of Greek, according to the community) were shown the sentences on randomized slides and instructed to read them in whistled modality, then the experiment was repeated in spoken modality [2]. Each target C_2 phoneme of the token word was coded and measured for pitch, intensity, and duration in spoken (50-300Hz) and whistled (800-4200Hz) modalities at the 40-60% section; each V_2 was coded and measured for pitch, intensity, duration, and features of the preceding C_2.

Results A mixed linear effects model in R shows that between these six measures, most contrasts are covered as significantly distinct ($p<.001$). We propose a reanalysis of the whistled consonant system in Greek to be (b p)(k g)(t)(d)(s z)(ts dz)(x)(ɣ)(v)(f)(θ)(ð)(r)(l)(m)(n). None of the six measures were able to distinguish the voicing contrast in labial and velar stops, or between the coronal fricatives and affricates, suggesting that whistled Greek might not keep these pairs of phonemes perceptually separate. Thus, the number of consonant contrasts diminishes from 20 in spoken Greek to 16 in whistled Greek, over double the prior reported number [4].





[1] Arvaniti, Amalia. (2000). *The phonetics of stress in Greek.* Journal of Greek Linguistics 1.

[2] Barnes, Jonathan, André Batchelder-Schwab & Okrah Oppong. 2023. Phrase level prosody of Akan-Twi in spoken and whistled modalities. *Phonetics and Phonology Europe 2023*.

[3] Kouneli, M., Meyer, J., & Nevins, A. (2013). *Whistled languages: Including Greek in the continuum of endangerment situations and revitalization strategies*. In M. Jones & S. Ogilvie (Authors), Keeping Languages Alive: Documentation, Pedagogy and Revitalization, 198-211.

[4] Meyer, J. (2015). Whistled languages. A Worldwide Inquiry on Human Whistled Speech; Springer: Heidelberg, Germany; New York, NY, USA.

[5] Rialland, A. 2005. Phonological and Phonetic Aspects of Whistled Languages. *Phonology* 22(2), 237-271.

[6] Trudgill, P. (2003). Modern Greek dialects: A preliminary classification. Journal of Greek Linguistics, 4(1), 45–63.

[7] Xiromeritis, N., & Spyridis, H. C. (1994). An Acoustical Approach to the Vowels of the Whistling Language of the Village Antias in the Greek Island of Evia. Acta Acustica united with Acustica, 80(5), 425-430.

7. Mechanisms of temporal prediction for speech segmentation: Evidence from nonword detection

Laurence White[1], Sarah Knight[1], Sven Mattys[2], Robert Lennon[1], Julie Morris[1]
[1]Newcastle University, [2]University of York

Spoken word boundaries are marked, in many languages, by lengthening of onset consonants [1, 2], with listeners using this lengthening to detect word boundaries in continuous speech [3]. Given the temporally linear flow of speech, the identification and interpretation of word-initial lengthening presumably entails predictive processing by listeners [4]. Thus, in order to perceive a word-onset consonant as relatively long requires the listener to have prior expectations, based on foregoing speech rate, about the duration of upcoming sounds. Indeed, foregoing speech rate has already been shown to affect listeners' phonetic judgements, e.g., about the presence of reduced syllables or location of lexical stress [5, 6].

We examined how foregoing utterance timing affects listeners' perception of localised durational cues to word boundaries in a series of studies using a novel nonword detection paradigm. Listeners heard diphone-synthesised twelve-syllable nonsense carrier utterances (e.g., *dumipako<u>libeku</u>binudafolu*), with trisyllabic nonword targets (e.g., *libeku*) embedded on half of the trials. **Utterance Position** of embedded targets was either early, medial or late in carrier utterances. In the baseline Flat timing condition, all segments were 120ms. In other timing conditions, all segments were 120ms except the specific duration-cueing segments. The segments were set to 170ms, to provide word boundary cues which were either congruent or incongruent with embedded target words. Crucially, lengthening of the target-word-initial onset consonant (*Syl1-C*) was expected to provide a strong cue to the preceding word boundary. The range of different **Timing** conditions is exemplified in the Figure 1 and Figure 2 captions.

In Studies 1 and 2, an auditory target probe (e.g., *libeku*) was presented after the utterance on each trial, with listeners responding to indicate whether they had heard the targets in the foregoing utterance. In all studies, we analysed target detection accuracy using generalised linear mixed effects models, with likelihood ratio tests to determine the predictive power of the experimental factors **Timing** and **Utterance Position**. As shown in Figure 1, Study 1 detection accuracy was at chance level (50%) for early-embedded targets, an apparent memory effect, with early acoustic traces disrupted by later syllables. Medial-embedded targets were detected above chance, but with no evidence of differential timing effects. In utterance-late position, however, listeners benefited from target-initial consonant lengthening ($p < .05$ *vs* other timing conditions). Study 2 attempted to replicate and extend this effect of utterance position on the interpretation of durational cues. We used a target-initial consonant lengthening cue as before, but also composite vowel+consonant lengthening (see Figure 2 caption) in some timing conditions, which we expected might further promote detection of the local word boundary. Study 2 detection accuracy was predicted by **Utterance Position**, $X^2(2) = 31.66$, $p < .001$, and **Timing**, $X^2(4) = 17.68$, $p = .001$: as in Study 1, early-embedded targets were detected at chance and medial-embedded targets were detected above chance, but with no accuracy differences due to timing (Figure 2). There was also an **Utterance Position** x **Timing** interaction, $X^2(8) = 17.87$, $p = .022$. This was attributable to target detection for the late-embedded *Syl1-V Syl2-C* condition being poorer than for all other late timing conditions ($p < .01$), apart from *Syl0-V*. Thus, once again, differential use of timing cues was only evident in later utterance positions.

Three further studies examined how durational cue use is mediated: a) memory effects, via pre-presented probes in Studies 3 and 4; b) listeners' development of lexical knowledge over trials, with no target repetitions in Study 5. Taken together, findings support the interpretation that listeners' expectations about speech sound duration are influenced by distal speech rate, with temporal predictions only possible given sufficient entrainment to foregoing utterances.

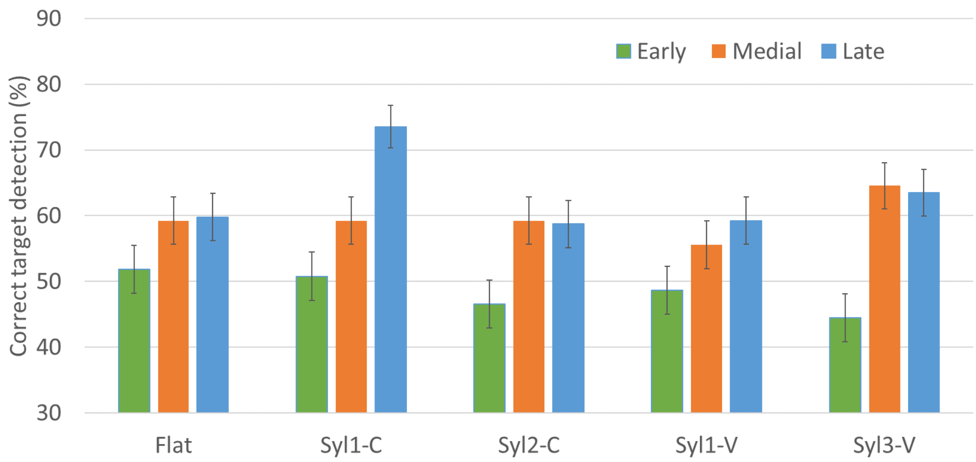


Figure 1. *Study 1 target detection accuracy.* ***Timing*** *conditions Flat: all segments 120ms; Syl1-C: first consonant 170ms (e.g., libeku); Syl2-C: second consonant 170ms (libeku); Syl1-V: first vowel 170ms (libeku); Syl3-V: third vowel 170ms (libeku).* ***Utterance Position*** *conditions Early vs Medial vs Late targets in carrier utterances. Bars one standard error.*

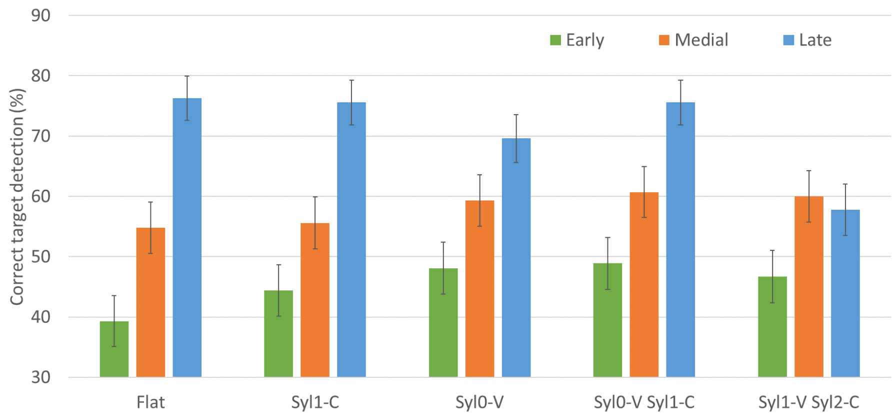


Figure 2. *Study 2 target detection accuracy.* ***Timing*** *conditions As Study 1, except Syl0-V: pre-target vowel 170ms (e.g., molibeku); Syl0-V Syl1-C: pre-target vowel & first consonant both 170ms (molibeku); Syl1-V Syl2-C: first vowel & second consonant both 170ms (libeku).* ***Utterance Position*** *conditions as above. Bars one standard error.*

[1] Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. 2004. Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology 6* (pp. 143-161). Cambridge University Press.

[2] Blum, F., Paschen, L., Forkel, R., Fuchs, S., & Seifart, F. 2024. Consonant lengthening marks the beginning of words across a diverse sample of languages. *Nature Human Behaviour*, *8*, 2127-2138.

[3] White, L., Benavides-Varela, S., & Mády, K. 2020. Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? *Journal of Phonetics*, *81*, 100982.

[4] White, L., & Malisz, Z. 2020. Speech rhythm and timing. In C. Gussenhoven & A. Chen (Eds.), *Oxford Handbook of Prosody* (pp. 167-182). Oxford University Press.

[5] Dilley, L.C., & Pitt, M.A. 2010. Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*, 1664-1670.

[6] Reinisch, E., Jesse, A., & McQueen, J.M. 2011. Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech*, *54*, 147–165.

8. Toward a unified articulatory account of speech and sign transitions

Reed Blaylock, Zed Sevcikova Sehyr
Chapman University

Introduction. Articulatory transitions, in both speech and sign, carry important perceptual information. Formant transitions in speech arise from temporally overlapping articulatory movements and provide cues about upcoming sounds [1]. Similarly, in signed languages, the movement of a signer's hands toward a sign's lexically-specified place of articulation may offer anticipatory information about the upcoming sign [2, 3]. However, it remains unclear whether these transitional movements in speech and sign follow similar kinematic principles and whether they can be accounted for within a unified articulatory framework. In this study, we investigate kinematic properties of transitional movements in isolated signs in American Sign Language (ASL) to examine whether these properties are consistent with previous observation of spoken transitions.

Method. We examined the transitional movement of the wrist from its initial rest position to its lexically-specified target location, measuring the *time to peak velocity*—the amount of time it takes for the wrist to reach its fastest speed relative to the whole duration of the transitional movement. Using 1,034 one-handed signs from ASL-LEX 2.0 [4], we applied pose estimation with MediaPipe [5] to track the dominant hand wrist movements. Transition onsets and offset were automatically detected based on wrist velocity. The transition phase was identified around the first major velocity peak (PVEL); transition onset was found where velocity was 10% of the difference between PVEL and the nearest preceding velocity minimum, and transition offset was found where velocity was 10% of the difference between PVEL and the nearest following velocity minimum. We calculated the *time to peak velocity* as the proportion of the time taken to reach maximum speed relative to the full duration of the transition phase

Results. Our results reveal that wrist movements in sign transitions follow a velocity trajectory similar to previously reported trajectories of spoken articulations [6]: wrist velocity peaks roughly halfway through the movement (45.8% of total duration) before decelerating into the target position. This trajectory resembles the behavior of a critically damped mass-spring system, a common model used to describe movement of the speech articulators (red line in Figure 1). The automatically detected transition offsets and aligned well with corresponding manually coded sign onsets in ASL-LEX 2.0.

Discussion. Our results support the idea that a unified articulatory account of transitions in speech and sign may be fruitful. Articulatory frameworks such as Articulatory Phonology [7], which conceptualize phonological units as dynamical action units (often modelled with mass-spring systems), could be extended to account for transitional movements in both modalities. In such an approach, transitions between signs would be considered phonological actions and part of the sign [2], carrying structured phonologically relevant information. This reconceptualization of sign transitions would help explain why transition information can facilitate lexical access [3], much like coarticulatory cues in speech. Future research should further explore the informational content of sign transitions, their role in sign perception, and how they contribute to the organization of signed phonology, especially within connected discourse.

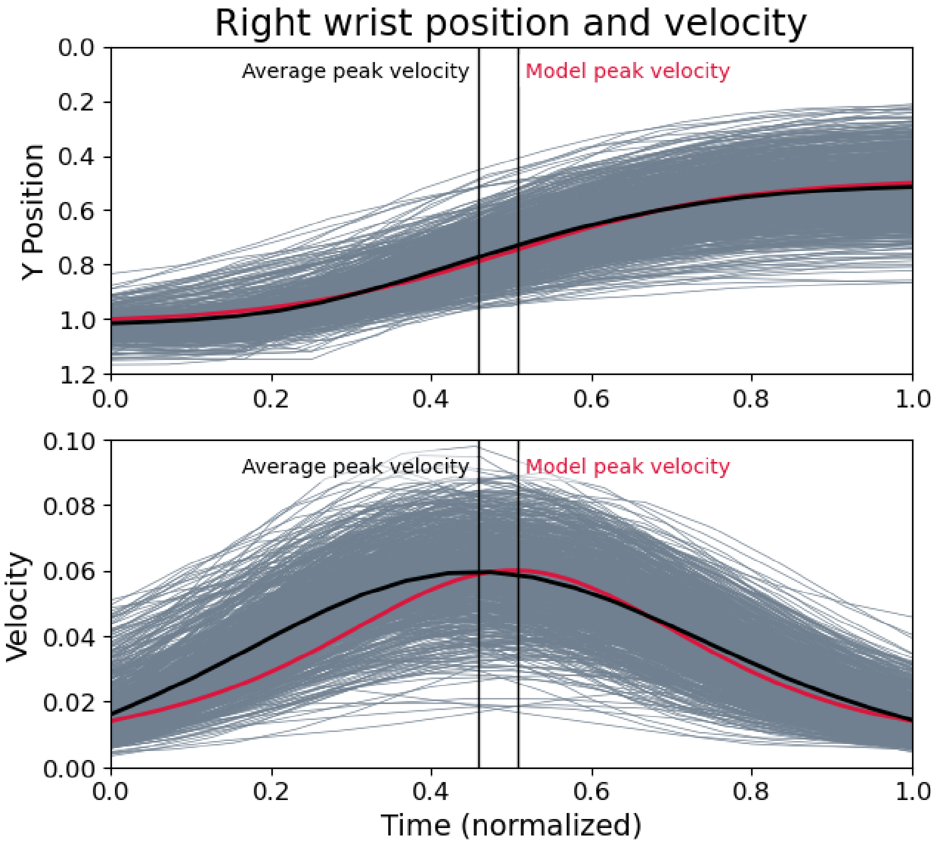


Figure 1. *Gray: Time-normalized position and velocity of the dominant wrist in 825 one-hand signs. Black: Average position trajectory and velocity profile of the 825 signs; peak velocity occurs just before 50%. Red: A vocal articulation modelled as a critically-damped mass-spring system with a soft spring; peak velocity occurs slightly after 50%.*

[1] Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs: General and Applied, 68(8), 1.

[2] Jantunen, T. (2015). How long is the sign? Linguistics, 53(1).

[3] Brozdowski, C., & Emmorey, K. (2023). Using transitional information in sign and gesture perception. Acta Psychologica, 236, 103923.

[4] Sevcikova Sehyr, Z., Caselli, N., Cohen-Goldberg, A. M., & Emmorey, K. (2021). The ASL-LEX 2.0 Project: A database of lexical and phonological properties for 2,723 signs in American Sign Language. Journal of Deaf Studies and Deaf Education.

[5] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019, June). Mediapipe: A framework for perceiving and processing reality. In Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR) (Vol. 2019).

[6] Sorensen, T., & Gafos, A. (2016). The Gesture as an Autonomous Nonlinear Dynamical System. Ecological Psychology, 28(4), 188–215.

[7] Browman, C. P., & Goldstein, L. M. (1986). Towards an Articulatory Phonology. Phonology Yearbook, 3, 219–252.

[8] Jantunen, T. (2013). Signs and Transitions: Do They Differ Phonetically and Does It Matter? Sign Language Studies, 13(2), 211–237.

9. Cue weighting in Mandarin sibilants perception and imitation by native Mandarin speakers and naïve English speakers: a pilot study

Kuilin Li, Meghan Clayards, Morgan Sonderegger
McGill University

How do naïve individuals imitate the new sounds of a new language when it has more categories than their native language? This study takes English and Mandarin sibilants as an example. English has a two-way sibilant contrast between alveolar /s/ and postalveolar /ʃ/, while Mandarin has a three-way contrast between alveolar /s/, retroflex /ʂ/ and alveolo-palatal /ɕ/ (Ladefoged & Wu, 1984). Native Mandarin speakers use two cues to distinguish /ɕ/ from other sibilants (Hauser, 2023; Li, 2006) – the second formant (F2) transition of the subsequent vowel (most important) and the spectral properties of fricative noise, e.g., centre of gravity (COG). In contrast, native English speakers primarily use fricative cues to distinguish between /s/ and /ʃ/ in English, while vowel transitions are also differentiated (Delattre, 1962) but have lesser effect on the perception of the distinction (Whalen, 1981, 1991). Thus, producing Mandarin sibilants contrasts may present a challenge for naïve English listeners.

The perception of Mandarin sibilants has been surprisingly understudied. Polish has a similar three-way sibilant contrast (Jassem, 2003) and Polish listeners have been found to use formant transitions as a key cue (Nowak, 2006). Moreover, naïve English listeners also rely solely on vowels and ignore frication difference when perceiving Polish /ɕ/-/ʂ/ (McGuire, 2007). However, while the COG values of Polish /ʂ/ and /ɕ/ are almost the same, standard Mandarin maintains a COG distinction between /ɕ/ and other sibilants (Hauser, 2023; Chiu, 2009). Thus, we cannot infer Mandarin perceptual cue weights based on Polish cue weights.

This study tests production and perception of Mandarin sibilants in native (Beijing) Mandarin and naïve English participants to answer the following questions: (1) How do native Mandarin listeners/speakers use cues when perceiving and producing Mandarin sibilants? (2) How do naïve English listeners/speakers use cues when perceiving and imitating Mandarin sibilants? If they use their English cue weighting strategy, we expect them not to use the vowel cue or treat it as a secondary cue. (3) Are both groups' cue weighing strategies at the perception and production correlated? If they use the consistent strategy in perception and production, then they are correlated. Otherwise, there is no link between them.

10 Mandarin participants completed a two-alternative forced-choice identification task and a production task. They categorized two 7-step fricative*7-step vowel continua (/sa/-/ɕa/, /ʂa/-/ɕa/), and then read /sa/, /ɕa/ and /ʂa/ each for 5 times. 10 English participants performed the same identification task and then imitated natural productions of /sa/, /ʂa/ and /ɕa/ tokens (1 token each) produced by a female native speaker, 20 times.

The results of the identification task (Fig 1) showed that Mandarin listeners use both cues, favouring vowel cues in the /sa/-/ɕa/ continuum, but weighting both cues almost equally in the /ʂa/-/ɕa/continuum. English speakers relied solely on vowel cues in /sa/-/ɕa/ and on fricative cues in /ʂa/-/ɕa/ (quite different from Polish). Overall, the English group struggled with /sa/-/ɕa/ distinction. The production results (Fig 2, right) showed that Mandarin speakers produced distinctive fricatives and vowels among the three sibilants, consistent with Hauser (2023) and Li (2006). English speakers produced clearly distinguishable /sa/ and /ʂa/. As for /ɕa/, almost all English speakers produced formant transitions distinct from /sa, ʂa/, but their imitation of the frication had large individual differences (Fig 2, left).

We built Bayesian logistic regression models for each task and group and calculated their cue weightings (not shown here), which are consistent with Fig 1 and Fig 2. To measure individual correlations between the perception and production tasks, we built a multivariate Bayesian model for both groups, which showed that there was no correlation between the tasks, but that the differences in participants affected the model significantly. Therefore, the lack of correlation might be due to too few participants and too much individual variation.

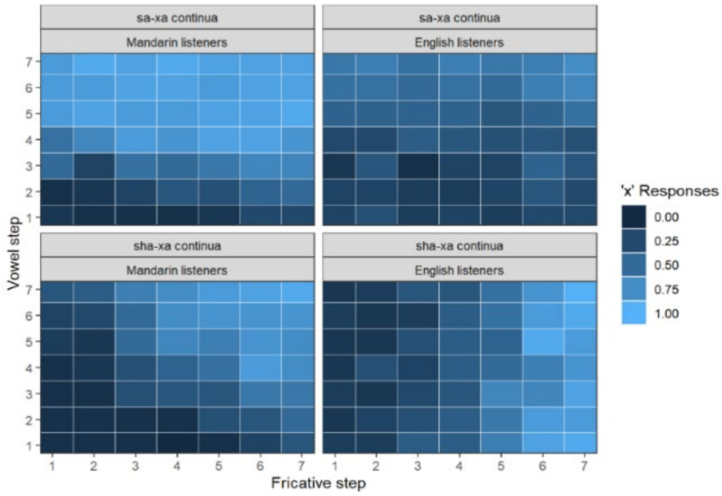


Figure 1. *Percentage of /ɕ/ responses as a function of the vowel step and fricative step, faceted by group and continua.*

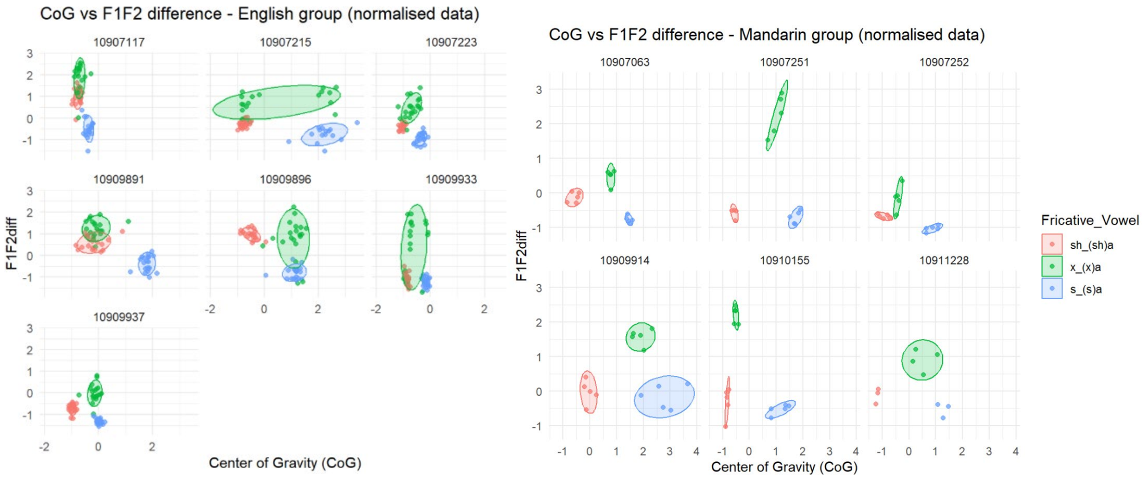


Figure 2. *Production (Mandarin group) and imitation (English group) results across each participant (after data clean).*

[1] Chiu, C. (2009). Acoustic and auditory comparisons of Polish and Taiwanese Mandarin sibilants. *Canadian Acoustics*, *37*(3), 142-143.

[2] Delattre, P. C., Berman, A. L., & Cooper, F. S. (1962). Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia linguistica*, *16*(1-2), 104-122.

[3] Hauser, I. (2023). Differential cue weighting in Mandarin sibilant production. *Language and Speech*, *66*(4), 1056-1090.

[4] Jassem, W. (2003). Polish. Journal of the international Phonetic Association, 33(1), 103-107.

[5] Ladefoged, P., & Wu, Z. (1984). Places of articulation: an investigation of Pekingese fricatives and affricates. *Journal of Phonetics*, *12*(3), 267-278.

[6] Li, F. (2008). The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese (Doctoral dissertation, The Ohio State University).

[7] Nowak, P. M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of phonetics*, *34*(2), 139-152.

[8] Whalen, D. H. (1991). Perception of the English/s/–/ʃ/distinction relies on fricative noises and transitions, not on brief spectral slices. *The Journal of the Acoustical Society of America*, *90*(4), 1776-1785.
[9] Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]–[š] boundary. *The Journal of the Acoustical Society of America*, *69*(1), 275-282.

10. More than just the sum of its parts: prominence perception in increasing context

Chigusa Kurumada[1], Janne Lorenzen[2], Stefan Baumann[2]
[1]*University of Rochester,* [2]*University of Cologne*

In conversation, subtle acoustic and prosodic variations effectively signal the informational status of referents (e.g., given vs. new) [1-2]. Typically, what is new is encoded with increased perceptual prominence, e.g., acute pitch movements and longer duration, while what is given is attenuated. Previous work has also identified as many as 17 signal-based and lexico-syntactic cues that impact the perception of relative prominence and reference resolution [3]. However,

a key question remains: *How do listeners process these cues within their surrounding context?* Prosodic features of speech are known to be perceived relative to their context [4]. For example, the prominence of the noun "parade" in "I saw a parade yesterday" likely depends on the pronunciation and semantics of its preceding context—both proximal ("a") and distal. Furthermore, its nuclear accent status depends on its subsequent context ("yesterday").

To answer this question, we elicited naive listeners' prominence judgments on a noun situated in an increasing amount of context. Do their judgments change and become clearer as more contextual cues are provided? Alternatively, are some cues more important than others, or are all cues necessary for listeners to make reliable prominence judgments?

We used naturalistic recordings of three female native speakers of German, taken from a corpus of read narratives [5]. Each speaker produced a series of sentences containing a noun once as new and once as given (**Example 1**). We spliced out the target noun with varying sizes of surrounding context to create four conditions: (1) noun only [N only], (2) noun + following verb [N+V], (3) article + noun + verb [A+N+V], and (4) an entire sentence [Sentence] (**Figure 1**). The given and new items within each speaker were equivalent in their grammatical role and pitch accent type and produced with a prenuclear and a nuclear accent, respectively.

30 native speakers of German listened to pairs of a given and a new item in sequence and selected the one that was more prominent than the other (**Figure 2**). The order of two items was counterbalanced, and each order was repeated four times per condition (8 trials * 3 speakers * 4 conditions = 96 responses / listener). The four conditions were blocked and presented in a fixed order, i.e., each listener heard the target nouns in an increasingly larger context, from the noun alone to a whole sentence. A total of 2880 responses were analyzed using a generalized linear mixed effects regression in R, with the fixed effects of condition (sliding-difference coded) and speaker (contrast-coded) and an interaction term between the two. A final model included random by-subject intercepts and slopes for all fixed effects.

Results indicated that native listeners consistently selected the new item as more prominent than the given item, even in the absence of context (**Figure 3**). That is, naturalistic productions of nouns seem to carry sufficient information about their relative prominence and differential information statuses of the referent. Important to our current question, the effect of condition was significant only at the level of [Sentence] ($\hat{\beta} = 2.49,\ p < 0.02$). All the other levels (= [N only], [N+V], and [A+N+V]) were not significantly different from each other. This was surprising as the stimuli in the [A+N+V] and the [Sentence] conditions were identical in their proximal context. *Only when listeners heard the target noun in a prosodically and semantically meaningful sentence did the contextual cues to prominence have a facilitating effect*.

We thus provide novel evidence that the contextual influence on linguistic judgments of prominence is not just the sum of its decomposable parts. Listeners appear to draw on their semantic and structural knowledge in addition to bottom-up perceptual cues to make holistic judgments [6]. In our full presentation, we will further discuss: (a) acoustic/prosodic analyses of the nouns and their context, (b) individual differences between the three speakers' productions, and (c) how more subtle differences in information status (such as the givenness of a lexical form vs. the givenness of a referential identity) may be perceived in context.

(1) Example narrative for the target word “Parade”

Um seinen Abschluss zu feiern, hat Julian einen Städtetrip unternommen. Er ist mit dem Zug nach Paris gefahren, hat im Hotel eingecheckt und wollte dann die Stadt erkunden. An dem Tag hat Julian eine **Parade** [new] eingeplant. Er wusste, dass ein Feiertag war und ist sofort in die Innenstadt gefahren. Dort hat Julian die **Parade** [given] angeschaut.

To celebrate his graduation, Julian went on a city trip. He took the train to Paris, checked into his hotel, and then planned to explore the city. That day, Julian planned to go on **a parade**. He knew it was a holiday, so he went straight to the city center. There, Julian watched **the parade**.

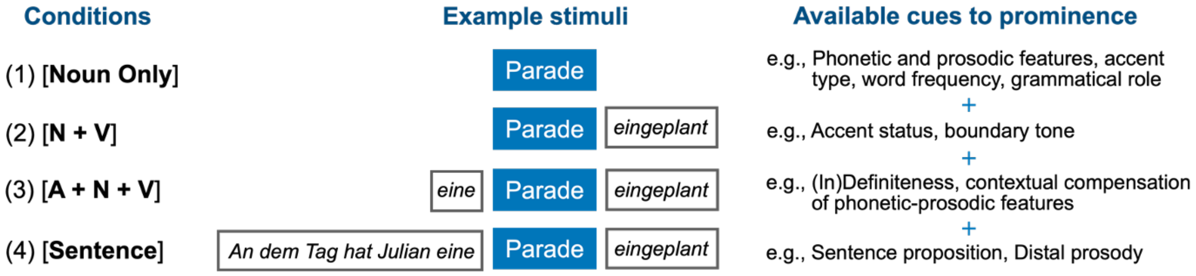


Figure 1. *Four conditions were created with increasing sizes of context of the target noun.*

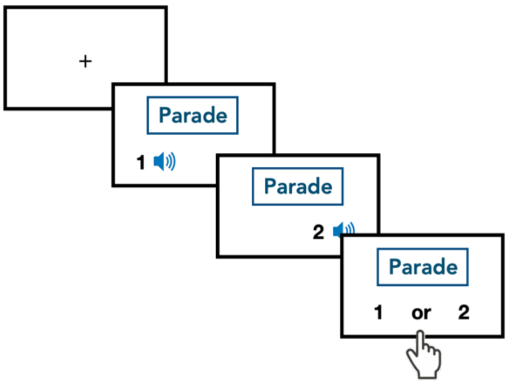


Figure 2. *Listeners heard a given and a new item in sequence and selected one that was more prominent than the other.*

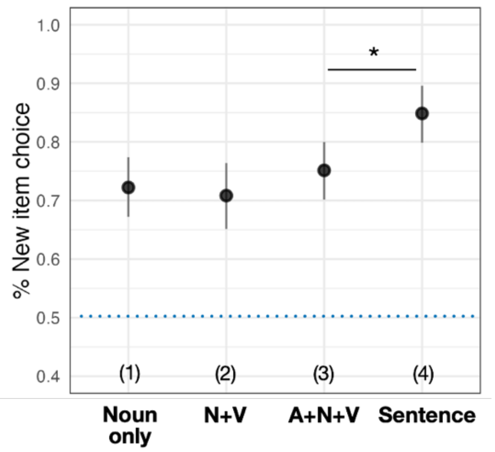


Figure 3. *Percentages of “new” item choice in the four conditions. Dotted line indicates 50% chance performance.*

[1] Aylett, M. & Turk, A. 2004. The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang Speech: 47*(1), 31-56.

[2] Rietveld, A.C.M., & Gussenhoven, C. 1985. On the relation between pitch excursion size and prominence. *J. Phon, 13*(3), 299-308.

[3] Baumann, S. & Winter, B. 2018. What makes a word prominent? Predicting untrained German listeners’ perceptual judgments. *J. Phon., 70,* 20-38.

[4] Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In Elordieta & Prieto (Eds.) *Prosody and Meaning*. Mouton de Gruyter, 239-269.

[5] Baumann, S., & Lorenzen, J. 2024. Boosting or inhibiting: how semantic-pragmatic and syntactic cues affect prosodic prominence relations in German. *PLOS One, 19*(4).

[6] Cole, J., Hualde, J.I., Smith, C.L., Eager, C., Mahrt, T., de Souza, R.N. 2019. Sound, structure and meaning: the bases of prominence ratings in English, French and Spanish. *J. Phon. 75*, 113–147.

11. Pronoun accentuation produces interference effects in memory for referential alternatives

Sophie Repp, Heiko Seeliger, Sven Weber
University of Cologne

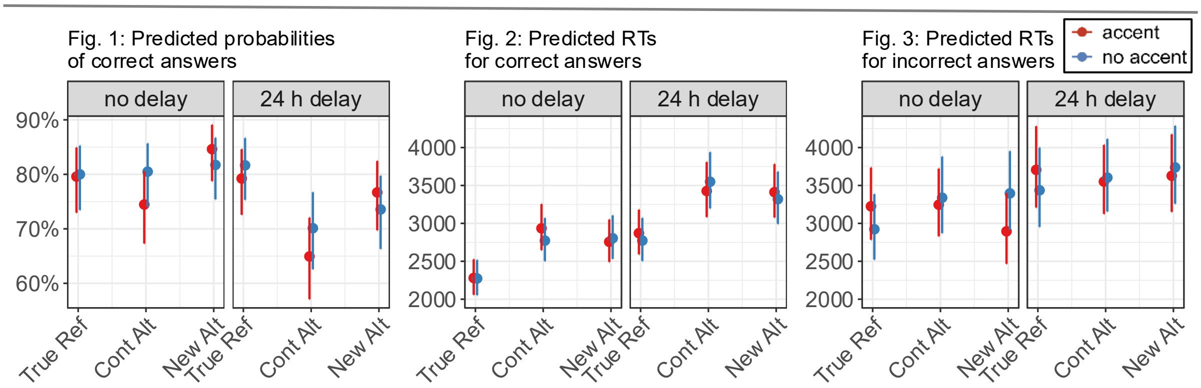
Prosodic prominence influences the representation of referents in discourse memory: High prosodic prominence of a noun may improve recall of the denoted referent and of information associated with it. This effect has been shown for prominence indicating focus 0–0 and contrastive topichood 00. Both information-structural settings involve (contrastive) alternatives. Better recall for focus is thought to stem from a richer semantic representation, due to a more specific memory trace for the focussed referent, and/or the additional representation of the alternatives. Better recall for contrastive topics has been linked to a structured memory representation with topics as anchors 0. Prominence-induced memory effects have not only been shown for the focus/topic itself but also for the alternatives 0–0. The memory representation can be long-lasting, e.g. a day 0. ♦ Memory-enhancing effects of prosodic prominence so far have only been tested for nouns. If similar effects arise with pronouns is an open issue. When a pronoun is uttered, the pronominal referent is accessed in memory. Accenting a pronoun may trigger a contrastive interpretation, i.e. indicate the presence of alternatives 0–0. So we may expect that pronoun accentuation also enhances memory for the denoted referent. ♦ We present data from a recall study (German). 96 participants listened to stories containing a sentence with a subject *d*-pronoun (1); *d*-pronouns are used in colloquial speech 0 and are accented more naturally than personal pronouns. Each story introduced a set of referential alternatives (e.g., animal breeders), then named two of the referents (rabbit breeder, horse breeder), followed by a third referent. The *d*-pronoun in the ensuing target sentence referred to the third referent and was either unaccented or carried a L+H* accent. The story ended with an evaluative statement. Participants listened to 30 stories (Latin sq.) mixed with 30 fillers. In the recall task, which followed directly after the presentation of all stories or with a 24h delay (between-participants) they verified written statements (Latin sq.): the target sentence with (2i) the original subject referent ($T_{TrueRef}$), (2ii) one of the contextual alternatives as subject ($T_{ContAlt}$) or (2iii) a plausible unmentioned alternative (T_{NewAlt}). ♦ Our (G)LMM analysis showed that ***recall accuracy*** (Fig.1) was higher for $T_{TrueRef}$ than for $T_{ContAlt}$ /T_{NewAlt}, and lower for $T_{ContAlt}$ than T_{NewAlt}. This suggest that the proposition expressed by $T_{TrueRef}$ is remembered best, and that *contextual referential alternatives* are also stored in memory turning them into lures for an illusory recall of $T_{ContAlt}$ (esp. after 24h); unmentioned referents are safer rejects. For $T_{ContAlt}$, accuracy was lower when the *d*-pronoun was accented, indicating that accentuation raises the discourse prominence of the contextual alternatives, raising their lure quality. Accuracy dropped after 24h except for $T_{TrueRef}$, indicating a long-term advantage for true propositions. ***Reaction times (RT)*** *for correct answers* were faster for $T_{TrueRef}$ than $T_{ContAlt}$/T_{NewAlt}; the 24h delay increased RTs (Fig. 2). The former effect confirms enhanced memory for true propositions, the latter confirms memory decay with time. *RTs for incorrect answers* to $T_{TrueRef}$ were slower when the *d*-pronoun was accented; for $T_{ContAlt}$/ T_{NewAlt} they were faster with accentuation (Fig. 3). We propose that the accent makes the mentioned alternatives more prominent *and* activates likely alternatives. This slows down the incorrect rejection of $T_{TrueRef}$, and speeds up the incorrect acceptance of T_{Alt}: participants are tricked into this decision faster than without an accent. The effect vanishes after 24h. ♦ Overall, the prosodic prominence of a pronoun has similar memory effects for the true proposition as nominals do. For contextual alternatives, pronouns show the opposite effect. Note that pronouns must be interpreted with reference to the previous discourse. So an accented pronoun induces the *re*-activation both of the pronominal referent and the alternatives. Arguably, this leads to greater competition, with the alternatives producing interference effects.

(1) Sample item listening phase (left and right context: English translation only)
The results from the animal breeding contest in Altdorf have just come in. A few of our people were there, remember. For example, our rabbit breeder and our horse breeder took part. Our pigeon breeder had the most beautiful animals.

Am Ende hat {der/DER$_{L+H*}$} echt viele Preise gewonnen.
at end has he really many prizes won
'At the end he really won lots of prizes.'
That was quite impressive.

(2) Sample item recall task

i. **TrueRef** *Der Taubenzüchter hat echt viele Preise gewonnen.*
ii. **ContAlt** *Der Kaninchenzüchter/Pferdezüchter hat echt viele Preise gewonnen.*
iii. **NewAlt** *Der Hühnerzüchter hat echt viele Preise gewonnen.*
the {pigeon//rabbit/horse//chicken} breeder has really many prizes won



[1] Sanford, A.J.S., Sanford, A.J., Molle, J., & Emmott, C. 2006. Shallow processing and attention capture in written and spoken discourse. *Discourse Processes* 42(2), 109–130.
[2] Fraundorf, S.H., Watson, D.G., & Benjamin, A.S. 2010. Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *JMemLang* 63(3), 367-386.
[3] Gotzner, N., & Spalek, K. 2019. The life and times of focus alternatives: Tracing the activation of alternatives to a focused constituent in language comprehension. *Language and Linguistics Compass* 13(2). e12310.
[4] Yan, M., Calhoun, S., & Warren, P. (2023). The role of prominence in activating focused words and their alternatives in Mandarin: Evidence from lexical priming and recognition memory. *Language and Speech*, 66(3), 678-705.
[5] Kember, H., Choi, J. Yu, J. & Cutler, A.. 2021. The processing of linguistic prominence. *Language and Speech* 64(2), 413–436.
[6] Repp, S., & Drenhaus, H. 2015. Intonation influences processing and recall of left-dislocation sentences by indicating topic vs. focus status of dislocated referent. *Language, Cognition and Neuroscience 30*, 324-346.
[7] Venditti, J., Stone, M., Nanda, P., & Tepper, P. 2002. Discourse constraints on the interpretation of nuclear-accented pronouns. *Speech Prosody* 2002, 675–678.
[8] Brown-Schmidt, S., Byron, D.K., & Tanenhaus, M.K. 2005. Beyond salience: Interpretation of personal and demonstrative pronouns. *JMemLang 53*(2), 292–313.
[9] Kaiser, E. 2011. Salience and contrast effects in reference resolution: The interpretation of Dutch pronouns and demonstratives. *Language and Cognitive Processes 26*(10), 1587–1624.
[10] Hinterwimmer, S. 2019. Prominent protagonists. *Journal of Pragmatics 154*, 79-91.

12. Mind the gap: studying multifaceted turn-taking cues in L2 interviews

Aviad Albert[1], Francesco Cangemi[2], Hae-Sung Jeon[3], Simon Wehrle[1], Jane Setter[4], Martine Grice[1]
[1]University of Cologne, [2]The University of Tokyo, [3]University of Central Lancashire, [4]University of Reading

The study of turn-taking in social interaction has been gaining attention and has accumulated many fascinating findings in recent decades [1–3]. This paper examines turn-taking in a specific genre (interview), treating the turn as multifaceted and multimodal, being composed of diverse cues in both the auditory and visual domains.

Turn-taking conventions differ between genres. In an interview setting, the turns are built into the structure of the communicative procedure, with question–answer pairs highly prevalent throughout a conversation [4]. We focus here on interviews with second-language learners. Two types of expectation emerge in such settings, leading us to propose a distinction between: (a) *informational* expectation (give viable and satisfactory answers to questions); and (b) *interactional* expectation (collaborate in the timely co-construction of the exchange between questions and responses). We hypothesize that if the informational expectation is met (i.e. a viable response is ready within the optimal time window), the interactional expectations are also likely to be met. However, if the interlocutor struggles to provide a rapid viable response (thereby not meeting informational expectations), we expect interactional cues to play an increasingly important role. As points of reference, we assume that an optimal exchange gap is around 200 ms [2] and that 700 ms marks the upper limit for acceptable gaps [5].

In *Praat* and *ELAN*, we annotated a subset of ICNALE [6], a 3.5 m-word, publicly available corpus of L2 English that features video recordings of interviews between a single male native Japanese speaker and 100 Japanese learners. Our annotation scheme provides for a closer look into the response mechanisms at play. We include in our annotation body language (*visual*; head, face and hand gestures), as well as auditory cues, which may be non-speech sounds (*vegetative*; e.g. lip-smacks, in-breaths [7]), non-lexical sounds (*vocal*; e.g. filled pauses such as “uhm”), or non-viable responses (*verbal*; i.e. speech material not directly addressing the issue at hand, with generic expressions such as "well, I think...”), leading up to *viable* responses that directly address the question. We focus on the 24 interviews featuring the least proficient (n = 11) and the most advanced learners (n = 13) in the corpus (CEFR: A2 and B2, TOEIC: 200-400 and 800-900, respectively). The present study analyzed two questions that were asked in all interviews, one polar (yes/no) question and one open (WH) question (see Figure 1). We measured the time lag between the end of the question and the annotated onsets of the learners’ responses.

As might be expected, the time-to-viable-response is longer with the open question: a median gap of 6007 ms in the open question, as compared to 512 ms in the polar question. Responses were rarely initiated with visual cues, although they did sometimes occur before the end of the question. Vegetative cues initiated responses more often, with a tendency towards the optimal 200 ms gap (202 ms median gap in open questions). Vocal cues were closer to the 700 ms upper bound (924 ms median gap in open questions) and, lastly, verbal cues appeared with a median gap of 2028 ms in open questions. Individual results are presented in Figure 1. We also found an effect of proficiency, as the beginner learners took almost twice as long (median 9383 ms) as the advanced learners (median 4855 ms) to produce a viable response to the open question (cf. [8]), with no clear group difference for the polar question.

This study provides preliminary evidence for a sequential handling of communicative expectations that prioritizes interactional expectations (signaling willingness to respond), in the absence of a viable answer within the relevant time window. It is shown that multifaceted cues, including vegetative sounds, which have been ignored in most related work (e.g. [3]), are employed by speakers to fill the silence until a viable answer can be produced.

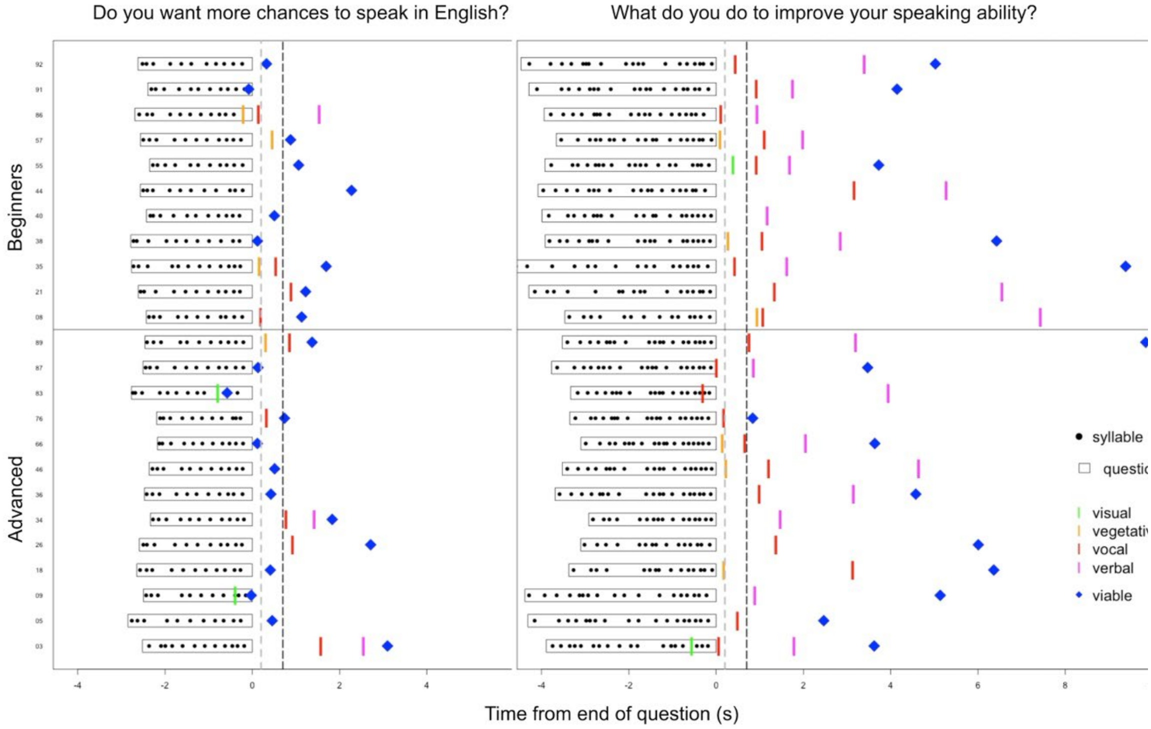


Figure 1. *Response times after a polar (yes/no) question (left) and an open (WH) question (right). Each row represents one learner. The x-axis represents time in seconds, with t=0 corresponding to the endpoint of the questions asked by the interviewer. Color-coded points in time represent the onset of the different types of cues that learners produced in response to questions (see list of cues below). A grey dashed vertical line marks 200 ms (putative optimal gap) and a black dashed vertical line marks 700 ms (putative upper gap bound) in both plots.*

List of cues:

- Visual (green): gestures, e.g. turn head, raise shoulders, open mouth
- Vegetative (orange): sounds, e.g. cough, lip-smack, swallow, in-breath
- Vocal (red): filled pauses and hesitations, e.g. "uhm", "er", "mmmh"
- Verbal (magenta): material not directly addressing the question, e.g. "well I think"
- Viable (blue): material directly addressing the question, e.g. "yes", "I try to sing English songs"

[1] Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.

[2] Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, 731. https://doi.org/10.3389/fpsyg.2015.00731

[3] Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: A corpus study. *Frontiers in Psychology*, 6, 509. https://doi.org/10.3389/fpsyg.2015.00509

[4] Wiens, A. N. (2017). *The Interview: Research on Its Anatomy and Structure*. Routledge. https://doi.org/10.4324/9781315132723

[5] Roberts, F., & Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America*, 133(6), EL471–EL477. https://doi.org/10.1121/1.4802900
[6] Ishikawa, S. (2023). *The ICNALE Guide.* Routledge.
[7] Torreira F, Bögels S, Levinson S. C. (2015) Breathing for answering: the time course of response planning in conversation. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.00284
[8] Wehrle, S., & Sbranna, S. (2023). Longer silence duration in L2 conversations is modulated by proficiency: Evidence from turn-timing in German and Italian. *Workshop on L2 Fluency, Speech Prosody 2024*, Leiden (https://bit.ly/48iA7Dl)

13. How meaning affects the duration of Japanese homophonous words

Motoki Saito[1], Ruben van de Vijver[2]
[1]Eberhard-Karls-Universität Tübingen, [2]Heinrich-Heine-Universität Düsseldorf

A central hypothesis of linguistic theory is that the relation between sound and meaning is arbitrary [1]. Research has brought to light an enormous amount of evidence that there are many systematic relations between the phonetic realization of a word and its meaning. Differences in phonetic details of words correlate with differences in parts-of-speech [2], morphological differences [3,4], and the meaning of words [5,6].

English homophones show systematic phonetic differences [7]. However, English has no phonemic length distinctions. This raises the question whether we would also see such effects in homophones in languages with a phonemic length distinction. One such language is Japanese, which additionally has a wealth of homophones.

To investigate systematically different phonetic realizations as a function of semantics in Japanese, we created a data set of 310,574 word tokens from a corpus of spontaneous Japanese of about 500,000 words, annotated for Parts-of-Speech, meaning, and transcribed in Japanese orthography. This data set included 653 homophonous noun types, and 38,075 homophonous noun tokens. In addition, we assessed whether these were present in a pre-trained fastText [8] word embeddings, and used only those words for which we have an embedding.

We used the words from the data set and their meanings as represented by embeddings as a basis for our analysis. This was done in the framework of the Discriminative Lexicon Model (DLM) (Baayen et al., 2019). The DLM contains three ingredients: word forms (in our case the words from the data set), word meanings (in our case their embeddings) and associations between them. The associations are estimated by means of multivariate, multiple linear regression [9]. These estimates of the associations make it possible to predict which word form is predicted by which meaning (as a model of word production), and which meaning is predicted by which word form (as a model of comprehension).

For the homophones in our data set we calculated whether the associations from meaning to word form were correlated with the duration of the homophone. We called this measure the unconditional semantic support of a word form. Since longer words have more segments and therefore receive more semantic support in total if semantic support for each segment was summed up, the amount of unconditional semantic support for each word was normalized by dividing the total amount of semnatic support by the number of segments in the word. An analysis with Generalized Additive Mixed-effects Models [10] showed that there was a positive correlation between the semantic support of a word and its length (see Figure 1). This effect was independent of an effect of word frequency. The better a word form is supported by a word meaning, the longer the word is.

We conclude that even in Japanese, which has phonemic length, phonetic detail can be used to distinguish homophones. This adds to the growing number of findings that attest to a direct link between form and meaning, further eroding the idea that their relation is arbitrary.

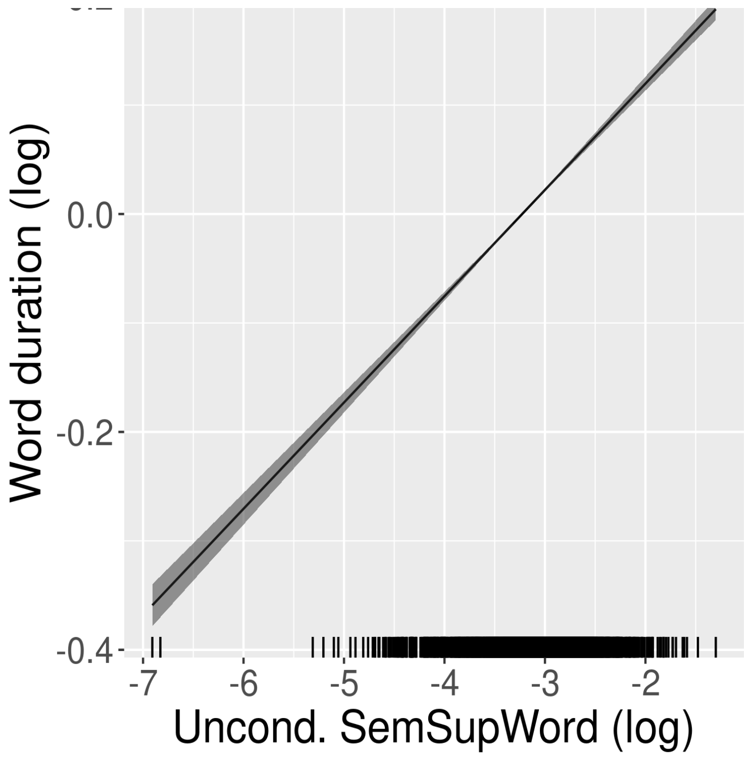


Figure 1. *Estimated partial effects of unconditional semantic support for word on word duration.*

[1] de Saussure, F. (1916). *Course in general linguistics*. McGraw-Hill.

[2] Lohmann, A. (2018a). Cut (n) and cut (v) are not homophones: lemma frequency affects the duration of noun–verb conversion pairs. *Journal of Linguistics*, 54(4), 753–777. https://doi.org/ 10.1017/s0022226717000378.

[3] Chuang, Y.-Y., & Baayen, R. H. (2021). *Discriminative learning and the lexicon: NDL and LDL*. Oxford: Oxford University Press.

[4] Plag, I. (2018). *Word-formation in English*, 2nd edition. Cambridge University Press.

[5] Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. *Complexity*, 2019(1), 1–39.

[6] Saito, M., Tomaschek, F., & Baayen, R. H. (2023). Articulatory effects of frequency modulated by semantics. In M. Schlechtweg (Ed.), *Interfaces of phonetics*. Berlin: De Gruyter.

[7] Lohmann, A. (2018b). Time and thyme are not homophones: A closer look at Gahl's work on the lemma-frequency effect, including a reanalysis. *Language*, 94(2), e180–e190. https : //doi.org/10.1353/lan.2018.0032.

[8] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.

[9] Nieder, J., Chuang, Y.-Y., van de Vijver, R., & Baayen, H. (2023). A Discriminative Lexicon Approach to Word Comprehension, Production and Processing: Maltese Plurals. *Language*, 99(2), 242–274. https://doi.org/10.31234/osf.io/rkath.

[10] Wood, S. N. 2006. Generalized additive models: An introduction with R. Boca Raton: CRC Press.

14. How do we feel about /x/? the role of iconicity and indexicality for language attitudes

Simon David Stein
Heinrich Heine University Düsseldorf

French sounds beautiful and romantic, German sounds harsh and aggressive—or so many people think. But why? Sociolinguists today agree that such language attitudes are an *indexical* phenomenon (Peirce 1958; Silverstein 2003). In this view, language attitudes result not from language itself but from differences in power, prestige, and sociocultural stereotypes (Giles and Niedzielski 1998). Linguistic features, e.g., particular sounds, point to groups of speakers, to their perceived ethnicity, region, socioeconomic status, and to their supposed character traits, like friendliness or harshness (see, e.g., Bayard et al. 2001; Coupland and Bishop 2007; Lippi-Green 2012). This activates aesthetic judgments about language, like "German sounds harsh".

However, the resurgence of iconicity research in the last years has rekindled interest in the question whether at least partly, some attitudes could also be an *iconic* phenomenon. In this view, aesthetic judgments may not only derive from social attributes of speakers but may be directly activated by properties of sound. For instance, voiced obstruents may be perceived negatively because of the articulatory challenge in producing them (Kawahara et al. 2021) or trilled /r/ may be associated with roughness because its discontinuous phonetics resembles rough textures (Winter et al. 2022). Despite several recent studies on both social and phonetic factors (e.g., Anikin et al. 2023; Hilton et al. 2022; Mooshammer et al. 2023; Reiterer et al. 2020), it remains an open question which specific features can carry which specific meanings.

To explore this question, this study put to the test the voiceless velar fricative /x/, one of the most notorious candidates hypothesized to make language sound unpleasant. In a novel, highly controlled experimental approach using newly created languages, 500 listeners of different language backgrounds each rated 3 out of 15 target stimuli with /x/ and 3 out of 15 control stimuli without /x/ on ten semantic differential scales (pleasantness, beauty, softness, shape, education, intelligence, friendliness, ordinariness, goodness, eroticism). Stimuli were matched in properties like mean sonority, CV ratio, voicing, syllable structure and syllable probability, and their remaining phonemic inventory. Audio files were generated with Amazon Polly using different language optimizations (Dutch, Arabic, German, Polish, Spanish) and different voices. Mixed models regressed ratings on condition (target or control) and on covariates encoding, e.g., how exposed listeners had been to /x/, how familiar the stimulus language felt to them, which real language they thought it was most similar to, and demographic information.

Results show that ratings on different scales are not, or only weakly, affected by whether /x/ is present. However, in interactions with exposure, stimuli with /x/ are rated significantly worse (e.g., uglier, more evil, less erotic) if the listener had more likely been exposed to /x/. This may indicate that these exposed listeners are more aware of respective stereotypes, and thus rate languages with /x/ worse despite being more used to them. Indeed, the strongest predictors are sociocultural, not phonetic-phonological, in nature: Listeners rate languages worse on almost all scales if they perceive the language as being less familiar, if they are male, and, crucially, if they *felt* the language sounded similar to a specific language group associated with "guttural sounds" or "harshness", such as languages from the Middle East or Germanic languages.

The results have implications for the interplay of iconicity and indexicality, and of social and aesthetic meaning. While language attitudes may partly be iconically affected by the phonological and phonetic makeup of language, this effect is, at least in the case of /x/, both modulated and overshadowed by sociocultural predictors. Judgments manifest through the indexical route, through which listeners rely on previous experience with languages and groups of speakers. The study also serves as a proof of concept based on which other linguistic features can be tested. The hope is that being more informed about why people feel the way they feel about language will, ultimately, help us combat linguicism and linguistic discrimination.

Anikin, Andrey, Nikolay Aseyev & Niklas Erben Johansson. 2023. Do some languages sound more beautiful than others? *Proceedings of the National Academy of Sciences of the United States of America* 120.17: e2218367120. https://doi.org/10.1073/pnas.2218367120.

Bayard, Donn, Ann Weatherall, Cynthia Gallois & Jeffery Pittam. 2001. Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics* 5.1: 22–49. https://doi.org/10.1111/1467-9481.00136.

Coupland, Nikolas & Hywel Bishop. 2007. Ideologised values for British accents. *Journal of Sociolinguistics* 11.1: 74–93. https://doi.org/10.1111/j.1467-9841.2007.00311.x.

Giles, Howard & Nancy Niedzielski. 1998. Italian is beautiful, German is ugly. In Laurie Bauer & Peter Trudgill (eds.), *Language myths*, 85–93. London: Penguin.

Hilton, Nanna H., Charlotte Gooskens, Anja Schüppert & Chaoju Tang. 2022. Is Swedish more beautiful than Danish? Matched guise investigations with unknown languages. *Nordic Journal of Linguistics* 45.1: 30–48. https://doi.org/10.1017/s0332586521000068.

Kawahara, Shigeto, Mahayana C. Godoy & Gakuji Kumagai. 2021. English speakers can infer Pokémon types based on sound symbolism. *Frontiers in Psychology* 12: 648948. https://doi.org/10.3389/fpsyg.2021.648948.

Lippi-Green, Rosina. 2012. *English with an accent: Language, ideology and discrimination in the United States*, 2nd edn. London; New York: Routledge. https://doi.org/10.4324/9780203348802.

Mooshammer, Christine, Dominique Bobeck, Henrik Hornecker, Kierán Meinhardt, Olga Olina, Marie Christin Walch & Qiang Xia. 2023. Does Orkish sound evil? Perception of fantasy languages and their phonetic and phonological characteristics. *Language and Speech*: 238309231202944. https://doi.org/10.1177/00238309231202944.

Peirce, Charles S. 1958. *Selected writings.* New York: Dover.

Reiterer, Susanne M., Vita Kogan, Annemarie Seither-Preisler & Gašper Pesek. 2020. Foreign language learning motivation: Phonetic chill or Latin lover effect? Does sound structure or social stereotyping drive FLL? In Kara D. Federmeier & Hsu-Wen Huang (eds.), *Adult and second language learning* (Psychology of Learning and Motivation 72), 165–205. Cambridge; San Diego; Oxford; London: Elsevier. https://doi.org/10.1016/bs.plm.2020.02.003.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23.3-4: 193–229. https://doi.org/10.1016/s0271-5309(03)00013-2.

Winter, Bodo, Márton Sóskuthy, Marcus Perlman & Mark Dingemanse. 2022. Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports* 12.1: 1035. https://doi.org/10.1038/s41598-021-04311-7.

15. What frequency effect reflects depends on what is measured

Ivan Yuen, Bistra Andreeva, Omnia Ibrahim, Bernd Möbius
Saarland University

Linguistic predictability is pervasive and has been shown to influence acoustic realization, with shorter duration for predictable units (e.g., Arnon & Cohen-Priva, 2013; Aylett & Turk, 2004). Yet these effects have been mostly focused on a specific linguistic level, rather than across levels. One type of predictability is 'frequency of occurrence', which could occur at the level of word (word frequency) and syllable (syllable frequency). Word and syllable frequency effects are assumed to arise from the ease of retrieval leading to fast response latency (RT). But only syllable frequency effect is postulated to arise from a mental syllabary that mediates phonological and phonetic encoding (e.g., Cholin et al., 2006). Although frequency effects have been reported to manifest in RT and acoustic duration, few studies examine both measures to get a better understanding of the role frequency plays in the phonological and phonetics processes that RT and acoustics reflect during spoken language production.

To investigate this issue, the current study examined the effect of high vs. low frequently-occurring monosyllabic (e.g., *Kind* 'child' vs. *Gift* 'poison') and disyllabic words (e.g., *Fehler* 'mistake' vs. *Feder* 'feather') in German, while controlling for the stressed vowel (as underlined). Target words contained either short or long monophthongs. High frequency monosyllabic words were manipulated to have low syllable frequency, whereas high frequency disyllabic words covary with stressed syllable frequency. Word frequency was estimated from CELEX and SUBTLEX-DE, while syllable frequency (i.e. surprisal) from a syllable-based unigram language model. Twenty monolingual German adults participated in a question-answer generation task. Prior to testing, participants named a series of pictures corresponding to the target words. At test, the pictures were used to elicit the target words that will be incorporated in the sentence response. Target words were elicited in 2 utterance positions: medial vs. final. We expect shorter acoustic vowel duration and faster RT in high than low frequently-occurring words and syllables. In light of the conflicting frequency manipulation for monosyllabic words, we expect the frequency effects across word and syllable to be cancelled out. But this is not the case for disyllabic words, as we expect cross-level cumulative frequency effects.

Two measures were taken and analysed using lmer (Bates, 2015) in R (R Core, 2022): stressed vowel duration and RT as measured from the beginning of the prompt question to the beginning of the verbal response. Results of vowel duration revealed a significant 4-way interaction with Word frequency, Number of syllables, Utterance position and Vowel type ($F = 5.1$, $df = 1, 1579$, $p = .02$*). To better understand the interaction, separate by-Number-of-syllables analyses revealed a Word frequency effect on short vowels in utterance-final monosyllabic words, but no overall effect in disyllabic words. Low frequency monosyllabic words increase duration of short vowels (Fig. 1). Such effects were not observed for long vowels. Results of RT revealed significant main effects: Word frequency ($F = 4.5$, $df = 1, 22$, $p =.04$*), Sentence duration ($F = 19.6$, $df = 1, 1233$, $p < .0001$***), and a significant Number of syllable-by-Utterance position interaction ($F = 5$, $df = 1, 21$, $p = .04$*). The number of syllables affects RT as a function of word frequency, with the effect in monosyllabic words. The interaction arises from faster RT for low frequency monosyllabic words, although they have conflicting high syllable frequency. This suggests RT to be primarily driven by syllable frequency during phonological/phonetic encoding. However, this effect was observed only in utterance-medial position, suggesting incremental speech planning. In sum, our findings do not support the view that frequency effects across levels are cumulative, in terms of acoustic duration and RT. Its implications will be discussed in terms of the multifaceted nature and role of frequency in speech production.

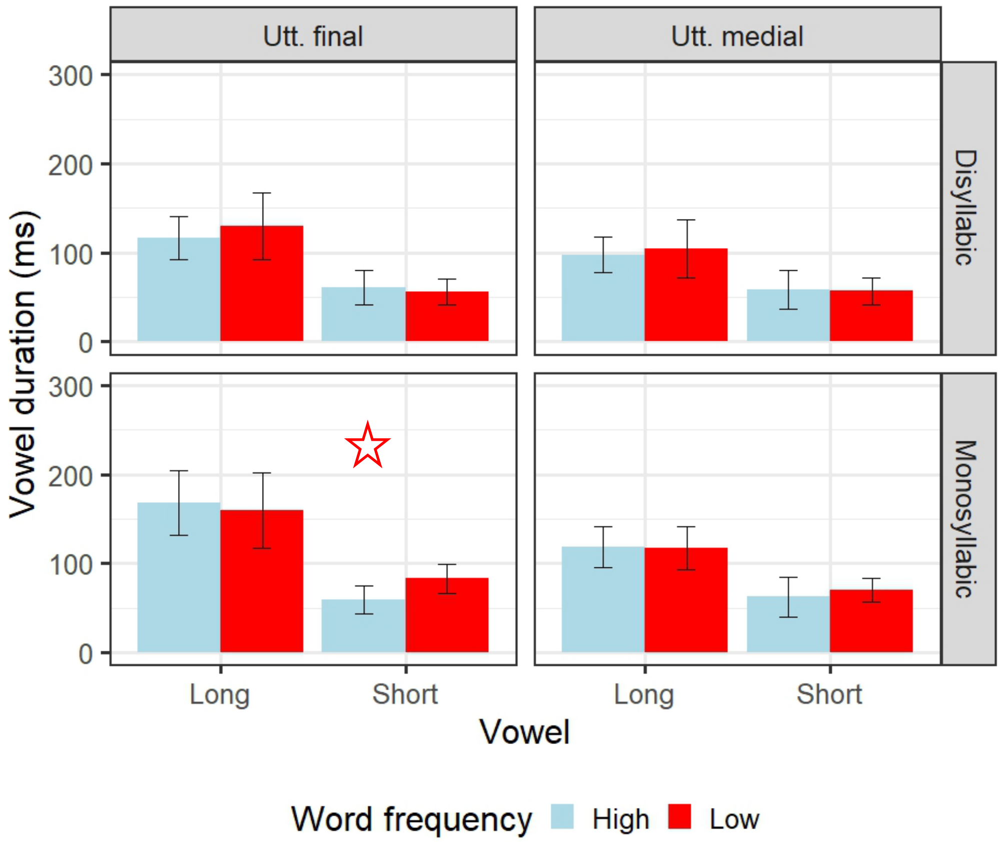


Fig. 1. *Duration of long (i.e. tense) and short (i.e. lax) stressed vowels in target monosyllabic and disyllabic words with high vs. low frequency in utterance-medial vs. final positions, with + /- 1 SD*

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56 (3), 349-371.

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech. *Language and Speech*, 47, 31-56.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Cholin, J., Levelt, W.J.M., & Schiller, N.O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99, 205-235.

R Core Team (2022). R: A Language and Environment for Statistical Computing

16. The acoustics of tone in Binumarien noun stems (Kainantu, Trans-New Guinea)

Renger van Dasselaar, Chenzi Xu
University of Oxford

This study examines the tonal patterns of disyllabic noun stems in Binumarien, employing fieldwork data and phonetic-acoustic analysis. It also establishes a pipeline for developing and applying language technologies to a low-resource language, benefitting future research.

Binumarien (endonym: *Afaqinna Ufa*) is a Kainantu language (Trans-New Guinea) spoken in the Eastern Highlands Province of Papua New Guinea [1], with around 1200 speakers. Most children grow up with Binumarien as their L1, along with Tok Pisin and regional languages.

In this study, speech data were collected between February and April 2023 in Binumarien village, from two male L1 speakers who were born and raised there. The speakers were recorded using a Tascam DR-10 audio recorder with a clip-on Rohde miniature microphone (Mono), and a Marantz PMD 661 solid state audio recorder (Stereo). The recorders ran simultaneously at 48 kHz, 24 bps. The Tascam recordings in WAV format were used here. The participants were asked to pronounce each noun stem four times: twice in isolation and twice with a suffix. The nouns were extracted from texts and a word list collected during a field trip in 2018-19 for a master's thesis [3] and from a literacy booklet [4]. In addition to pronouncing each item, speakers were asked to whistle each word for an impressionistic identification of tone. When the interviewer was in doubt, participants were asked to group words based on tone patterns.

All the annotated intervals in the recordings, totalling 1.9 hours, were used to train a preliminary acoustic model with Montreal Forced Aligner [5], which facilitates word- and segment-level time alignment between audio and text. Upon inspection, about 15% of the word-level alignments of nouns were manually corrected. Then sound intervals of disyllabic noun stems in citation form were extracted, within which f0 measurements every 10 milliseconds in voiced regions were obtained using Parselmouth library [6, 7] (floor: 75Hz, ceiling: 300Hz). The f0 measurements in Hz were normalised to semitones relative to the speaker mean, and the corresponding time values in seconds were linearly scaled to the range [0,1], to enable the comparison of the contour shapes across speakers and word items.

Impressionistically, we find four tones on syllables on the surface: high (H), low (L), falling (F), and rising (R) (see Table 1 for illustrations). F occurs only on long vowels and diphthongs, while R appears on these as well as on stem-final short vowels and occasionally between a L and a H. Whether F and R should phonologically be seen as separate tonal units or as combinations of underlying L and H is to be investigated.

Acoustically, the f0 contours of 50 disyllabic noun stems (coded in different colours) are shown in Figure 1, grouped in eight preliminary clusters. The eight clusters were based on the visual inspection of the surface f0 contours. Variations within each cluster can be partially attributed to the differences in syllable structure among these words. Some clusters may appear similar, but serve to differentiate word meanings. For example, the f0 contours of the minimal pair of *aandau* are illustrated in Figure 2, where the two f0 patterns, LH and LF, are affiliated with different word meanings ('white hair' and 'animal').

This empirical study contributes to our knowledge of the Binumarien tonal system. We aim to collect more data from additional speakers in the future to generalise our findings and gain deeper insights into the tonal system, distinct from many Southeast Asian varieties.

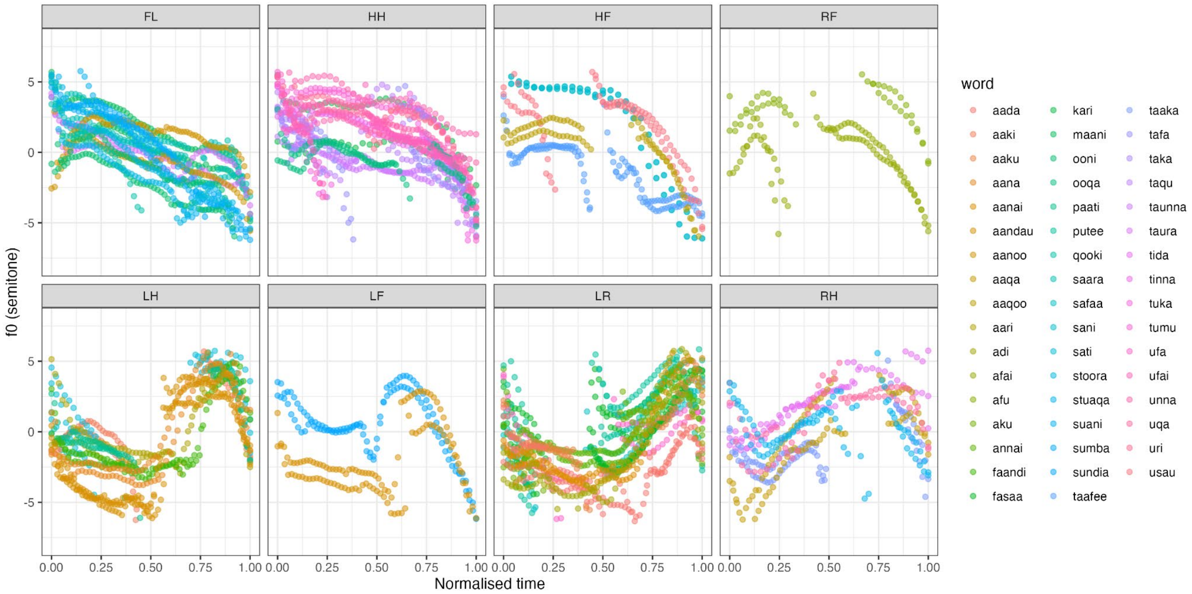


Figure 1. *The f0 contours of 50 disyllabic noun stems produced in citation form.*

H vs L	áínáíná ‘family’	ainá íná ‘thing
H vs F	a:ndáú ‘white hair’	a:ndáu ‘animal’
H vs R	tá:ká ‘t.o. tree’	sǎ:rá ‘centipede’
L vs R	faiqǐ ‘child’	faíqí ‘man’
L vs F	fe:fã ‘arrow’	dê:na ‘hummingbird’

Table 1. *Contrastive Tones in Binumarien Noun Stems.*

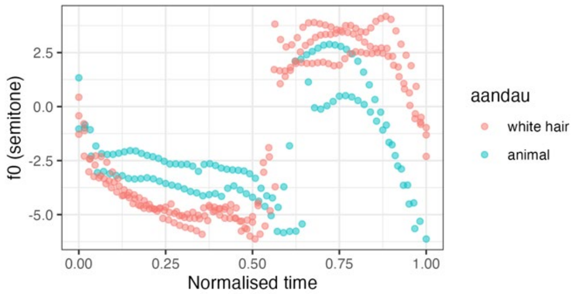


Figure 2. *The f0 contours of all aandau tokens produced by a speaker.*

[1] Pawley, A., & H. Hammarström. 2017. The Trans New Guinea family. In Bill Palmer (Ed.), *The languages and linguistics of the New Guinea area: A comprehensive guide*. Berlin: De Gruyter, 21–196.

[2] Oatridge, D. & J. Oatridge. 1965. Phonemes of Binumarien. In Frantz, Frantz, Oatridge, Oatridge, Loving, Swick, Pence, Staalsen, Boxwell & Boxwell (Eds.), *Papers in New Guinea Linguistics*. Canberra: Australian National University, 13-22.

[3] van Dasselaar, R. 2019. Topics in the Grammar of Binumarien: Tone and Switch-reference in a Kainantu Language of Papua New Guinea. Master’s thesis, Leiden University.

[4] Aadoo. 1973. *Oosana Oosana Aandau Ufa - Animals and Birds.* Summer Institute of Linguistics.

[5] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech*.

[6] Jadoul, Y., Thompson, B., & de Boer, B. 2018. Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics, 71*, 1-15. https://doi.org/10.1016/j.wocn.2018.07.001

[7] Boersma, P., & Weenink, D. 2021. Praat: doing phonetics by computer [Computer program]. Version 6.1.38, retrieved 2 January 2021 from http://www.praat.org/

17. Impacts of visual feedback on L2 learner monophthong vowel production in Spanish

[1]Heather M. Offerman, [2]Louis Gehrke
[1]University of Arkansas [2]University of Texas at Austin

Research suggests that pronunciation instruction should be developed for second language (L2) learners (Derwing & Munro, 2005). The use of *visual feedback* (VF) interventions as a means of instruction for adult L2 learners in producing more target-like realizations has been proposed by several previous works in the field (Olson, 2014). VF is broadly defined as "the use of visual representations of speech sounds or articulatory movements (or both) as a pedagogical tool for teaching or improving second language (L2) pronunciation" (p. 240; Offerman & Olson, 2023). L2 learners view NS productions of a sound or sounds visually represented (typically on a screen) along with their own productions of the same sound or sounds for comparison, aiding them in identifying non-target-like productions (Dlaska & Krekler, 2008). Historically, VF has primarily focused on suprasegmental features of the L2, such as intonation contours (Chun, 1989; de Bot, 1983; Hardison, 2004). More recently, other researchers have focused the use of visual feedback on segmental features (vowels and consonants) in languages such as Japanese and Spanish (Motohashi-Saigo & Hardison, 2009; Okuno, 2013; Olson & Offerman, 2021).

Indirect VF has been increasingly tested and used as more free and accessible visual analysis software has become available, such as Praat (Boersma & Weenink, 2019). Along with displaying suprasegmental features, segmental features are also visible in Praat via spectrogram and sound wave displays (Saito, 2007). This is categorized as a form of indirect VF, as it displays acoustic visualizations of a speaker's speech (i.e., not direct representations of articulators and only corresponding to speech articulators). While VF studies involving romance languages have primarily focused on consonants, some VF studies have produced mixed results when testing vowel production, finding that VF interventions had little to no impact on L2 vowel production (e.g., for French and Spanish vowels; Olson, 2022; Ruellot, 2011). However, as vowel production can impact intelligibility and has been noted to be challenging for L2 learners, it is crucial that vowel instruction be developed. One case that has yet to be tested involves monophthongs in L2 Spanish produced by L1 English speakers. In Spanish, stressed vowels /e/ and /o/ are never diphthongized, unless indicated by the orthography with 'i' and 'u' (/i/ and /u/). However, in American English, /e/ and /o/ tend to be produced as [eɪ] and [oʊ] in stressed positions. As L2 learners tend to produce stressed vowels /e/ and /o/ as diphthongs in their L2 Spanish, likely due to the influence of American English, the current study focuses implementing a VF paradigm involving formant patterns as a means of teaching these pronunciation differences.

Thirty-five intermediate-level L2 learners (Experimental = 17; Control = 18) from a large U.S. university participated in pretest, posttest, and delayed posttest recordings, which consisted of 20, bisyllabic tokens (10 per vowel) in CVCV order embedded in a carrier phrase and spontaneous speech task (i.e. picture task). Participants were given a language background questionnaire (Birdsong et al., 2012) and took part in two interventions per vowel /e/ and /o/, where they received VF through Praat on their own recordings, comparing them to native-speaker productions. Data collection has been completed, and phonetic analyses via Praat will measure reduction of diphthongization by assessing F2 values. Statistical analyses will be carried out via the lme4 package (Bates et al., 2014) in RStudio (Allaire, 2012), with group and time set as fixed effects and participant and vowel set as random effects. Previous pilot results indicate that participants can reduce diphthongization. It is anticipated that participants will significantly improve from pretest to posttest, but there will be a decline in performance for the delayed posttest.

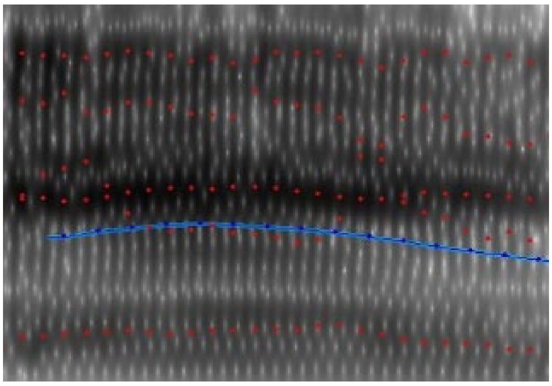


Figure 1. *Spanish [e]*

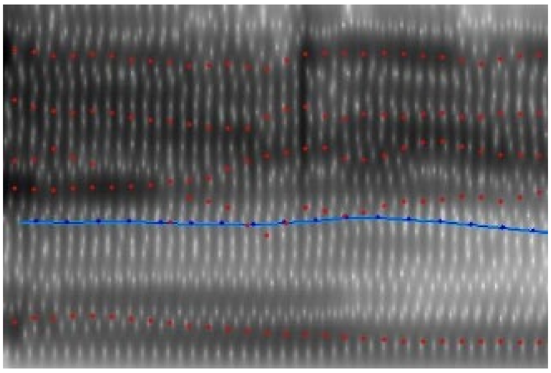


Figure 2. *American English [eɪ]*

[1] Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual language profile: An easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin*. https://sites.la.utexas.edu/bilingual/

[2] Boersma, P., & Weenink, D. (2019). Praat: Doing Phonetics by Computer (Computer program).Version 6.1. 02. Retrieved from http://www.praat.org/.

[3] Chun, D. (1989). Teaching tone and intonation with microcomputers. *CALICO Journal, 7*(1), 21–46. http://www.jstor.org/stable/24147465

[4] de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech, 26*(4), 331–350. https://doi.org/10.1177/002383098302600402

[5] Hardison, D. (2004). Generalization of computer assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology, 8*(1), 34–52. https://www.learntechlib.org/p/74458/

[6] Motohashi-Saigo, M., & Hardison, D. (2009). Acquisition of L2 Japanese geminates Training with waveform displays. *Language Learning & Technology, 13*(2), 29–47. https://www.learntechlib.org/p/74427/

[7] Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second language segmental production: The generalizability of pronunciation gains. *System*, *59*, 45–60. https://doi.org/10.1016/j.system.2016.03.003

[8] Okuno, T., & Hardison, D. M. (2016). Perception-production link in L2 Japanese vowel duration: Training with technology. *Language Learning & Technology, 20*(2), 61–80.

[9] Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning and Technology, 18*(3), 173–192. http://llt.msu.edu/issues/ october2014/olson. pdf

[10] Olson, D. J., & Offerman, H. M. (2021). Maximizing the effect of visual feedback for pronunciation instruction: A comparative analysis of three studies. *Journal of Second Language Pronunciation*, *7*(1), 89–115. https://doi.org/10.1075/jslp.20005.ols

Workshop: Heritage language phonetics and phonology

The “uh” gives you away - Phonetic aspects of filler particles in bilingual heritage speakers’ narrations

Marlene Böttcher, Margaret Zellers
Kiel University

Filler particles (FPs) like *uh* occur frequently in spontaneous discourse yet are often not considered linguistic items [cf. 1]. However, FPs show language specific forms [1, 2] which have to be acquired in both the first and the second language and are also susceptible to cross-linguistic influence. While FPs' high frequency, especially in spontaneous speech [3], might help promote their acquisition, their prosodically non-salient form [4, 5] can lead them to be easily missed by listeners [6]. The acquisition of language specific FP forms might, therefore, pose challenges, especially for bilinguals with different amounts and kinds of input across their two languages. As simultaneous or early sequential bilinguals, heritage speakers (HS) grow up with a heritage language (HL) and the majority language (ML) [7]. The different societal status of their two languages results in HL input predominantly in a familial setting and ML a more diverse input in almost every sphere of the public life.

FPs have been described for several languages as having a centralized, yet, language specific quality [3]. FPs in English are reported with a low central vowel [8], in Russian with low or front central vowel [9], while FPs in German show high variation [2]. Research on bilinguals’ FP vowels suggest that bilingual speakers produce language-specific FP vowels in both their languages, yet, with acoustic differences from monolingual speakers [e.g. 10].

The current research addresses the question whether bilingual HSs produce language specific FPs in their two languages as part of their linguistic repertoire and whether they differ in FP vowels compared to monolinguals, adding to a “heritage accent” [7, p. 10]. To do so, data from the RUEG corpus [11] of bilingual HSs living in the U.S. and in Germany, both with Russian as a HL, is considered and compared to monolingual data of their respective languages. Specifically, data produced by 50 adolescent speakers (10 from each speaker group), each producing a minimum of 5 FPs in either language, is investigated.

The acoustic analysis of the first (F1) and second (F2) vowel formants of FPs in Praat [9] reveals a large overlap of vowel qualities across HSs’ two languages both in Germany (Pillai = 0.03) and in the U.S. (Pillai = 0.04). HSs in the U.S. produce FP vowels within a smaller area within the vowel space in their English compared to their Russian, while HSs in Germany produce FP vowels with more variance in the F2 in their German compared to their Russian (see Fig. 1). The comparison with monolingual data suggests an effect of the language contact within HSs in both their HL and their ML (see Fig. 2). FP vowels in of HSs in the U.S. show lower variance and cover a smaller area compared to both monolingual speaker groups, while HSs in Germany show more overlap with the respective monolingual group.

The results from this study suggest a bidirectional influence of ML and HL in the phonetic form of FPs and provide insight into phonetic features potentially adding to a “heritage accent”.

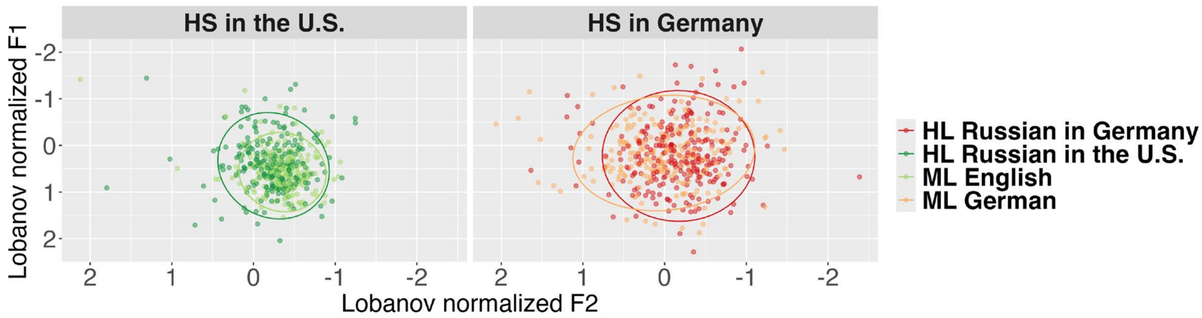


Figure 1: *Lobanov normalized formant values of filler particle vowels for the two heerittage speaker (HS) groups across the speakers' two languages* .

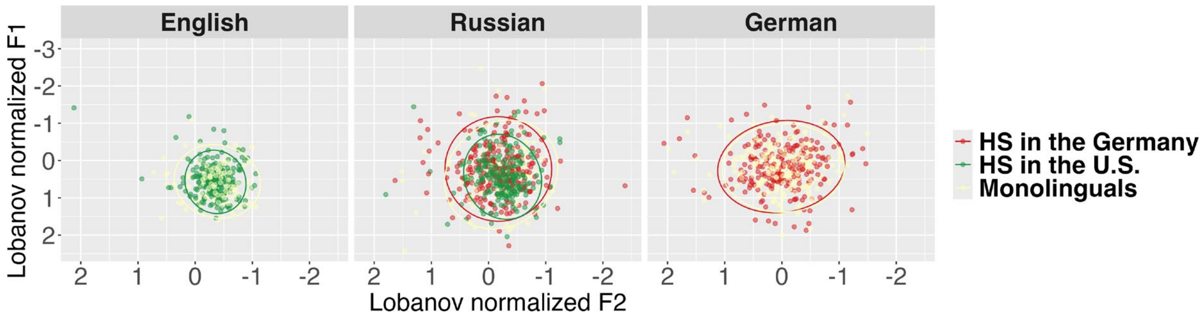


Figure 2: *Lobanov normalized formant values of filler particle vowels of heritage speaker(HS) groups compared to monolingual speakers across languages.*

[1] Clark, H. H. and FoxTree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. Cognition, 84 (1), pp. 73-111

[2] Candea, M., Vasilescu, I., & Adda-Decker, M. (2005). Inter-and intra-language acoustic analysis of autonomous fillers. Proceedings Recueil Des Communications, pp. 47-52

[3] Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., and Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. Language and speech, 44(2), pp. 123-147

[4] Belz, M. (2021). Die Phonetik von *äh* und *ähm*. Akustische Variation von Füllpartikeln im Deutschen. Springer Berlin Heidelberg.

[5] Braun, A., Elsässer, N., and Willems, L. (2023). Disfluencies revisited – are they speaker specific? Languages, 8(3):155.

[6] Niebuhr, O. and Fischer, K. (2019). Do not hesitate! - unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. Proc. Interspeech 2019, pp. 544-548

[7] Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and cognition*, *23*(1), pp. 4-20.

[8] Shriberg, E. (2001). To 'errrr'is human: ecology and acoustics of speech disfluencies. *Journal of the international phonetic association*, *31*(1), 153-169.

[9] Stepanova, S. (2007). Some features of filled hesitation pauses in spontaneous Russian. In *Proceedings of ICPhS* 2007, pp. 1325-1328

[10] García-Amaya, L., & Lang, S. (2020). Filled pauses are susceptible to cross-language phonetic influence: Evidence from Afrikaans-Spanish bilinguals. *Studies in Second Language Acquisition*, *42*(5), pp. 1077-1105.

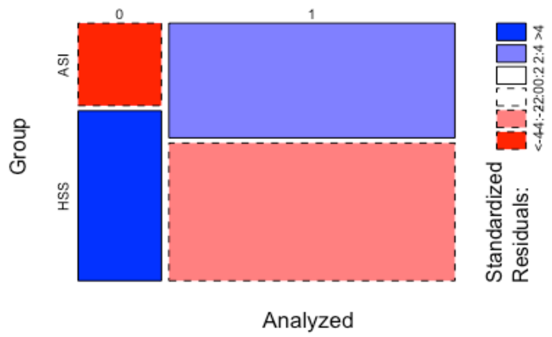
[10] Lüdeling, A., Alexiadou, A. Allen, S. et al. (2024). RUEG Corpus (1.0) Zenodo https://doi.org/10.5281/zenodo.11234583

[11] Boersma, P. and Weenink, D. (2021). Praat: doing phonetics by computer. version 6.2.03. http://www. praat. org/.

Vowel production and gender comprehension in Spanish-English bilinguals

Laura Colantoni, Laura Escobar, Ana T. Pérez-Leroux, Danielle Thomas, Crystal Chen, Alexei Kochetov
University of Toronto

We analyze word-final unstressed Spanish vowels /a e o/ produced by early and late Spanish-English bilinguals with the goal of determining to which extent variability in vowel production due to contact with English may impact learning and representation of Spanish gender, as proposed by the modular interaction hypothesis [1, 2]. Spanish gender is quasi-transparent with over 60% of nouns ending in /o a/ aligning as masculine or feminine respectively, and /e/-ending nouns (13%) being predominantly masculine [3]. These vowels merge and centralize in heritage Spanish-English speakers [4, 5]. Moreover, in some Spanish varieties (Mexican and Andean), word-final vowels may devoice after voiceless obstruents [6, 7]. Thus, contact with English and variety-specific phonetic features conspire to obscure the input, making phonology-based gender attribution [8] more difficult for heritage speakers. If variability in vowel production is a predictor of gender accuracy, participants who merge vowels in production should diverge in gender performance. Adult Heritage Speakers of Spanish (HSs; n=23; living in North America before the age of 11), and adults Spanish Immigrants (ASIs; n=25; who moved to Canada after 14) performed two tasks. Production data were obtained from a picture naming task with 244 targets ending in unstressed /a e o/. Non-analyzable vowels (devoiced, creaky) were identified first, and F1-F2 values (Nearey [9]-normalized) were extracted at midpoint from voiced vowels and the Bhattacharyya affinity (BA) scores [10, 11] were computed. To study gender comprehension, an eye-tracking task, implemented in a Tobii Pro Fusion system at 120 Hz, presented three images with an unspliced audio prompt. Stimuli were controlled for gender, animacy, noun morphology, and determiner (definite/possessive), counterbalanced across participants. We report here results on proportion of gaze fixations to target extracted from a window that included the word ending +300ms. Our focus is on gender-transparent nouns preceded by a possessive determiner (e.g. *mi pal**a*** 'my shovel' vs. *mi pal**o*** 'my stick'), where the final vowel offers the only cue to gender. Preliminary results show that, although the proportion of devoiced vowels was higher in the HS group (Fig.1), between-group differences were not significant. BA scores (Fig.2) were also similar between groups (HSs: [a-o] = .89; [a-e] = .85; ASIs: [a-o] = .90; [a-e] = .85). Results of the eye-tracking task (Fig.3) revealed no significant group effects but a significant group*alternation effect, indicating that HSS fixate less to vowel-alternating nouns, compared to ASIs. Due to wide range of individual variation in both tasks, the correlation between the results of the vowel overlap analysis and the eye-tracking task turned out not to be significant. Ongoing analyses explore the role of a third factor, lexical performance. Although our preliminary analyses of the groups' performance do not fully support the modular interaction hypothesis, the within-group variability observed in the HSS group suggest that the highly variable input to which heritage populations are exposed combined with absence of formal education in the minority language, may impact the acquisition of grammatical categories.



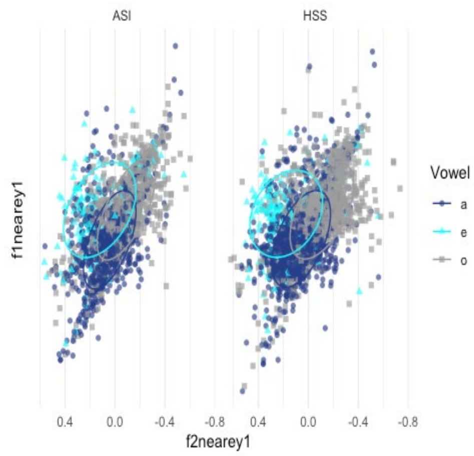


Figure 1. *Mosaic plot of analyzable vowels* Figure 2. *Vowel plots (HSs, right, Controls, left)*

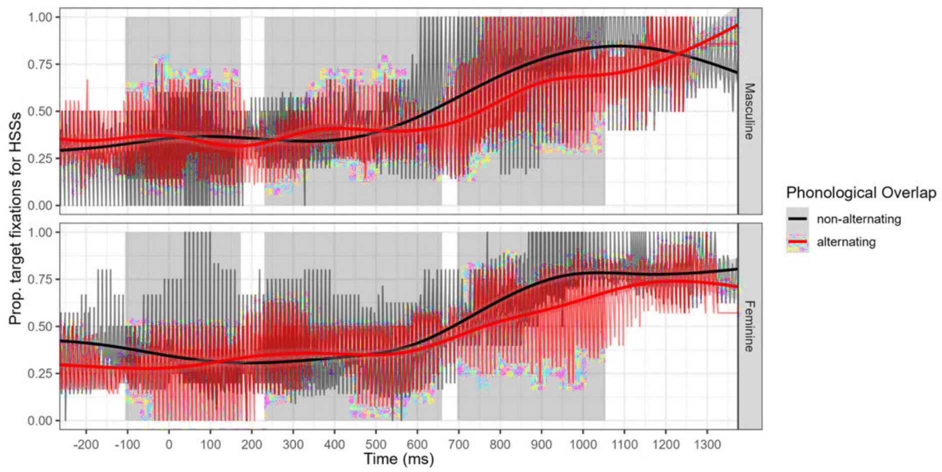


Figure 3. *Proportion of fixation to target in alternating vs. non-alternating nouns (HSS group)*

1. Colantoni, L., R. Martínez, N. Mazzaro, A. T. Pérez-Leroux, and N. Rinaldi. 2020. A Phonetic Account of Spanish-English Bilinguals' Divergence with Agreement. Languages, 5, 4, 58. https://doi.org/10.3390/languages5040058
2. Pérez-Leroux, A. T., López, Y. Á., Barreto, M., Cuza, A., Marinescu, I., Yang, J., & Colantoni, L. (2023). The Phonetic and Morphosyntactic Dimensions of Grammatical Gender in Spanish Heritage Language Acquisition, Heritage Language Journal, 20(1), 1-37. doi: https://doi.org/10.1163/15507076-bja10017
3. Clegg, J. H. (2011). A frequency-based analysis of the norms for Spanish noun gender. *Hispania*, *94*(2), 303-319.
4. Menke, M. (2010). Examination of the Spanish vowels produced by Spanish-English bilingual children. *Southwest Journal of Linguistics*, *28*(2), 98-135.
5. Mazzaro, N., Colantoni, L., & Cuza, A. (2016). Age effects and the discrimination of consonantal and vocalic contrasts in heritage and native Spanish. In C. Tortora, M. den Dikken, I. Montoya, & T. O'Neill (Eds.), *Romance Linguistics 2013: Selected papers of the 43th Linguistic Symposium on Romance Languages (LSRL), New York, 17-19 April, 2013* (pp. 277-300). John Benjamins.
6. Moreno Fernández, F. (2000). *Qué español enseñar* [What Spanish to teach]. Arco Libros.

7. Dabkowski, M. (2018). *Variable vowel reduction in Mexico City Spanish*. [Unpublished doctoral dissertation]. Ohio State University.
8. Culberston, J.., H. Jarvinen, F. Heggarty, and K. Smith. 2019. Children's sensitivity to phonological and semantic cues during noun class learning evidence for a phonological bias. *Language* 95: 268–93.
9. Nearey, T. M. (1977). *Phonetic Feature Systems for Vowels*. Dissertation, University of Alberta, Edmonton, AB, Canada. Reprinted 1978 by the Indiana University Linguistics Club.
10. Johnson, D. E. (2015). Quantifying Overlap with Bhattacharyya's Affinity. Paper presented at NWAV 44, Toronto, ON, Canada, October 25. Available online: https://danielezrajohnson.shinyapps.io/nwav_44 (accessed on 28 August 2020).
11. Strelluf, C. (2016). Overlap among back vowels before /l/ in Kansas City. *Language Variation and Change* 28: 379–407.

Final /t d/ deletion in American English spoken by monolinguals and heritage Spanish speakers

Karolina Broś
University of Warsaw

Final /t d/ deletion is a well-known phenomenon of English phonology whereby /t d/ are deleted when preceded by another consonant in connected speech, e.g. best man [bes mæn] or kept going [kep goʊɪŋ]. The process has been reported for many varieties of English, from American, through Chicano English to British dialects (Coetzee, 2004; Baranowski & Turton, 2020). However, different studies have led to contrary findings concerning the phonological vs. morphological conditioning of the process, and the completeness of deletion is still subject to debate (Purse, 2019; Purse and Turk, 2019).

The discrepancies found between previous studies can be due to methodological limitations. Sociolinguistic studies based on large corpora lack control over all the factors involved, while controlled articulatory studies may result in process suppression by speakers and are usually based on very small numbers of participants. Importantly, some research suggests that complete /t d/ deletion may not be as frequent as reported by sociolinguists. Instead, we have either full [t d] or weakened, undershot pronunciations that result in the sound being inaudible (Purse, 2019). This paper aims at finding the middle ground between a fully controlled study based on isolated phrases and a spontaneous speech elicitation. I use whole sentences in which final /t d/ are contextualized in both a reading and a repeating from memory condition to control for orthographic bias. This should result in more naturalistic productions than word/phrase lists.

On the heritage language part, it is known that Hispanic English has developed features of its own that are intermediate between Spanish and English in line with the assumption that prolonged language contact may lead to “interdialectal forms” (Trudgill 1986:62). Some examples include changes in the /ai/ diphthong influenced by Southern American monophthongization (Wolfram et al., 2011) or changes in the VOT of stops in heritage Spanish speakers’ English and Spanish (Kim, 2011). It is worth noting that Spanish syllable structure does not allow complex codas at the end of a word, and many Spanish varieties weaken or fully delete even simple codas in this position (Hualde, 2005). Thus, it is only natural for Spanish speakers to simplify final clusters. This is precisely what happens in second language acquisition, especially in past tense forms. If there is an influence or interference of Spanish on the English of bilingual Spanish-English speakers, we might expect that /t d/ deletion will be more frequent or follow a different pattern than in American monolinguals (Santa Ana, 1996). The present study addresses this issue by comparing 20 monolingual and 20 heritage Spanish speakers of American English from California.

The participants were asked to produce a total of 120 sentences repeated twice and constructed in a way that makes it possible to compare minimal pairs of sound sequences. Each sentence is different and contains a target phrase starting with a verb ending in a bilabial sound in either the past simple or the present tense, followed by another bilabial or a neutral context (vowel). A total of 40 verbs were used. Taking the verb stop as an example, we have:

1) She finally stopped making noise, 2) She stopped eating to raise a toast, 3) Stop making excuses and start working, 4) I had to stop eating junk food. The analysis of those productions includes auditory and acoustic determination of whether the /t d/ is pronounced or not, the drop in intensity and the subsequent increase towards the following vowel, and the duration of the sequence with and without an intervening /t d/. The idea behind the analysis is that, in cases where e.g. [t] is inaudible in stopped making, if it is completely deleted and there is no tongue tip raising gesture then we would expect the preceding [p] to be of the same duration as when the [t] is present (audible) or as in a control condition (e.g., stop eating, stop making). If [p]

duration is longer, this means that it includes the inaudible [t] at least in some part and hence deletion is not complete.

Baranowski, M., & Turton, D. (2020). TD-deletion in British English: New evidence for the long-lost morphological effect. *Language Variation and Change, 32*(1), 1-23.

Coetzee, Andries. (2004). *What it means to be a loser: non-optimal candidates in optimality theory* (Doctoral dissertation). University of Massachusetts, Amherst.

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.

Kim, Ji Young. (2011). L1-L2 phonetic interference in the production of Spanish heritage speakers in the US. *The Korean Journal of Hispanic Studies*, 4, 1–28.

Purse, Ruaridh. (2019). The Articulatory Reality of Coronal Stop 'Deletion'. In *Proceedings of the 19th International Congress of the Phonetic Sciences* (XIX). Melbourne, Australia.

Purse, Ruaridh, & Turk, Alice. (2016). '/t d/ deletion': Articulatory gradience in variable phonology. Paper presented at LabPhon 2016.

Santa Ana, Otto. (1996). Sonority and syllable structure in Chicano English. *Language Variation and Change, 8*, 3–91.

Trudgill, Peter. (1986). *Dialects in Contact*. Oxford, U.K.: Blackwell.

Wolfram, Walt, Kohn, Mary E., & Callahan-Price, Erin. (2011). Southern-Bred Hispanic English: An Emerging Socioethnic Variety. In J. Michnowicz & R. Dodsworth (Eds.), *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics* (pp. 1–13). Somerville, MA: Cascadilla Proceedings Project.

Preaspiration in Modern vs. North American heritage Icelandic

Meike Rommel[1], Tanja Kupisch[2], Christiane Ulbrich[3], Nicole Dehé[1]
[1]University of Kostanz, [2]Lund University, [3]University of Cologne

We study preaspiration in heritage Icelandic as spoken in Manitoba/Canada (North American Icelandic; NAmIce) as compared to Modern Icelandic as spoken in Reykjavík/Iceland (ModIce) to investigate its robustness under conditions of extreme language contact and, specifically, whether it is subject to cross-linguistic influence (CLI) from English or, conversely, maintained or even overemphasized given its contrastive value in Icelandic.

Preaspiration is a salient phenomenon of ModIce. It occurs on (historical) geminates (*hattur* [ˈhahtʏr̥] 'hat'), and on stops before lateral /l/ and nasals /m, n/ (*epli* [ˈɛhplɪ] 'apple', *opna* [ˈɔhpna] 'to open') (Árnason 2011), and is preceded by a short vowel (V). L1 Icelandic children acquire preaspiration comparatively late, by 5-6 years, but before school age (Sigurjónsdóttir 2002, 2008; Másdóttir 2008). While preaspiration does occur in North American English (Hejná et al. 2021), it is infrequent and its status is different from preaspiration in ModIce. Late acquisition and its near absence in English could make it vulnerable to CLI, while its contrastive function could have the opposite effect.

Based on previous literature studying heritage language (HL) phonology (Polinsky 2018 among many others), we address the following hypotheses (H):

H1: Preaspiration is retained in moribund NAmIce, but realized differently from ModIce.

H2: The differences between ModIce and NAmIce follow strategies previously observed for related segmental phenomena in HLs. Possible strategies employed in NAmIce include lengthening of the preceding vowel or other phases of the plosives, overemphasized realization, and post-aspiration instead of preaspiration.

We conducted a picture-naming task with 60 stimuli in three conditions according to Icelandic syllable structure and presence/absence of aspiration, following Árnason (2011): (1a): short V, preaspiration; (1b): short V, half-long C, no preaspiration; (1c): long V, no preaspiration; example pictures in Figure 1.

a.	hattar [ˈhahtar̥]	'hats'	VhCV
b.	haddar [ˈhatˑar̥]	'hair, hairdo (Pl.)'	VCˑV
c.	hatar [ˈhaːtar̥]	'hates'	VːCV

We present data from 20 ModIce speakers, recorded in Reykjavík (2023), and 20 NAmIce speakers, recorded in Manitoba (2023/24). The recordings were annotated in Praat (Boersma & Weenink 2025). Duration measurements were carried out for the pre-aspiration phase, the vowel preceding the target consonant, and breathy parts of the vowel and the individual plosive phases.

Results show that NAmIce speakers are aware of the distribution of preaspiration and realize it, but vary in its realization as compared to ModIce speakers, confirming H1. For example, some NAmIce speakers realize a breathy phase after the vowel (voiced friction) like ModIce speakers, but unlike ModIce speakers, this is not followed by voiceless friction (preaspiration proper) (Figures 2 vs. 3). We also find that NAmIce speakers overemphasize preaspiration, resulting in a longer preaspiration phase compared to ModIce (H2). Moreover, our results confirm the relevance of the segmental environment (preceding vowel, place of articulation of target plosive; Hejná et al. 2021) for the production of preaspiration in both ModIce and NAmIce. More generally, these results indicate that typologically marked structures, if phonemically relevant, may be retained in extreme language contact settings, undergoing changes independently of the contact language.

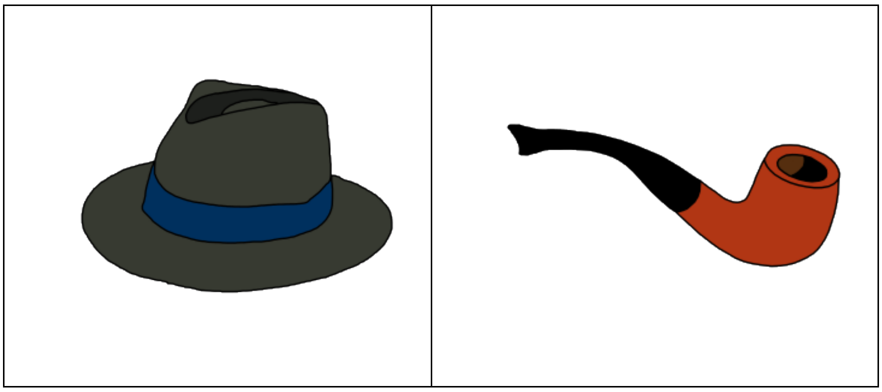


Figure 1. *Example stimuli; left-hand panel: hattur [ˈhahtʏr̥] (‘hat’), see (1a); right-hand panel: pípa [pʰiːpa] (‘pipe’), see (1c).*

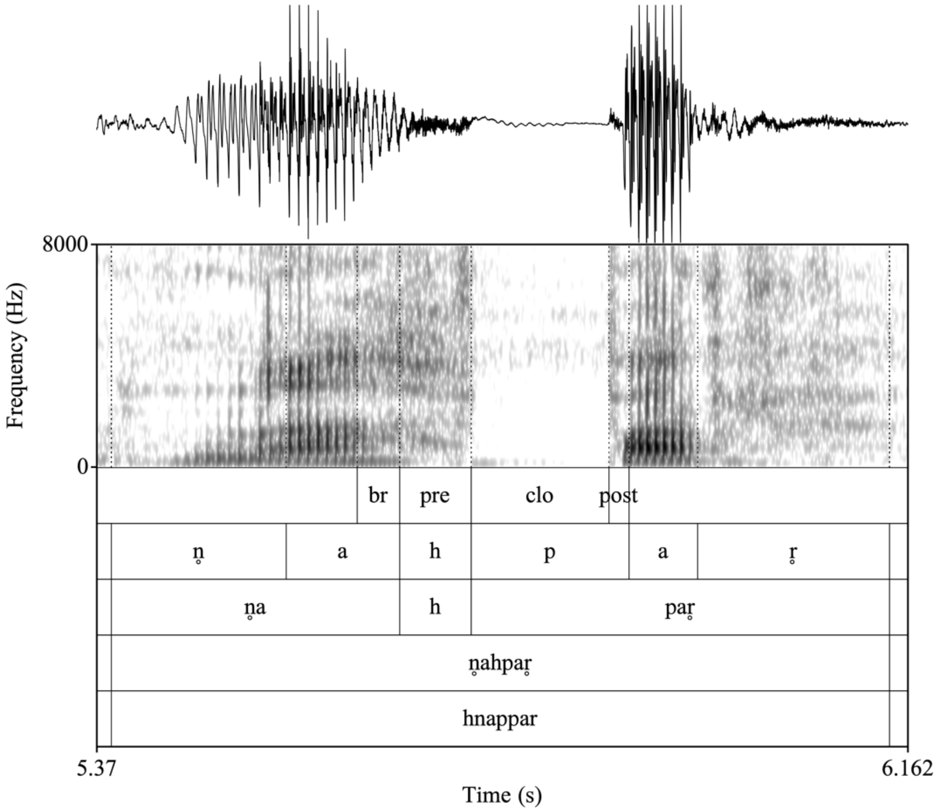


Figure 2. *ModIce speaker R05_o, male, 85 years old; hnappar ‘buttons’ (br = breathy, pre = preaspiration, clo = closure phase, post= post-aspiration).*

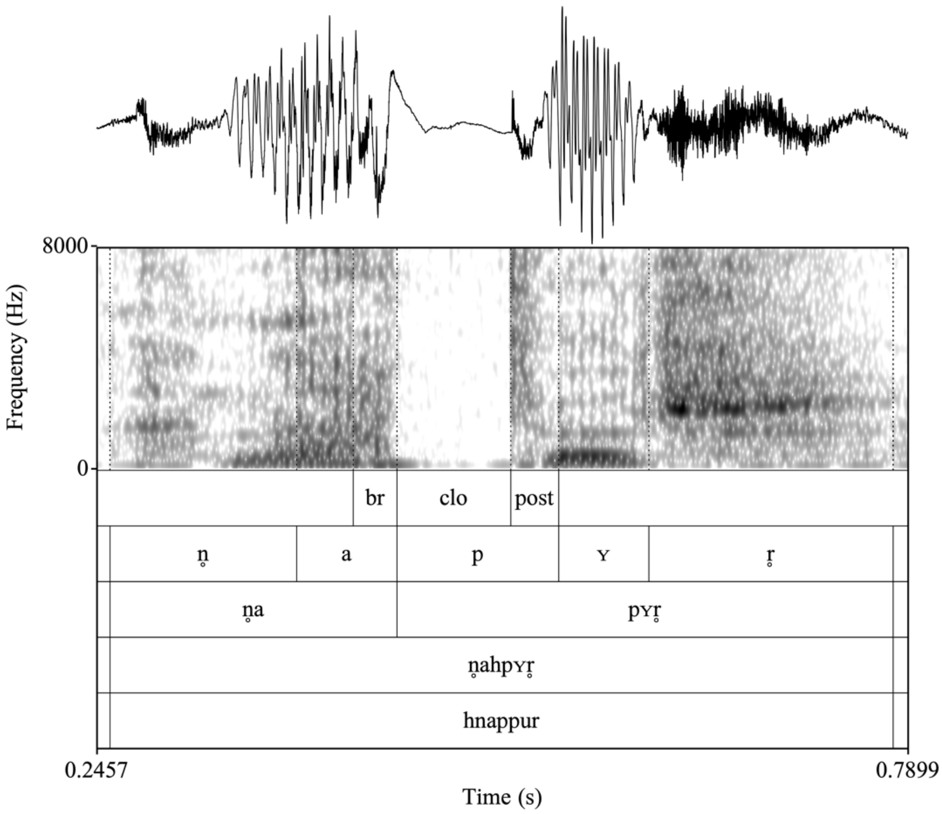


Figure 3. *NAmIce speaker CA19_o, male, 67 years old; hnappur 'button' (br = breathy, pre = pre-aspiration, clo = closure phase, post= post-aspiration).*

Árnason, K. (2011). *The phonology of Icelandic and Faroese.* Oxford: Oxford University Press.

Boersma, Paul &Weenink, David (2025). Praat: doing phonetics by computer [Computer program]. Version 6.4.27, retrieved 27 January 2025 from https://www.fon.hum.uva.nl/praat/

Hejná, M., Kaźmierski, K., & Guo, W. (2021). Even Americans pre-aspirate. *English World-Wide 42(2)*, 200-226.

Polinsky, M. 2018. Heritage Languages and Their Speakers. Cambridge University Press, Cambridge.

The early vowel production by heritage Cantonese speakers

Chen Lan, Peggy Mok
The Chinese University of Hong Kong

The present study investigates the acoustic patterns of eleven Cantonese vowels produced by heritage speakers (HSs) in early childhood and compares the heritage language (HL) patterns with that of the age-matched homeland children, aiming to explore how HSs acquire Cantonese vowels in the early stage of language development, whether and how their development differs from the homeland development, and what factors accounting for the HL patterns. Thirty HSs aged between 2;0 and 6;11 (HL children) who were born in Canada and grew up in or around Vancouver, with at least one of the parents growing up speaking Hong Kong Cantonese, participated in a picture-naming task conducted in a sound-proof room. Data collected from thirty homeland children of the same ages who were born and raised in Hong Kong were included for comparison. Tokens of the seven long vowels in open syllables and in nasal coda environments and of the four short vowels in nasal or stop coda environments were produced by each child twice in isolation. Acoustic measurements were done using *FormantPro* [1] in Praat [2], and the statistical analysis were conducted in R [3].

Results indicated that both HL and homeland children were able to produce stable vowel height distinctions in the first three years of age, while the vowel frontness continued changing and have not stablized even at the oldest age band (6;0-6;11). However, the acoustic features of individual vowels produced by HL children differed from that of the homeland children, and the divergence appeared early on. First, the two- and three-year-old HL children favored the front position of the vowel space while the homeland children centralized the vowels (see Figure 1). Even though the vowel space gradually expanded as age increased, the heritage vowel triangle was larger, fronter, and lower (see Figure 2). Second, the fronting of /aː/ and the raising of the pre-velar /ɛ/ were associated with heritage vowel production (see Figure 3), rendering their Cantonese being different from the homeland patterns. Third, HSs contrasted the long-short vowels mainly in vowel quality, while the homeland children contrasted them in both quality and quantity. These findings lend support to the influence of the ambient language on the HL development, for example, the early exposure to the majority language (ML) English and the quality of the HL inputs.

Our findings also revealed a positive relationship between the amount of HL inputs and outputs and the degree of similarity to the homeland patterns. In addition, compared to the HL children whose age of ML acquisition was at or above 24 months, those who started to be regularly exposed to the ML from birth performed better in producing the vowel distinctions, but the acoustic details of some vowels differed from the homeland pattern.

In conclusion, while the vocal development and the phonological process are relatively parallel, phonetic differences appear between HL children and their monolingual peers. The HL children form an HL-specific vowel production patterns containing variants between an element in the HL sound system and an element influenced by HL-ML interaction.

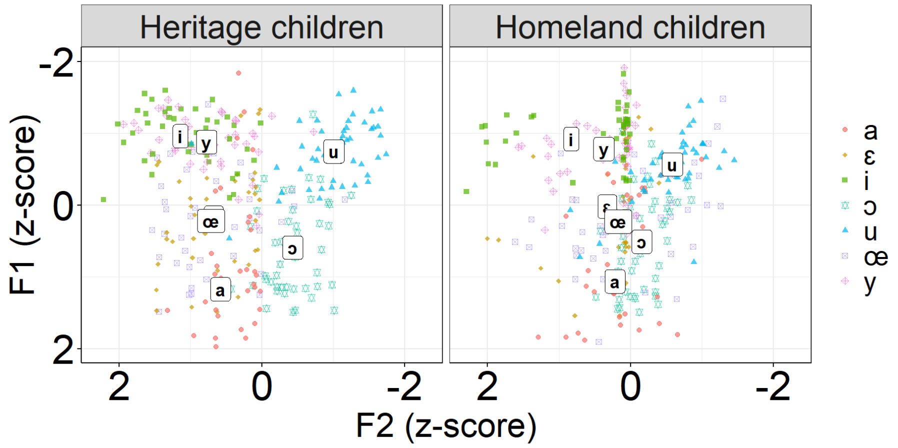


Figure 1. *F1-F2 plotting of the seven long vowels produced by heritage and homeland children aged between 2;0 and 3;11.*

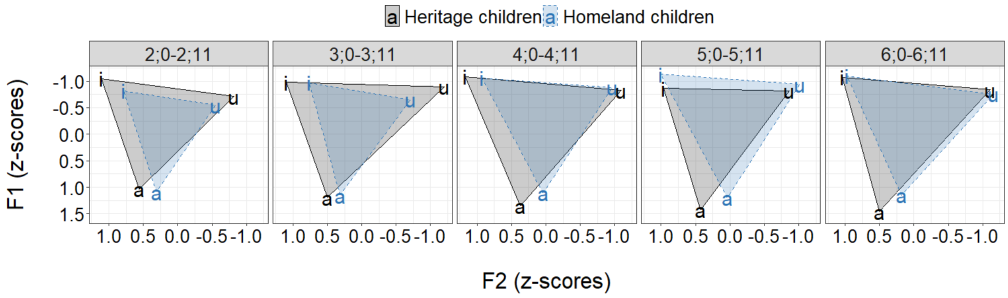


Figure 2. *Vowel space of heritage and homeland children of five age bands.*

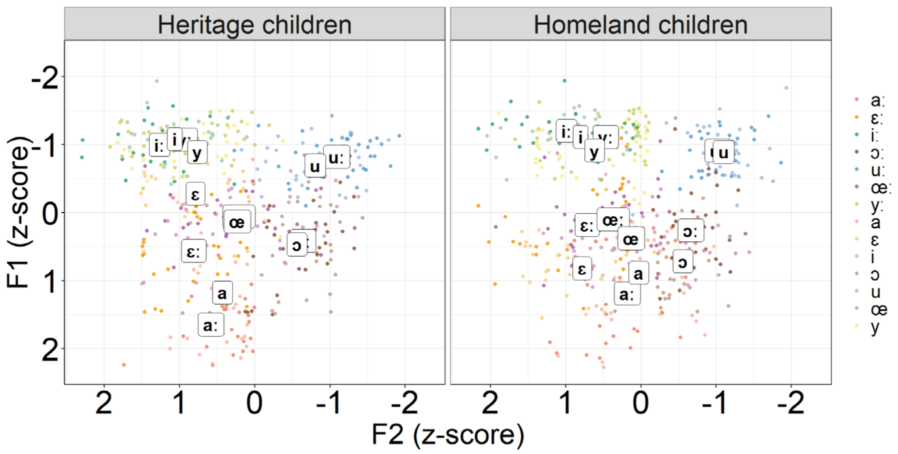


Figure 3. *F1-F2 position of the seven long vowels in open syllable and in closed syllables produced by heritage and homeland children from age 4;0.*

[1] Xu, Y. & Gao, H. 2018. *FormantPro as a tool for speech analysis and segmentation*. Revista de Estudos da Linguagem.

[2] Boersma, P. & Weenink, D. 2024. Praat: doing phonetics by computer [Computer program]. Version 6.4.04, retrieved from http://www.praat.org/.

[3] R Core Team. 2024. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Emerging intonation patterns in bilingual and monolingual child speakers of Australian English

Catalina Torres, Paola Escudero
University of Zurich, Western Sydney University

The acquisition of intonational phonology is one that starts early on but may require several years to fully master. Studies examining the speech prosody of monolingual children indicate that motor coordination and access to linguistic and cognitive resources could be adult-like starting from around seven years of age (Patel and Grigos 2006). Although prosodic structure in the speech of heritage language (HL) speakers tends to be an understudied topic (Ronquest & Rao 2018), there is evidence that HL children 9-years old can produce intonation contours typical of the acquired language varieties (Torres, 2024). However, vanishingly little is known about younger HL speaking children's intonation. In this study, we ask whether HL speaking children master the realisation of pitch accents and boundary tones in a comparable way to adult speakers. We focus on children (4-5 years) who are bilingual speakers of Spanish and Australian English (AE) together with a control group of monolingual peers in AE.

Australian English speakers use a variety of pitch accents and prosodic contours. Although the most common accent type is the simple H* other pitch accents such as !H*, L*, and L+H* are also attested in the literature (Fletcher & Loakes, 2018). The 'high-rising terminal' has often been claimed to be an intonation pattern which is particularly characteristic of Australian English. In this variety, these statement rises are realised mostly as low-onset high rises i.e. L* H-H% and they generally differ from rising intonation in yes-no questions which is more often realised as H* H-H% (Fletcher and Harrington, 2001).

The recordings used in the present study are part of a large corpus of child language speech in bilingual and monolingual children, as part of a longitudinal project on non-societal language maintenance and enhancement conducted in Australia (Escudero et al. 2024a/b). Data collection took place via Zoom with an experimenter conducting a session where children participated in the tasks with the help of a parent for session set up (Escudero et al. 2021).

In our acoustic analysis, we examine F0 peaks and contours of monosyllabic nouns. For this, twenty continuous points were extracted from target words. Further, we coded whether the selected part of speech was identified as a terminal word. We use generalised additive mixed models in our statistical investigation to examine the global F0 contour in the speech of five bilingual and three monolingual children.

Figure 1 visualises intonation contours from target words that were not in terminal position (no) and those that were (yes). Preliminary findings indicate that all children realise the commonly attested H* in pre-nuclear and nuclear positions. Interestingly, we find evidence for bilingual children showing high terminal rises whereas we don't find evidence for this pattern in the monolingual children. These observations suggest that children 4-5 years old can already master dialect typical intonation contours. Previous examination of our bilingual HL group showed that they realised vowels and consonants in a similar way to AE speaking adults (Diskin-Holdaway et al, 2024; Escudero et al, 2024c), indicating that bilingualism does not change their pattern of phonetic/phonological development.

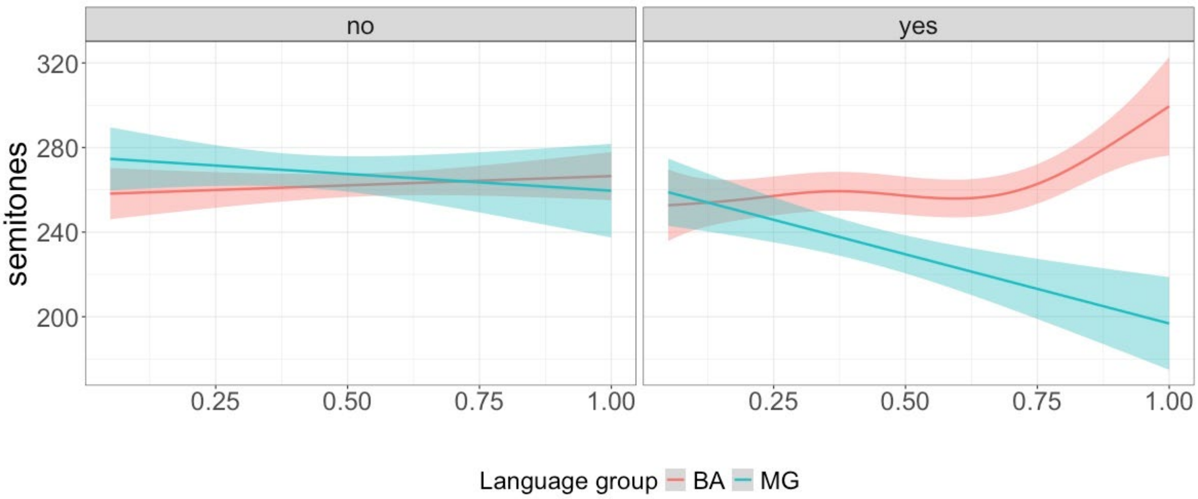


Figure 15 *Time normalised fundamental frequency contours in target words. 'No' denotes words that were not in terminal position, while 'yes' denotes words that were in terminal position. Monolingual children are in blue and bilingual children in red.*

Escudero, P., Escobar, G. P., Gallego, M. H., Diskin-Holdaway, C., & Hajek, J. (2024a). A database of multilingual child speech with recordings from a longitudinal project for multilingual education. In Proceedings of the Nineteenth Australasian International Conference on Speech Science and Technology (pp. 97-101).

Escudero, P., Escobar, G. P., Diskin-Holdaway, C., & Hajek, J. (2024b). Enhancing Heritage and Additional Language Learning in the Preschool Years: Longitudinal Implementation of the Little Multilingual Minds Program. OSF Preprints. September, 16, 20204.

Escudero, P., Li, W., & Diskin-Holdaway, C. (2024c). Have four-year-olds mastered vowel reduction in English? An acoustic analysis of bilingual and monolingual child storytelling. In Proceedings of the 19th Australasian International Conference on Speech Science and Technology, SST.

Escudero, P., Pino Escobar, G., Casey, C. G., & Sommer, K. (2021). Four-year-old's online versus face-to-face word learning via eBooks. Frontiers in psychology, 12, 610975.

Fletcher, J., & Harrington, J. (2001). High-rising terminals and fall-rise tunes in Australian English. Phonetica, 58(4), 215-229.

Fletcher, J., & Loakes, D. (2018). Pitch accent variation and realization in interactive discourse in Australian English. In Proceedings of the 17th Australasian International Conference on Speech Science and Technology (pp. 93-96).

Diskin-Holdaway, C., Li, W., & Escudero, P. (2024). Bilingual preschoolers' phonetic variation keeps up with monolingual peers: The case of voiceless plosives in Australian English. In Proceedings of the 19th Australasian International Conference on Speech Science and Technology, SST.

Ronquest, R., & Rao, R. (2018). Heritage Spanish phonetics and phonology. The Routledge handbook of Spanish as a heritage language, 164-177.

Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question–statement contrast in 4, 7 and 11 year-old children. Speech Communication, 48(10), 1308-1318.

Torres, C. (2024). Portuguese and German Intonation Contours in a Two-Way Immersion School. Languages, 9(2), 54.

Accent perception in bilingual children acquiring Italian

Paulina Speck[1], Tanja Kupisch[1,2], Svenja Krieger[1]
[1]University of Konstanz, [2]Lund University

Early bilingual language acquisition involves the challenge of developing two different phonological systems (e.g., Fabiano-Smith & Goldstein, 2010; Kehoe & Girardier, 2020). Some phonemes are common to both languages, some unique to one or distributed differently over the acoustic space. Focusing on the heritage language, this study seeks to answer whether bilingual children nevertheless develop perceptual systems akin to those of monolingual children or whether they 'perceive with an accent'?

Most previous studies on global foreign accent in heritage speakers have focused on adults and whether their speech sounds accented to monolingual listeners (e.g., Kupisch et al. 2014, Flores & Rato 2016). Only a few studies have investigated the perceived accent of children growing up bilingually (e.g., Wrembel et al. 2019, Kupisch et al. 2021, Kupisch et al. 2024, Laméris et al. 2024). Generally, heritage speakers sound less accented than late bilinguals, but more accented than monolingual speakers, and there is individual variation, largely determined by language exposure, but also age of onset in the majority language and age at testing. It is unclear whether these findings on production are mirrored in perception.

Our study is based on an accent rating task with 57 Italian-speaking children, of which 27 grew up monolingually in Italy (mean age 8.3, range 5-12), while 30 grew up bilingually in Germany (mean age 9.1, range 5-12). The children had to judge short speech samples (9-11 seconds) from 40 children between the age of 6 and 9 years with the same language background (20 monolingual, 20 bilingual). While listening to the speech samples, participants were presented with illustrations of boys and girls that aligned with the gender of the speakers. For each sample, they had to decide whether the child spoke like an Italian boy/girl or not.

The results show no difference between the monolingual and bilingual raters when judging the speech of monolingual children, but minor differences when they judge the speech of bilingual children (Figure 1). For the group comparison with respect to accent perception, we also compared how the two rater groups (bilingual and monolingual raters) perceived the accent of bilingual and monolingual speakers depending on their age (Figure 2). When listening to monolingual children, only age affected the ratings ($\beta = -0.16$, $SE = 0.07$, $z = -2.46$, $p < .05$), while there was no rater group difference ($\beta = -0.03$, $SE = 0.29$, $z = 0.01$, $p = .91$). In the case of bilingual speakers, both age ($\beta = 0.27$, $SE = 0.07$, $z = 3.86$, $p < .001$) and rater group affected accent perception ($\beta = 0.76$, $SE = 0.32$, $z = 2.38$, $p < .05$). Moreover, the bilingual child raters were significantly affected by the time they had spent in Italy ($\beta = 0.82$, $SE = 0.40$, $z = 2.04$, $p < .05$), with more time spent in Italy resulting in more perceived foreignness Overall, these results show very subtle differences in accent perception between monolingual and bilingual children, modulated by the time they spent in Italy.

Figure 1. *Frequency of perceived accent of monolingual and bilingual raters in Italian across monolingual and bilingual speakers.*

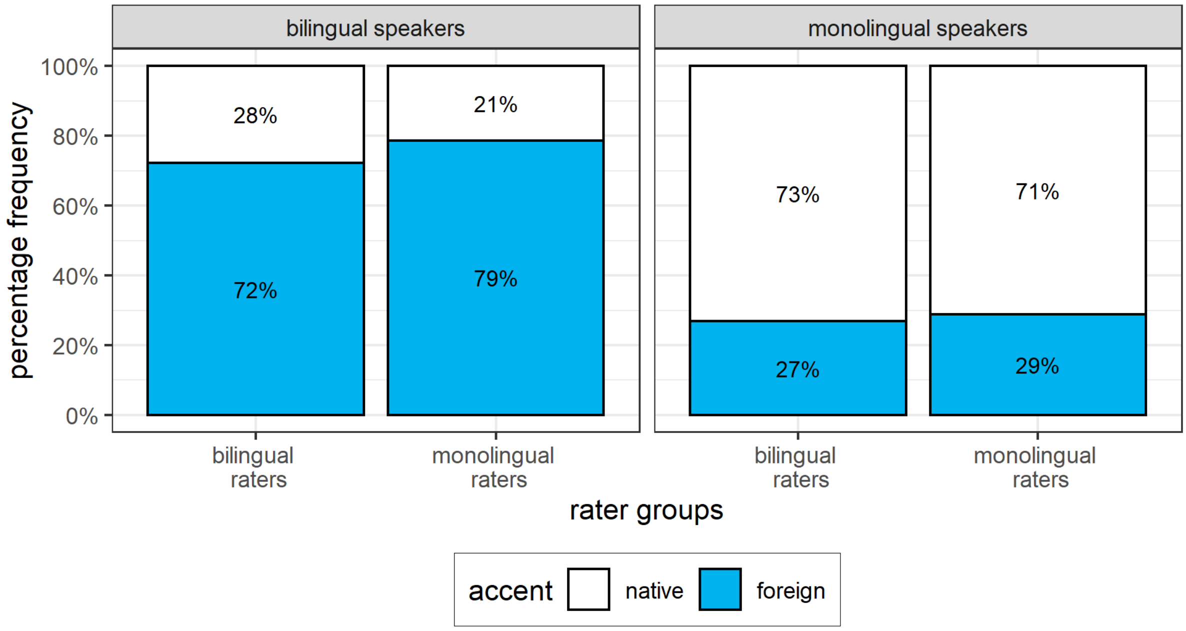
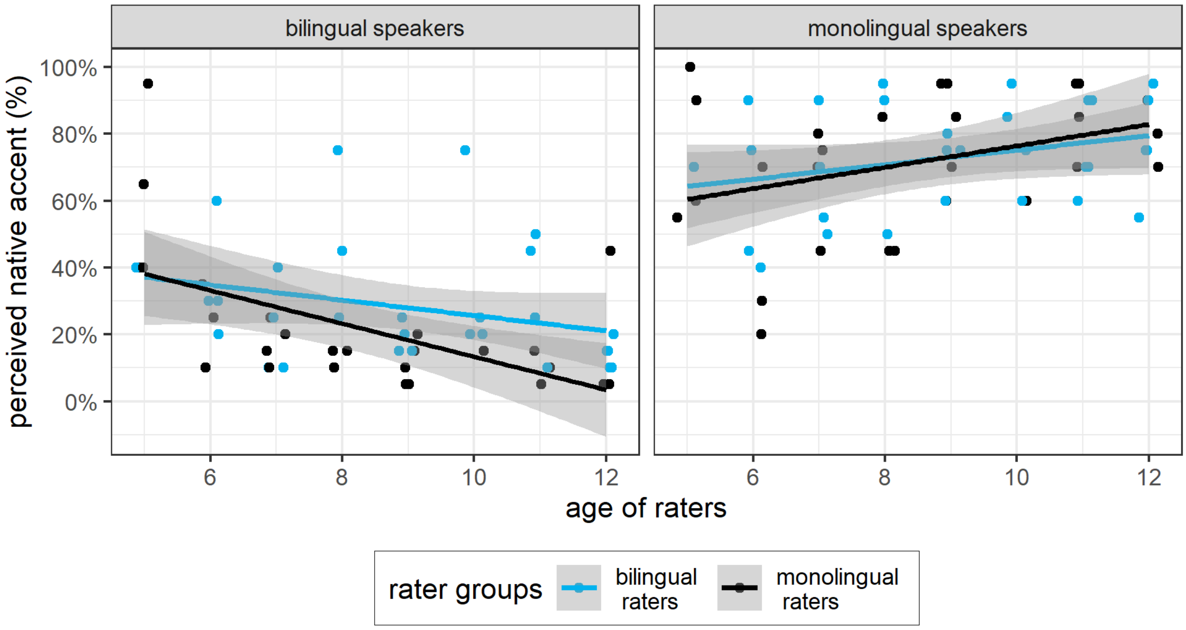


Figure 2. *Perceived native accent across rater and speaker groups taking into account the age of raters.*



Fabiano-Smith, L., & Goldstein, B. (2010). Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*, *53(1)*, 160–178.

Flores, C. & Rato, A. (2016). Global Accent in the Portuguese Speech of Heritage Returnees. *Heritage Language Journal*, *13(2)*, 161-183.

Kehoe, M., & Girardier, C. (2020). What factors influence phonological production in French speaking bilingual children, aged three to six years? *Journal of Child Language*, *47*, 945–981.

Kupisch, T., Barton, D., Hailer, K., Klaschik, E., Stangen, I., Lein, T., & Van de Weijer, J. (2014). Foreign Accent in Adult Simultaneous Bilinguals. *Heritage Language Journal*, *11(2)*, 123-150.

Kupisch, T., Kolb, N., Rodina, Y., & Urek, O. (2021). Foreign accent in pre- and primary school heritage bilinguals. *Languages*, *6(2)*, 96.

Kupisch, T., Canzi, M., Ferin, M., Geiss, M., Reiber, M., & Speck, P. (2024). How bilingual experience shapes accents at primary school age. *Heritage Language Journal, 21*, 1-30.

Laméris, T., Kubota, M., Kupisch, T., Cabrelli, J., Snape, N., & Rothman, J. (2024). Language change in Japanese-English bilingual returnee children over the course of five years: Evidence from accent-rating. *Second Language Research,*1-26.

Wrembel, M., Marecka, M., Szewczyk, J., & Otwinowska, A. (2019). The predictors of foreign-accentedness in the home language of Polish-English bilingual children. *Bilingualism: Language and Cognition, 22(2)*, 383-400.

Marked vs unmarked questions in heritage speakers of Italian

Valentina De Iacovo[1], Paolo Mairano[2], John Hajek[3]
[1]Università di Torino, [2]Université de Lille, [3]The University of Melbourne

Intonation conveys information at various linguistic and extralinguistic levels [1]. In Italian, as in other languages of the world, it also provides information about the speaker's regional origin [2]. Previous studies proved that L2 Italian speakers have more difficulty in recognizing certain more marked intonational patterns as interrogative (e.g., rising-falling patterns) compared to patterns closer to the L1 variety or patterns that are universally less marked for questions (i.e., rising or falling-rising patterns) [3]. In this study, we focus on heritage speakers (HS) of Italian, who have as their dominant language the language of the host country (in our case Australian English) [4]. We aim to test whether HS of Italian in Australia used different intonation patterns when speaking Italian and English, specifically, whether, in Italian, they will use the intonation pattern based on their familiar and/or geographical background in Italy. In this work, we created a production task of sentences in Italian and English based on two modalities (declarative vs interrogative) and we administered it to 20 HS of Italian. Our hypothesis is that HSs will show different intonation patterns (falling-rising vs rising-falling contour) based on the language employed (IT or EN). Results show that different intonation patterns are used, with a more homogeneous pattern in English and more variety in Italian.

[1] Fónagy, I. 1989. On status and functions of intonation, *Acta Linguistica Hungarica*, vol. 39, no. 1/4, pp. 53–92.
[2] Sorianello, P. 2011. *Prosodia: modelli e ricerca empirica*. Carocci Editore.
[3] De Iacovo, V., and Mairano, P. (2024). The effects of regional Italian prosodic variation on modality identification by L1 English learners. In *Speech Prosody* pp. 1–4.
[4] Benmamoun, E., Montrul, S. and Polinsky, M. "Heritage languages and their speakers: Opportunities and challenges for linguistics". Theoretical Linguistics, 39(3-4): 129-181, 2013. https://doi.org/10.1515/tl-2013-0009
[5] B. Gili Fivela, C. Avesani, M. Barone, G. Bocci, C. Crocco, M. D'Imperio, R. Giordano, G. Marotta, M. Savino, and P. Sorianello, "Intonational phonology of the regional varieties of italian," in Intonation in romance. Oxford University Press, 2015, pp. 140–197.

Cluster Analysis of Bulgarian Judeo-Spanish Calling Contours

Mitko Sabev[1], Bistra Andreeva[1], Christoph Gabriel[2], Jonas Grünke[3]
[1]Saarland University, [2]Johannes Gutenberg University Mainz, [3]University of Regensburg

In a production experiment, 17 bilinguals (6 female, mean age 87.33, SD 9.99; 11 male, aged 78.00 ±9.32) performed a production task in Judeo-Spanish (**JSp**) and Bulgarian (**BB**), along with 14 Monolingual Bulgarian (**MB**) speakers (10 female, aged 78.91 ±8.70; 4 male, aged 83.00 ±10.49). They were asked to call a girl named /kaˈlina/ from a distance in three contexts: neutral (for dinner), positive (to open her presents) and negative (for a telling-off). Fundamental frequency and vowel midpoint formant frequencies were measured in Praat. Hierarchical clustering was applied to identify distinct f_0-contour types, and an optimal number of clusters was determined based on information cost measures [1, 2]. A multivariate conditional inference tree was constructed to assess how CLUSTER (contour type) and DURATION vary with situational CONTEXT, linguistic VARIETY and GENDER [3]. Linear mixed effects models were used to compare F1 and F2 frequencies of initial- vs final-syllable /a/ vowels.

Cluster analysis and evaluation identified four optimal clusters, plotted in Fig. 1, and the tree in Fig. 2 shows that CLUSTER and DURATION are sensitive to all predictors – CONTEXT, GENDER and VARIETY. One contour dominates the negative context for all varieties and genders: cluster **A**, which starts with a slight rise into the stressed (second) syllable, followed by a steep and deep fall (H* L-L%). Neutral and positive contexts are not differentiated, but contours vary across genders and varieties. In all varieties, men use overwhelmingly contour **B**, an appreciable climb into the stressed syllable and then a smooth fall (L-H* !H-L%). Monolingual women use three contours in non-negative contexts: contour **B**, which is favoured by all men, along with the cross-linguistically well-attested calling tune **C** (L-H* !H-%), and contour **D**, a slow and rather shallow rise that doesn't peak until the final syllable (L+H* H-L%). In both their languages, bilingual women's first choice of non-negative contour is **B**, which is the one typical for men and one out of the three contours used by monolingual women. Intriguingly, the bilingual women's second choice is contour **A**, which is prevalent in negative contexts for everyone. Crucially, even though they may use the same contour shape in both negative and non-negative contexts, non-negative realisations have considerably longer durations, which substantially softens the call. One reason for bilingual women not employing more 'cheerful'-sounding contours in non-negative contexts may have to do with childhood stigma: several female Judeo-Spanish speakers pointed out—in separately recorded biographical interviews—that they were mocked by peers for their sing-song intonation.

Bulgarian is known for its pronounced and neutralising vowel reduction: when unstressed, the non-high vowels /a ɔ/ raise and merge with their higher counterparts, /ɤ u/; the process is categorical and not merely duration-driven [4]. After centuries of contact with Bulgarian, the local variety of Judeo-Spanish also exhibits appreciable reduction [5]. It has been observed that Bulgarian unstressed vowels resist reduction in the final syllables of calling phrases, which has been explained with the addition of an extra beat [6]. We report significantly higher F1 frequencies of /a/ in the final than in the initial syllables of all calling contours, confirming that a special post-lexical prominence is in place there, which causes unstressed vowels to behave like lexically stressed ones, retaining a degree of openness that no other unstressed vowels do.

Cluster analysis performed to objectively identify contour categories yielded different results from an analysis of a smaller data set based on auditory impression and f_0-span measurements [7]: clearer effects of context, gender and variety were revealed, and an additional contour was identified. There is little variation in the marked, negative context. Men show little variation in non-negative contexts as well, while women have more varied repertoires: in addition to men's default non-negative contour, monolingual women use two extra tunes, while bilingual women supplement the default contour with a lengthened version of the negative tune.

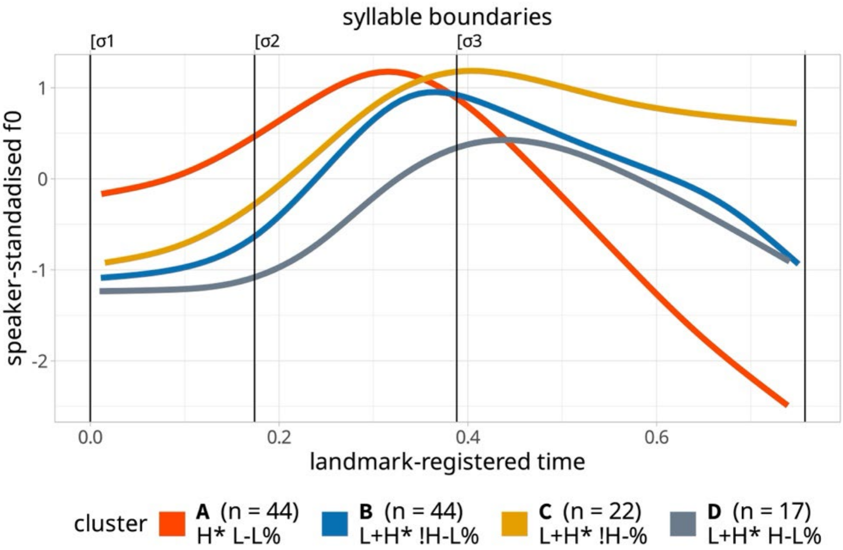


Figure 1. *Contour types identified by cluster analysis (GAM curves).*

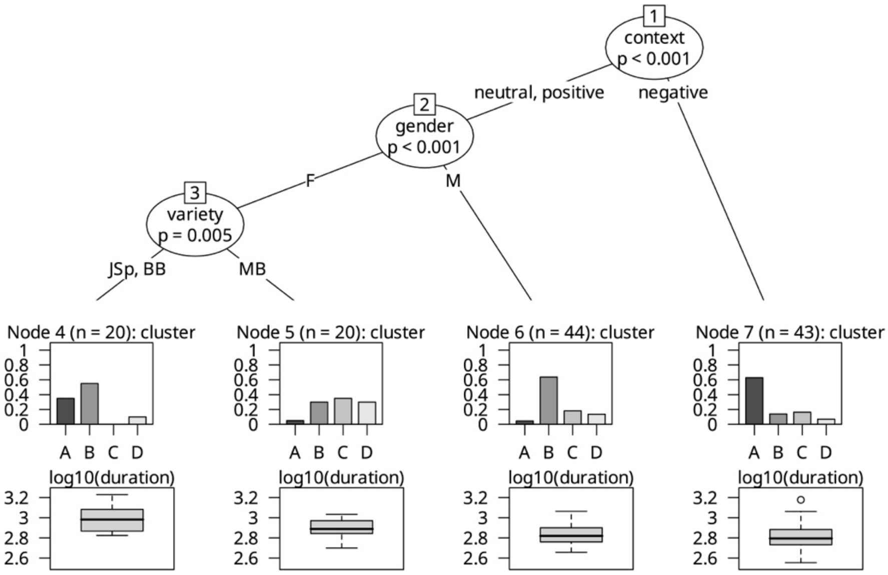


Figure 2. *Conditional inference tree: cluster + duration ~ context + variety + gender*

[1] Kaland, C. 2023. Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *JIPA* 53(1), 159–188.

[2] Kaland, C. & Ellison, T. M. 2023. Evaluating cluster analysis on f0 contours: an information theoretic approach on three languages. *Proc. 20th ICPhS*, 3448–3452.

[3] Hothorn, T. & Zeileis. A. 2015. partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research* 16, 3905–3909.

[4] Sabev, M. 2023. Unstressed vowel reduction and contrast neutralisation in western and eastern Bulgarian: A current appraisal. *JPhon* 99, 101242.

[5] Grünke, J., Sabev, M., Gabriel, C. & Andreeva, B. 2023. Vowel reduction in spontaneous Bulgarian Judeo-Spanish. *Proc. 20th ICPhS*, 2820–2824

[6] Andreeva, B. & Dimitrova, S. 2023. Vocative Intonation in Bulgarian. *Proc. 2nd Intl. Conf. on Tone and Intonation*, 29–33.

[7] Grünke, J., Andreeva, B., Gabriel, C. & Sabev, M. 2023. Vocative Intonation in Language Contact: The Case of Bulgarian Judeo-Spanish. *Languages* 8(4), 284.

Workshop: Moving beyond the means: Instability of acoustic-phonetic cues marking segmental contrasts

Acoustic instability is an outcome of interaction between coarticulatory resistance and contrast: Insights from Assamese, Malayalam, and Mongolian

Indranil Dutta
Jadavpur University

Interaction between coarticulation and phonological contrast is essential for modeling speech production. In this talke, we examine coarticulatory resistance across three typologically diverse languages - Assamese, Malayalam, and Khalkha Mongolian - to explore how articulatory constraints shape phonological systems. Using acoustic analyses and eye-movement data, we investigate how coarticulatory information is preserved, resisted, or exploited in speech processing.

In Assamese, a language where the dental-retroflex contrast has historically merged into an alveolar stop, locus equation slopes indicate that alveolar stops exhibit greater coarticulatory resistance than labials and velars (Thakuria et al., 2023). This challenges the expectation that a sparse contrast inventory allows for increased coarticulation (Recasens, 1984). Instead, the data suggest that articulatory complexity and motor constraints play a more dominant role in determining resistance patterns (Dutta et al., 2023).

Malayalam, with its dense coronal inventory, provides a compelling test case for coarticulatory effects on lexical access. An eye-tracking study on coronal stop contrasts (dental, alveolar, retroflex) demonstrates that mismatched coarticulatory cues hinder lexical retrieval, particularly when segments with greater coarticulatory resistance are involved (Seema et al., 2023). This aligns with previous work showing that coarticulatory consistency enhances lexical access efficiency (Beddor et al., 2013; Fowler, 1984). Additionally, Malayalam coronal stops show differential patterns of resistance in VC and CV transitions, reinforcing findings that coarticulatory resistance varies with phonetic context (Dutta et al., 2019; Recasens & Espinosa, 2009).

Khalkha Mongolian presents an intriguing case of vowel harmony and coarticulatory resistance. While vowel harmony operates in a left-to-right (carryover) direction, vowels in non-harmonic domains exhibit greater anticipatory coarticulation, counteracting the harmony process (Mitra et al., 2023). This suggests that coarticulatory resistance serves as a stabilizing force in maintaining phonological distinctions.

Taken together, these findings contribute to a broader understanding of how phonetic constraints shape phonological structure. The cross-linguistic perspective underscores the role of coarticulatory resistance in maintaining phonemic contrast, supporting models that integrate articulatory constraints and perceptual dynamic.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America, 133*(4), 2350–2366.

Dutta, I., Redmon, C., Krishnaswamy, M., Chandran, S., & Raj, N. (2019). Articulatory complexity and lexical contrast density in models of coronal coarticulation in Malayalam. *Proceedings of the 19th International Congress of Phonetic Sciences*.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics, 36*(4), 359–368.

Mitra, A., Krishnaswamy, M., & Dutta, I. (2023). Coarticulation and contrast in a vowel harmony system: Coarticulatory propensity in Khalkha Mongolian V-C-V sequences. *Proceedings of ICPhS 2023*.

Ohala, J. J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. *Third International Conference on Spoken Language Processing*.

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America, 76*, 1624–1635.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America, 125*, 2288–2298.

Seema, G. P., Krishnaswamy, M., Mishra, R., & Dutta, I. (2023). Mismatched coarticulatory information hinders lexical access of coronal stops in Malayalam. *Proceedings of ICPhS 2023*.

Svantesson, J.-O., Tsendina, A., Karlsson, A., & Franzén, V. (2005). *The phonology of Mongolian*. Oxford University Press.

Thakuria, R., Gogoi, P., & Dutta, I. (2023). Acoustic analysis of alveolar stop coarticulation in Assamese. *Proceedings of ICPhS 2023*.

Intensity and f_0 enhances phonological length contrast in Italian

Sankhadeep Ghosh, Indranil Dutta
Jadavpur University

Previous studies have shown that while duration is the primary cue for phonological quantity contrasts (Zmarich et al., 2011), acoustic features, such as amplitude, fundamental frequency (f_0), and voice quality of the surrounding vowels, often co-vary with duration (Payne, 2006).In Italian, duration plays a key role in predicting phonological length (Esposito & Di Benedetto, 1999; Benedetto & De Nardis, 2021b,a).

In our study, we corroborate the findings on duration in Italian based on data from the GEMMA corpus (Di Benedetto & De Nardis, 2022). However, we note several instances where singleton and geminate consonants exhibit overlapping durations. Specifically, singletons with durations comparable to those of geminates, and vice-versa are found in the the GEMMA corpus. In this study, we investigate how phonological length is cued when duration ceases to be the dominant cue in Italian. We find that in such cases, amplitude becomes the most crucial cue, followed by fundamental frequency, in the acoustic manifestation of phonological length irrespective of the vowel environments.

GEMMA corpus of Italian language consists of singleton and geminate pair of five manners of artic- ulation categories, namely stops (/p/, /t/, /k/, /b/, /d/ and /g/), affricates (/dz/, /Ã/, /ts/ and /Ù/), fricatives (/f/, /v/, /s/, /z/ and /S/), liquids (/l/, /r/ and /ń/), nasals (/m/, /n/, /ñ/) and glides (/j/ and /w/) in a symmetric intervocalic environment of vowels /a/, /i/ and /u/. So, the template for every instance recorded was of V_1 C(:)V_2, where V_1 and V_2 are the same vowel. This data was collected from six speakers, three males and three females. For this study we excluded the segment which have only singleton (/z/) or only geminate instances (/S/, /ń/ and /ñ/). Likewise, we also excluded the glides (/j/ and /w/) since they do not exhibit contrastive length.

Out of 1,832 samples, we found 392 samples that exhibit overlapping durations (See Figure 1). Four instances of overlapping nasal consonants (two instances for each /m/ and /n/) were excluded due to their very low occurrence rate. The remaining 388 instances were used as the dataset for the present study. The distribution of manners of articulation within this dataset is as follows: 67 instances of stops, 98 fricatives, 176 affricates, and 47 of liquids.

Our findings show that the V_1 nucleus intensity of affricates is higher in geminates than in their singleton counterparts ($p = 0.03441$). A similar pattern is observed for liquids, where the intensity at V_1 offset is significantly higher for geminates ($p = 0.0412$). These results align with the findings of Arora & Dutta (2011), who observed that perceptually longer sounds exhibit greater amplitude. However, for stops, a different trend is observed: the V_1 offset intensity is significantly greater for singletons than for geminates ($p = 0.02199$), particularly within the durationally overlapping range (see Figure 2). For overlapping fricative segments, intensity is not a significant cue for phonological length distinction.

f_0 in the following vowel is lower in singletons compared to their geminate counterparts. Figure 3 shows the results of the f_0 at the onset (10%) and nucleus (50%) of V_2. The p-values for stops at these two points are 0.026 and 0.0003, respectively, indicating statistically significant difference.

We argue that in the absence of the primary acoustic cue for phonological length contrast − duration, as seen in Italian − secondary cues, such as intensity, display an asymmetry across different manners of articulation (e.g., affricates, liquids, vs. stops), thus hindering the ability to draw a generalizable conclusion. Additionally, as can be seen in the density plots on the left of Figure 1, geminates exhibit greater durational variation relative to singletons. This implies greater instability of duration as a sole acoustic cue for marking phonological contrast.

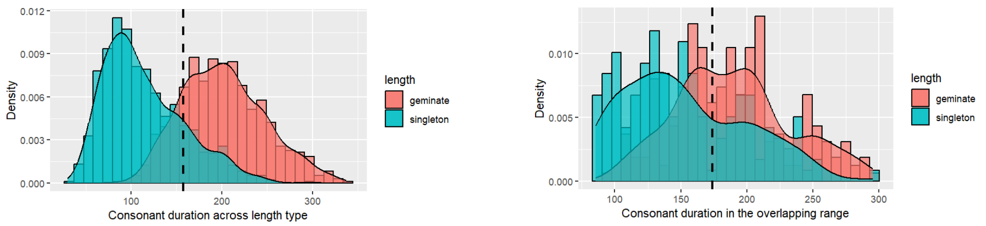


Figure 1: *Overlapping duration: Full dataset of 1832 samples (left) and Overlapping dataset of 392 samples (right)*

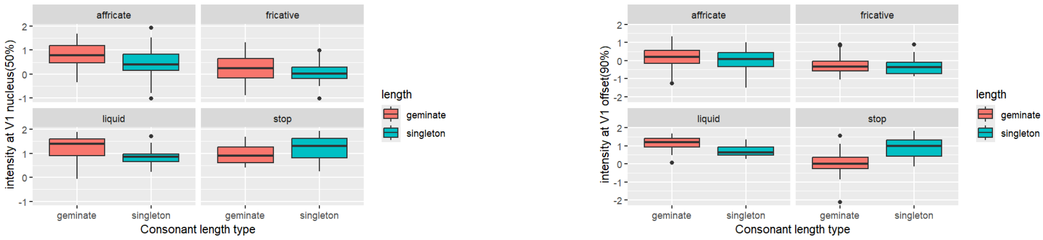


Figure 2: *Intensity at the nucleus and at the offset of V_1 for different manners of articulation*

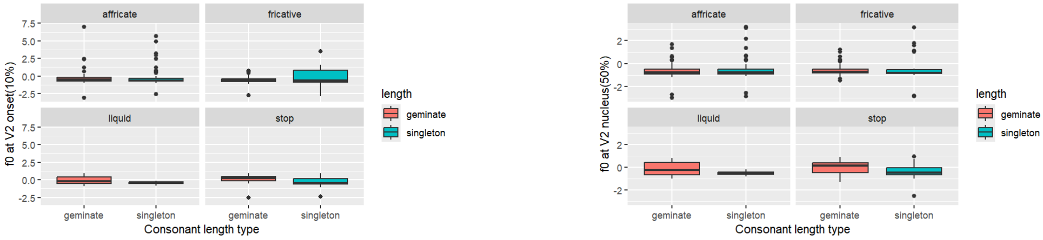


Figure 3*: f_0 at the onset and at the nucleus of V_2 for different manners of articulation*

Arora, A., & Dutta, I. (2011). Louder is longer: Amplitude conditioned lengthening in diasporic Siraiki. *The Journal of the Acoustical Society of America*, *130* (4).

Benedetto, M. G. D. & De Nardis, L. (2021a). Consonant gemination in italian: The affricate and fricative case. *Speech Communication*, *134*, 86-108. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167639321000753

Benedetto, M. G. D. & De Nardis, L. (2021b). Consonant gemination in italian: The nasal and liquid case. *Speech Communication*, *133*, 62-80. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167639321000765

Di Benedetto, M.-G., & De Nardis, L. (2022). The gemma speech database: Vcv and vccv words for the acoustic analysis of consonants and lexical gemination in Italian. *Data in Brief*, *43*, 108373. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352340922005704

Esposito, A., & Di Benedetto, M.-G. (1999, 11). Acoustical and perceptual study of gemination in italian stops. *The Journal of the Acoustical Society of America*, *106*, 2051-62.

Payne, E. (2006, 06). Non-durational indices in italian geminate consonants. *Journal of the International Phonetic Association*, *36*, 83 - 95.

Zmarich, C., Gili Fivela, B., Perrier, P., Savariaux, C., & Tisato, G. (2011, 08). Speech timing organi- zation for the phonological length contrast in Italian consonants. In (p. 401-404).

Perceptual correlates of lexical tone in Brajbhasha

Neetesh Kumar Ojha, Shakuntala Mahanta
Indian Institute of Technology, Guwahati, Assam, India

Brajbhasha, a northcentral Indo-Aryan language, exhibits traces of tonogenesis—historically attributed to the loss of the glottal fricative /h/ at various locations in the syllables of the NIA cognates—leading to a low-tone realization [1]. The production experiments had shown the existence of 2 tones, "Low" and "Neutral" [1]. The current study investigates the perception of lexical tones through a controlled experimental setup using a series of pitch-manipulated audio stimuli (words in isolation) designed to potentially simulate pitch contours corresponding to various types of low and/or neutral tones. This research comprises two experiments conducted on 182 participants aged 15–55 who categorized 144 auditory stimuli (consisting of 12 initially neutral words with 11 pitch-manipulated versions of each; see Figure 1) into "Neutral," "Low," or "Unsure" responses through a categorical perception test employing the 2AFC method [2][3]. Experiment 1 examines the interaction between the location of /h/ deletion in the NIA cognate and the type of vowel lowering of the stimulus in determining the participants' binary responses ('Low' vs 'Other'). Experiment 2 explores the role of demographic variables, including age group and access to education, in lexical tone perception. Results from the generalized linear mixed-effects [4] and multinomial logistic regression [5] models performed in RStudio [6], wherein we examined the probability of perceiving a "Low" tone, reveal significant findings to understand the underlying processes of tone perception, providing new insights into its dynamics in Brajbhasha.

One common aspect in the emergence of low/falling tones in tonal languages like Chinese [7], Vietnamese [8], and Tai [9] is the loss of the final /h/; however, the loss of /h/ in the current study is more comparable to that of Punjabi [10][11]. The current study confirms that Brajbhasha tonogenesis due to the loss of /h/ parallels that of Punjabi more substantially than others. Experiment 1 (see Figure 2) in the current perception study confirms that the loss of intervocalic and coda /h/ in NIA is responsible for the 'falling' contour as a separate tonal category in Brajbhasha. Further, experiment 2 (see Figure 3) reveals evidence of generational tone loss (or tonoexodus) in Brajbhasha due to the younger generation's access to education, i.e., contact with standard Hindi-Urdu and/or English that have a dominant status in society—similar to the tone loss in Deori [12] due to "bilingual contact" with a dominant non-tonal Assamese language.

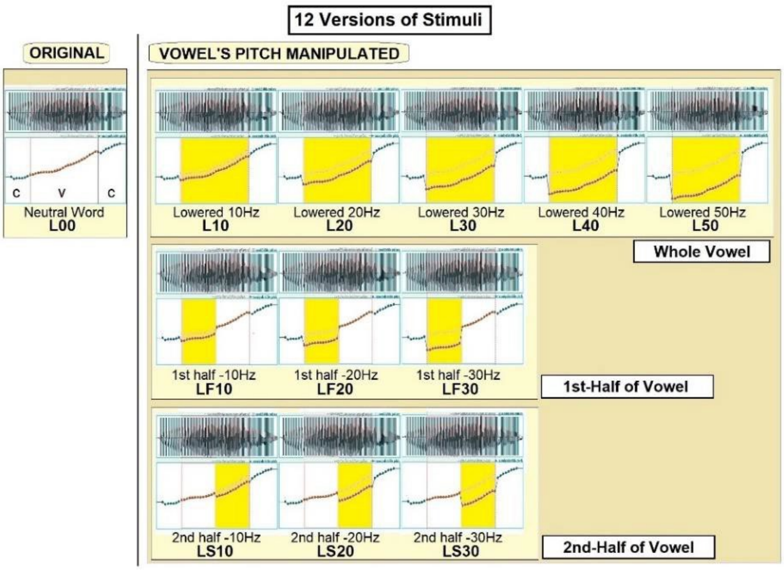


Figure 1. *Preparation of the audio stimuli per initially neutral word.*

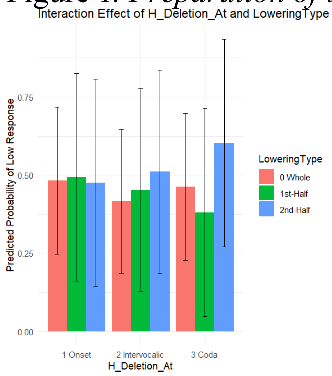


Figure 2. *Connection between location of /h/ deletion in NIA cognate and type of vowel lowering in audio stimulus.*

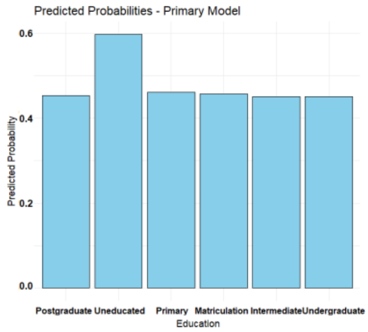


Figure 3. *Impact of access to education on perception of low tone by the participants.*

[1] Ojha, N.K. and Mahanta, S. (2023) 'Vowels, tones, and tonogenesis in Braj', in *Proc. TAI 2023*. pp. 73–77. https://doi.org/10.21437/TAI.2023-16

[2] Jogan, M. and Stocker, A.A. (2014) 'A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability', *Journal of Vision*, 14(3), pp. 20–20.

[3] Jenadeleh, M., Zagermann, J., Reiterer, H., Reips, U.D., Hamzaoui, R. and Saupe, D. (2023) 'Relaxed forced choice improves performance of visual quality assessment methods', in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, pp. 37–42.

[4] McCullagh, P. (2019) *Generalized linear models*. Routledge.

[5] Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. and Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R*. New York: Springer.

[6] R Core Team (2024) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

[7] Haudricourt, A.G. (1954) 'Comment reconstruire le chinois archaïque', *Word*, 10(2–3), pp. 351–364.

[8] Haudricourt, A.G. (1954) *De l'origine des tons en vietnamien*. Imprimerie Nationale.

[9] Hanbo, L. (2016) *Tonal development of Tai languages*. MA thesis. Chiang Mai: Payap University.

[10] Lata, S. and Arora, S. (2013) 'Laryngeal tonal characteristics of Punjabi—an experimental study', in *2013 International Conference on Human Computer Interactions (ICHCI)*. IEEE, pp. 1–6.

[11] Malik, M.A. and Kokub, I. (2020) 'Segmental study of Punjabi glottal fricative /H/: An acoustic analysis of glottal fricative in Majhi and Lehndi dialects of Punjabi'.

[12] Acharyya, P. and Mahanta, S. (2018) 'Production and perception of lexical tone in Deori', in *Sixth International Symposium on Tonal Aspects of Languages (TAL)*. pp. 93–97.

The instability of preaspiration as a cue to voicelessness: evidence from Burmese and Italian

Chiara Repetti-Ludlow
Carnegie Mellon University

Introduction. Preaspirated segments are cross-linguistically rare, a fact which some researchers argue is because their low acoustic salience leads to diachronic instability [1, 2]. However, it remains unclear whether preaspiration is actually acoustically instable, and if so, what conditions this instability. Exploring this topic is complicated by the fact that preaspiration is often contemporaneous with preceding vowels [3], meaning binary classifications of preaspiration are not always useful. To address this issue, the present study explores acoustic cues of aspiration (H1*-H2* and CPP), following previous on the analysis of aspiration/breathiness concurrent with vowels [4]. The acoustic cues are measured over the preceding vowel in two test cases: Burmese sonorant voicing and Italian stop voicing, both of which are known to involve a period of [5, 6]. Variation in these cues are examined with different conditioning factors, including gender, age, and prosody, with the goal of determining whether preaspiration is instable, and if so, what conditions this instability.

Methods. Production studies were carried out with 9 speakers of Burmese and 25 speakers of Florentine Italian. Speakers of each language were balanced for gender and age cohort (younger = < 30; older = >30). Participants read target words which contained voiced and voiceless sonorants in Burmese (N = 1686, e.g. [l̥a] 'sword') or stops in Italian (N = 5389, e.g. [di**t**a] 'fingers') in phrase-medial or phrase-final position. Data was then annotated in Praat, and aspiration cues (H1*-H2* and CPP) were measured over the last fifth of the preceding vowel. Thus, aspiration was not measured with a binary classification, but instead, cues related to aspiration were analyzed over the vowel. Finally, linear mixed-effects models were carried out in R to examine how these cues were related to voicing and whether there was an interaction with phrasal position, gender, and age cohort. In cases where there were significant interactions, post-hoc tests were carried out to determine whether the interaction resulted in voicing neutralization.

Results. In Burmese, results for H1*-H2* show an interaction between voicing and gender, such that men have significantly different H1*-H2* based on voicing, but women don't. There is also an interaction between voicing and age cohort, such that older participants have significantly different H1*-H2* based on voicing category, but younger participants don't. For CPP, there is an interaction between voicing and age cohort, such that older participants have significantly different CPP based on voicing category, but younger participants don't. These results are illustrated in Figure 1. Turning to Italian, results for H1*-H2* demonstrate a main effect of voicing (higher H1*-H2* for voiceless) and gender (higher for men), while results for CPP demonstrate main effects of voicing (lower for voiceless) and position (lower finally).

Discussion. Results indicate that in Burmese, H1*-H2* and CPP are instable as acoustic cues. While they are significantly correlated with voicing for the older cohort, this is neutralized for the younger cohort. Furthermore, while H1*-H2* was correlated with voicing for men, it was not for women. This could indicate that Burmese is undergoing a change, whereby acoustic cues to aspiration are not produced concurrently with the vowel, and women might be at the forefront of this change, as has been found in other sound changes [7]. Meanwhile, Italian demonstrates relative consistency with these acoustic cues, such that H1*- H2* and CPP are both correlated with voicing, but this does not vary based on conditioning factors. Taken together, these results demonstrate that cues aren't de facto instable, as the same cues that seem to be instable on the vowel in Burmese are robust in Italian. Instead, results suggest that when cues become less stable, it provides evidence of language change, which is conditioned by social factors.

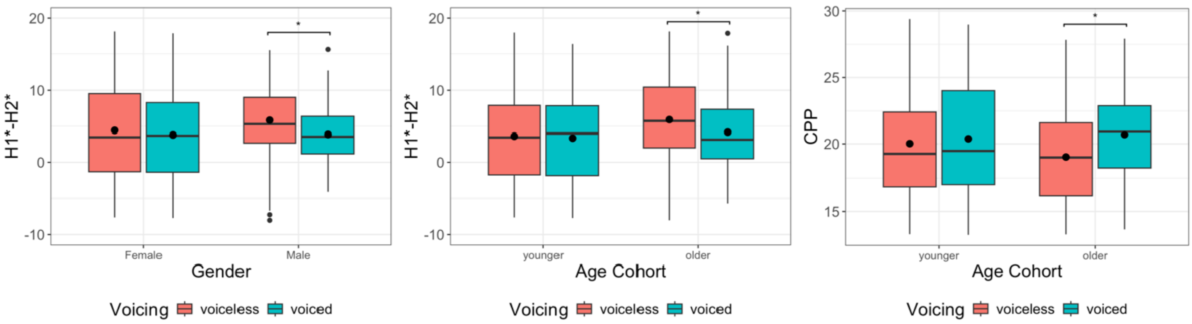


Figure 1. *Burmese results showing interactions between voicing and gender for H1*-H2*, and voicing and age cohort for H1*-H2* and CPP.*

[1] Blevins, J. 2018. Evolutionary phonology and the life cycle of voiceless sonorants. *Typological hierarchies in synchrony and diachrony*, 29-60.

[2] Silverman, D. 2003. On the rarity of pre-aspirated stops. *Journal of linguistics* 39(3), 575-598.

[3] Gordeeva, O., & Scobbie, J. 2010. Preaspiration as a correlate of word-final voice in Scottish English fricatives. *Turbulent sounds: An interdisciplinary guide* 21, 167-207.

[4] Berkson, Kelly Harper. 2019. Acoustic correlates of breathy sonorants in Marathi. Journal of Phonetics 73, 70–90.

[5] Bhaskararao, P., & Ladefoged, P. 1991. Two types of voiceless nasals. *Journal of the International Phonetic Association* 21(2), 80-88.

[6] Stevens, M., & Alvord, S. M. 2011. Consonant length in Italian: Gemination, degemination and preaspiration. In *Selected proceedings of the 5th conference on laboratory approaches to Romance phonology*, 21-32.

[7] Labov, W. 1990. The intersection of sex and social class in the course of linguistic change. *Language variation and change* 2(2), 205-254.

Tenseness in Jiuhe Bai

Meihao Wan, Peggy Mok
The Chinese University of Hong Kong

Introduction: Bai is a Sino-Tibetan language primarily spoken in Yunnan Province, China. Traditionally, it is considered to have a tense vs. lax register contrast in its tone system [1]. The most well-studied phonetic features distinguishing this register contrast in Bai are pitch and phonation type: tense tones are accompanied by nonmodal phonation which varies across speakers [2, 3]. However, previous studies have not observed differences in vowel quality which also plays an important role in the register contrast of other Sino-Tibetan languages, such as Hani [4] and Yi [5]. This study aims to investigate the phonetic cues of the register contrast in the production of Bai comprehensively.

Method: All data were collected in Jiuhe Township which belongs to Lijiang City, Yunnan Province. In Jiuhe Bai, T1, T3, T5 are tense tones; T2, T4, T6 are lax tones. 10 native Bai speakers participated in the experiments, all of whom were born and raised in Jiuhe. The stimuli consisted of 14 monosyllabic minimal pairs or near-minimal pairs of the 6 lexical tones. Participants were asked to produce them three times in a random order. Both acoustic and electroglottograph (EGG) signals were recorded. Acoustic analysis was conducted with VoiceSauce [6], extracting the relevant parameters for pitch, phonation and vowel quality (H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, CPP, HNR, Energy, SHR, f0, F1, F2).

Contact quotient (CQ) and speed quotient (SQ) were extracted from EGG signals using a MATLAB script. CQ refers to the ratio of the duration of vocal fold contact to the total duration of one vocal fold cycle. SQ is defined as the ratio of the contacting duration to the decontacting duration. A higher CQ indicates greater glottal constriction, indicating creaky and tense phonation, while a lower CQ means breathier phonation.

Results: Figure 1 shows the z-scored pitch contours of the tones in Jiuhe Bai using Generalized Additive Mixed Models (GAMM). T1 and T2 share similar pitch contours, while T3 has a sharper slope. T4 and T5 have the same ending, but the initial pitch of T5 is higher than that of T4. Figures 2 and 3 illustrate the CQ and SQ contours of different registers analyzed with GAMM. The CQ contour of tense syllables is lower than that of lax syllables, while the SQ contour of tense syllables is higher, suggesting a breathier phonation in tense syllables. Figure 4 shows the formants of the tense and lax vowels at vowel midpoint. The vowels with tense tones have higher F1 values than those with lax tones. A principal component analysis (PCA) was conducted to reduce the dimensionality of the spectral parameters of phonation. The first two PCs account for 59.1% of the variance in the data (PC1 = 30.7%, PC2 = 28.4%, as shown in Figure 5). The factor correlation loading for PC1 and PC2 are displayed in Figure 6. PC1 is mostly correlated with individual harmonics, such as H4*(0.40), A1*(0.38), H2*(0.38), H1*(0.36), PC2 is mostly correlated with periodicity and noise, such as HNR15(-0.36), HNR25(-0.36), HNR35(-0.32), and related to spectral tilt measures, like H1-A3* (0.34), H1-A1*(0.32). A mixed-effect logistic regression model was employed to evaluate the contribution of different cues to the register contrast. Table 1 shows that the register contrast is significantly related to all three dimensions: f0, F1 and PC1, which pertains to phonation. Based on the magnitude of the coefficients, F1 plays the most important role in distinguishing the contrast in Bai, followed closely by f0. Phonation plays a less important role than the two dimensions.

Discussion: In this study, the register contrast in Jiuhe Bai exhibits differences in all three dimensions: pitch, vowel quality and phonation. Tense vowels have higher F1 values than those of lax vowels, suggesting a retraction of the tongue root in the language. This underscores the instability of acoustic cues in encoding linguistics contrasts and provides insights into the interaction of multiple dimensions in the register contrast.

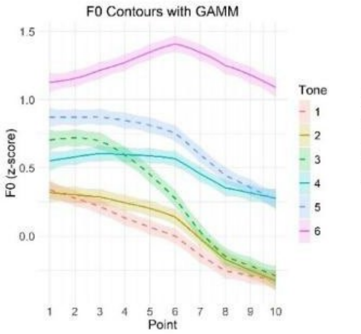


Figure 1. *F0 contours of lexical tones in Jiuhe Bai*

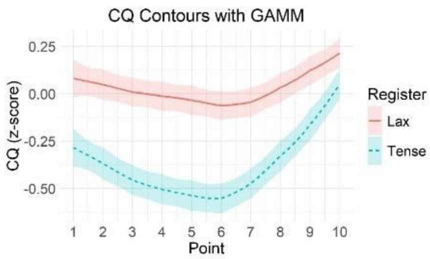


Figure 2: *CQ contours of different registers in Jiuhe Bai*

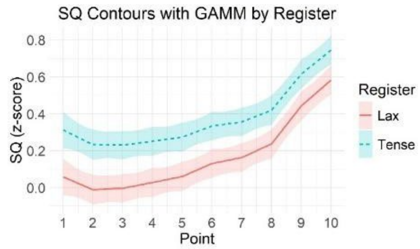


Figure 3: *SQ contours of different registers in Jiuhe Bai*

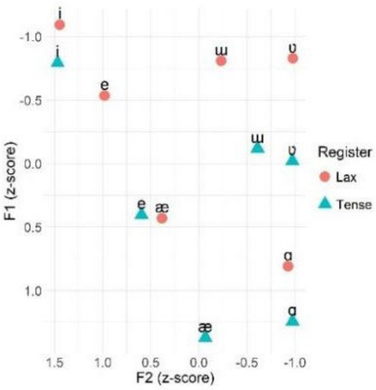


Figure 4: *Mean vowel formant frequencies of the register contrast in Jiuhe Bai*

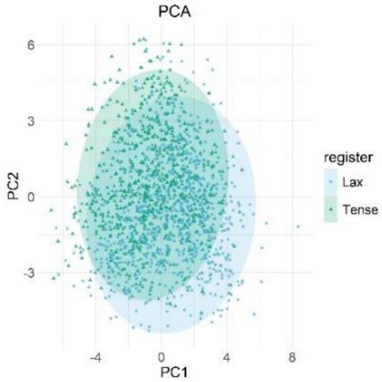


Figure 5: *PCA of the acoustic space*

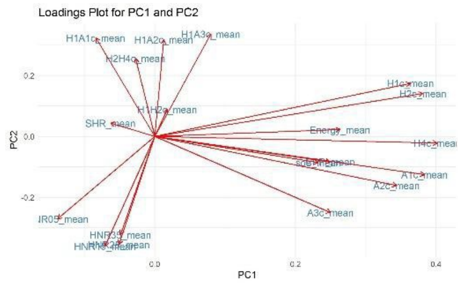


Figure 6: *The loadings for PC1 and PC*

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.02509	0.05592	-0.449	0.6536
PC1	-0.30248	0.07150	-4.231	2.33e-05***
PC2	0.01988	0.11101	0.179	0.8579
sF1_mean	0.50164	0.11023	4.551	5.34e-06***
sF2_mean	-0.09893	0.05902	-1.676	0.0937
strF0_mean	-0.46155	0.07881	-5.857	4.72e-09***
Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Table 1: *Results of mixed-effects logistic*

[1] Xu, L., & Zhao, Y. 1984. *An introduction to Bai language*. Beijing: The Ethnic Publishing House.

[2] Wang, F. 2015. Variations of laryngeal features in Jianchuan Bai. *Journal of Chinese Linguistics*, 43(1), 434-452.
[3] Li, X., & Wang, F. 2020. Phonation variation and strategy of tone distinction: the case of Meiba Bai. *Journal of Chinese Linguistics*, 48(2), 379–401.
[4] Maddieson, I., & Ladefoged, P. 1985. "Tense" and "lax" in four minority languages of China. *Journal of Phonetics*, 13(4), 433-454.
[5] Kuang, J., & Cui, A. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics*, 71, 194-214.
[6] Shue, Y.-L., P. Keating, C. Vicenik, K. Yu. 2011. VoiceSauce: A program for voice analysis, *Proceedings of the ICPhS XVII*, 1846-1849.

Vocalic signals are worth strengthening, but consonants may not be

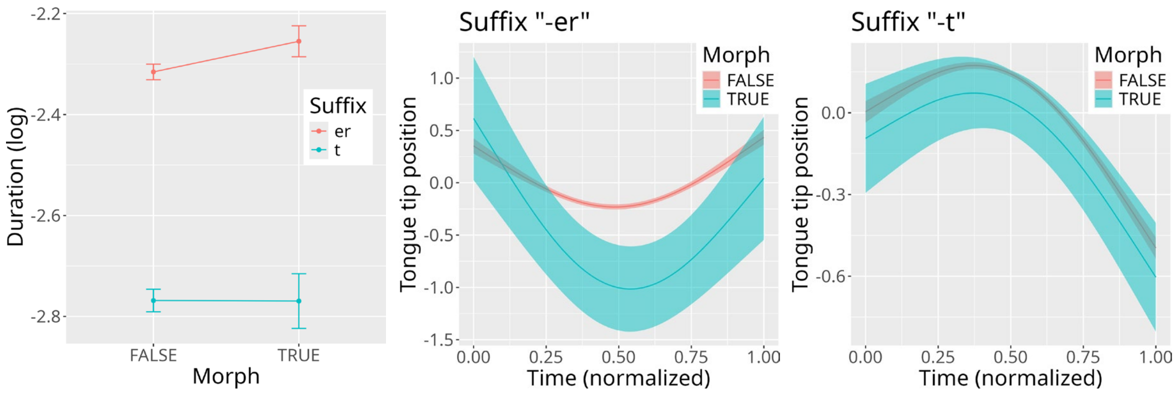
Motoki Saito[1], Jessie S. Nixon[2]
[1]Eberhard-Karls-Universität Tübingen, [2]Universität Bielefeld

Acoustic cues within phonemic units can vary systematically, according to a variety of factors (reviewed in Nixon and Tomaschek, 2023). For example, affixes have been reported to be associated mostly with clearer subphonemic realizations (e.g., Smith, Baker, and Hawkins, 2012; Strycharczuk and Scobbie, 2016), compared to the non-morphological counterparts of the affixes (e.g., *free**s*** [fri:z] vs. *freez**e*** [fri:z]), presumably because they carry their own meanings (e.g., Hay, 2007). However, the opposite effects of morphology on subphonemic realizations, namely shorter duration of suffixes compared to pseudo-suffixes, have also been reported (e.g., Song et al., 2013). We propose a possible confound that may explain this discrepancy: namely, the vocality of the segments under investigation. A difference in the production of a vowel can lead to clearer speech, compared to a difference in a (stop) consonant, given the same level of effort.

To test this hypothesis, German suffixes *-t* [-t] (e.g., *mach**t*** [maxt] "makes") and *-er* [-5] (e.g., *Arbeit**er*** [aKbaIt5] "worker") were investigated. Both suffixes have corresponding pseudo-suffixes, i.e. no "n-suffix counterparts (e.g., *Schach**t*** [Saxt] "shaft"; *Vat**er*** [fa:t5] "father"). All the words with these (pseudo-)suffixes were collected from a spontaneous speech corpus with duration and articulographic recordings (Arnold and Tomaschek, 2016). Two Generalized Additive Mixed-effects Models were fitted (Wood, 2017). One model tested duration and the other tested vertical tongue position during the word-final segment (i.e., [-t] or [-5]), as a function of vocality, the morphological status of the (pseudo-)suffixes, and the interaction between them. Additionally, word frequency, speech rate, and word duration were included as covariates as well as random effects of speaker. The tongue-position model also included the previous and following segments as random effects to take coarticulation into account.

The duration model showed significantly longer duration of the suffixal *-er*, compared to the pseudo-suffixal *-er*, while the duration of the word-final *-t* was not significantly different between the suffixal and pseudo-suffixal *-t* (Figure 1a). For articulatory realizations, the tongue tip trajectory of word-final *-er* was estimated to be significantly lower for the suffixal *-er*, compared to the pseudo-suffixal *-er* (Figures 1b and 1c). Since *-er* [-5] is a low vowel, lower tongue positions indicate clearer articulation. In contrast, the model predicted no significant difference between the tongue trajectories of the suffixal and pseudo-suffixal *-t* (Figure 1c).

These findings suggest vocalic segments are more likely to be enhanced by making them longer in duration and by articulating them more clearly, compared to (stop) consonants. The present study is the first to investigate phonetic enhancement effects of morphology in terms of inherent differences in acoustic/articulatory properties of segments. We propose that some acoustic cues can be enhanced with relatively little additional articulatory effort, while others would require an inhibitive level of additional effort to achieve equivalent enhancement. Here we investigated stop consonants versus vowels. The situation, however, may be different for fricatives (Schmitz and Baer-Henney, 2024). To get to the bottom of this complex question may require further investigation. Nevertheless, the results found here suggest the speakers seem to adjust their speech according to such inherent differences between vocalic and (stop) consonantal segments: vocalic cues are worth enhancing, while (stop) consonants may not be.



(a) Suffix duration. (b) Tongue trajectory for *-er*. (c) Tongue trajectory for *-t*.

Figure 1: (*a*). *Predicted suffix duration. The x-axis represents the morphological condition with "TRUE" and "FALSE" representing the real- and pseudo-suffixes respectively. The whiskers represent the confidence intervals. (b) and (c). Tongue tip trajectories for the word-final -er (middle) and -t (right). The x-axis represents normalized time with 0 and 1 corresponding to the onset and offset of word-final -er/-t. The blue and red curves represent the real- and pseudo- suffixes respectively. The shaded areas in represent the confidence intervals.*

Arnold, Denis and Fabian Tomaschek (2016). "The Karl Eberhards Corpus of spontaneously spoken southern German in dialogues — audio and articulatory recordings". In: *Tagungs- band der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*, pp. 9–11.

Hay, Jennifer (2007). "The phonetics of 'un'". In: *Lexical Creativity, Texts and Contexts*. Ed. by Judith Munat. Amsterdam/Philadelphia: John Benjamins, pp. 39–57.

Nixon, Jessie S and Fabian Tomaschek (2023). "Does speech comprehension require phonemes?" In: *The Handbook of Usage-Based Linguistics*, pp. 161–178.

Schmitz, Dominic and Dinah Baer-Henney (2024). "Morphology renders homophonous segments phonetically different: Word-final /s/ in German". In: *Proceedings of Speech Prosody 2024*, pp. 587–591.

Smith, Rachel, Rachel Baker, and Sarah Hawkins (2012). "Phonetic detail that distinguishes prefixed from pseudo-prefixed words". In: *Journal of Phonetics* 40.5, pp. 689–705. DOI: 10.1016/j.wocn.2012.04.002.

Song, Jae Yung, Katherine Demuth, Stefanie Shattuck-Hufnagel, and Lucie Ménard (2013). "The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultra- sound". In: *Journal of Phonetics* 41.3-4, pp. 281–295.

Strycharczuk, Patrycja and James M. Scobbie (2016). "Gradual or abrupt? The phonetic path to morphologisation". In: *Journal of Phonetics* 59, pp. 76–91. DOI: 10.1016/j.wocn.2016.09.003.

Wood, Simon N. (2017). *Generalized additive models: An introduction with R*. 2nd. Boca Raton, Florida, USA: CRC Press.

Instability of contrastive phonation cue weighting: patterns of individual differences

Xiaobei Chen
University of Chicago

A key challenge in phonation research arises from the non-uniform association between phonation types and other co-varying acoustic properties like pitch and vowel quality (Matisoff 1973; Maddieson & Ladefoged 1985; Esposito & Khan 2020). Examining individual differences in cue weighting offers crucial insights into the nature and variability of phonation. This study aims to address two previously understudied issues: (i) **speaker-specific patterns in phonation cue weighting**, and (ii) the **temporal dynamics** of tense and lax phonation contrasts within vowels in Hani (Yi branch, Tibeto-Burman).

Background. The two-way phonation contrast in Yi languages has been broadly documented as creaky vs. modal voice (Kong 1996; Zhu & Zhou 2008) or tense vs. lax from a typological perspective (Kuang & Keating 2014; Keating et al. 2023). While previous literature has reported group-averaged cue weighting patterns across languages, they have rarely explored individual variation within a speech community, particularly in terms of how phonation contrasts unfold over a vowel's temporal course.

Experimental method. The data presented in this paper were collected through on-site fieldwork in Qimaba, China (22°10'N, 102°50'E). 40 native Hani speakers (F22, M18) participated in the production experiment, ranging in age from 25 to 66 (mean = 48.75). The word list comprised 21 minimal pairs of frequently used monosyllabic words with a CV structure, featuring two phonation types (tense and lax), two tones (33 and 21), and six vowels (a, i, u, ɛ, o, ɤ), e.g. *'early'* /na̰^{21}/ and *'many'* /na^{21}/. Additionally, nine filler words were included, which shared similar consonants and vowels as the target tokens but carried a high tone 55, e.g. *'injured'* /na^{55}/. During the translation task, the experimenter orally elicited each word in Chinese in a randomized order. Participants were then required to translate each word into Hani and repeat it four times while being recorded[1]. In total, 8160 tokens were collected (51 words × 4 repetitions × 40 speakers).

Data analysis method. Acoustic parameters are extracted using VoiceSauce (Shue et al. 2011) at 1ms intervals throughout each vowel, including f0, vowel formants, vowel duration, phonation cues such as H1*, H2*, H1*-H2*, H1*-A1*, H1*-A2*, and H1*-A3*. Raw data is temporally and z-score normalized within speakers (Lobanov 1971), with a principal component analysis (PCA) performed on phonation measurements to reduce dimensions. Generalized additive mixed models (GAMMs) are then employed to capture the dynamic trajectories of these cues. On this basis, cue weights are quantified as coefficients from mixed-effects logistic regression and discriminant analysis (Shultz et al. 2012). Finally, k-means clustering is applied to identify patterns of individual differences in phonation cue weighting.

Sample results[2]. As examples, we present data from a 53-year-old female speaker (F53) and a 48-year-old male speaker (M48). They both produce lax vowels with higher **H1*-A1* and H1*-H2*** values. However, as shown in Figure 1, F53 displays a relatively delayed phonation distinction, while M48 differentiates tense and lax properties at vowel onset, with the contrast merging after time point 6. Regarding **pitch**, Figure 2 reveals that for F53, lax syllables tend to have higher f0, whereas M48 produces tense ones with a slightly higher f0. Figure 3 illustrates that M48's **vowel space** is more affected by phonation, with a decrease in F2 for back vowels in tense contexts and a reduction in F1 for the high-mid vowel /e/ compared

[1] The words were not produced in frame sentences because this study serves as a baseline for a larger project on prosodic effects.

[2] This is a new work in progress, with data analysis expected to be completed by May 2025.

to /e/. In contrast, F53's vowel qualities remain stable across phonations, except that /o̱/ is pronounced higher and more retracted than its lax counterpart.

Discussion. This case study reveals notable individual differences in the timing and extent of phonation distinctions, as well as in cue weighting strategies. These findings highlight speaker-specific realizations of phonation contrasts and suggest possible trajectories for sound change. To further explore these variation patterns, a cluster analysis will be conducted using data from all 40 participants. Potential challenges include (a) interpreting the clusters in relation to sociolinguistic factors and (b) selecting appropriate parameters to describe and cluster dynamic data.

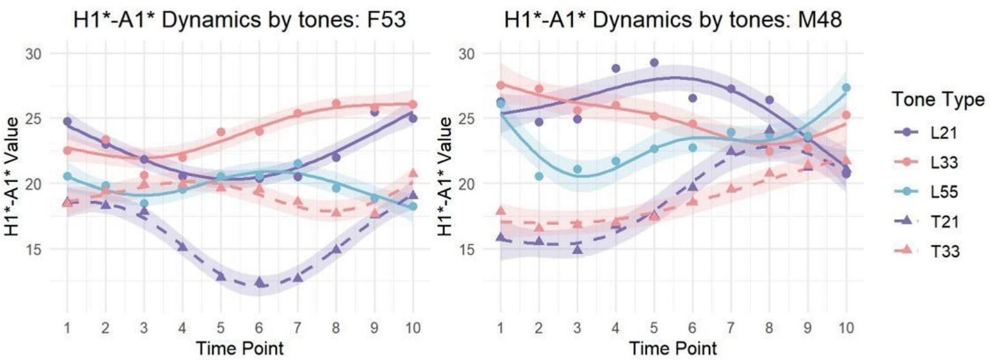


Figure 1. *H1*-A1* dynamics by tones (GAMs fit)*

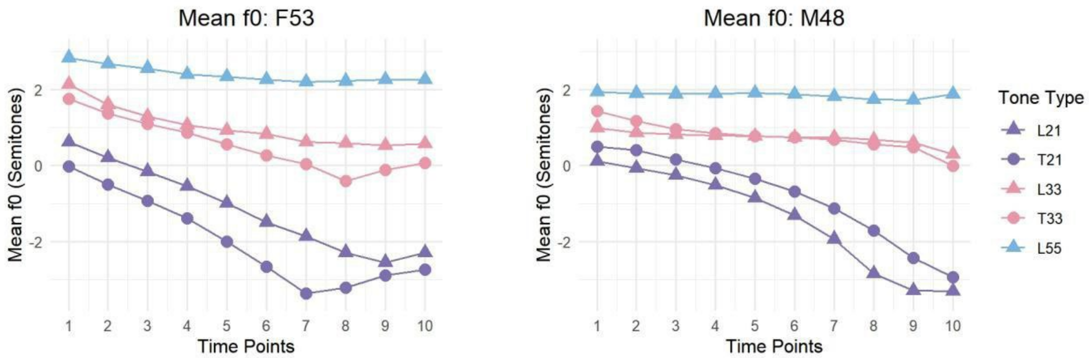


Figure 2. *Mean f0 in semitones (L: lax, T: tense)*

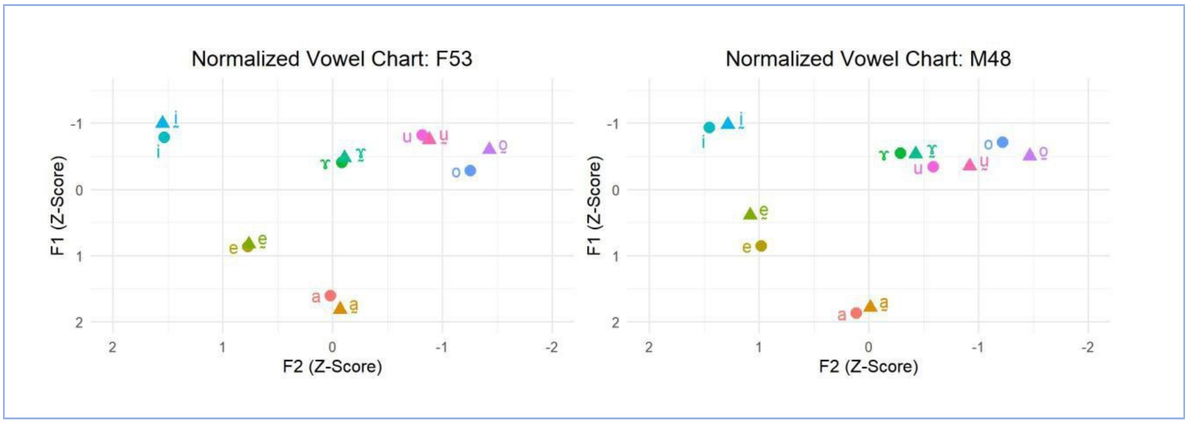


Figure 3. Z-score normalized vowel charts

Esposito, Christina M., and Sameer ud Dowla Khan. "The Cross-Linguistic Patterns of Phonation Types." *Language and Linguistics Compass*, vol. 14, no. 12, 2020, e12392.

Keating, Patricia, Jianjing Kuang, Marc Garellek, and Christina M. Esposito. "A Cross-Language Acoustic Space for Vocalic Phonation Distinctions." *Language*, vol. 99, no. 2, 2023, pp. 351-389.

Kong, Jiangping. "哈尼语发声类型声学研究及音质概念的讨论" [“Acoustic Study of Phonation Types in Hani and Discussion of the Concept of Voice Quality”]. 民族语文, vol. 1, 1996, pp. 40-46.

Kuang, Jianjing, and Patricia Keating. "Vocal Fold Vibratory Patterns in Tense versus Lax Phonation Contrasts."*Journal of the Acoustical Society of America*, vol. 136, 2014, pp. 2784–2797.

Maddieson, Ian, and Peter Ladefoged. "‘Tense’ and ‘Lax’ in Four Minority Languages of China." *Journal of Phonetics*, vol. 13, no. 4, 1985, pp. 433–454.

Matisoff, James A. "Tonogenesis in Southeast Asia." Consonant Types and Tone, vol. 1, *Southern California Occasional Papers in Linguistics*, 1973, pp. 71-96.

Shue, Y.-L., Keating, P., Vicenik, C., and Yu, K. "VoiceSauce: A Program for Voice Analysis." *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 2011, pp. 1846–1849.

Labiodental and interdental fricatives in Jordanian Arabic speakers with class III malocclusion: an acoustic analysis

Aziz Jaber

This study examines the acoustic properties of labiodental and interdental fricatives (/f, θ, ð, ðˤ/) in Jordanian Arabic speakers with Class III malocclusion compared to those with normal occlusion. It focuses on frication duration and center of gravity (COG) in CVVC sequences with long vowels (/ā, ū, ī/). The study included 28 participants—14 with Class III malocclusion and 14 with normal occlusion—balanced for gender. Speech data were analyzed using Praat, and statistical analysis was conducted in SPSS (version 27.0) using Two-way Multivariate Analysis of Variance (MANOVA) and repeated measures tests.

Results showed significant differences between groups. Frication duration for voiced interdental fricatives (/ð, ðˤ/) was significantly longer in Class III individuals (M = 125, SD = 49.076) than in controls (M = 104, SD = 23.399), $F(1, 620) = 16.713$, $p < 0.001$. Voiceless fricatives (/θ, f/) also had longer durations in Class III participants (M = 171, SD = 70.61) than in controls (M = 160, SD = 41.195). The significant effect of voice and occlusal classification, $F(2, 620) = 172.064$, $p < 0.001$, underscores airflow disruption due to mandibular protrusion.

For COG, voiceless fricatives (/θ, f/) had significantly lower values in Class III malocclusion (M = 2417, SD = 1746.46) than in normal occlusion (M = 5097, SD = 2395.97), indicating a posteriorized tongue position ($F(1, 620) = 114.005$, $p < 0.001$). Conversely, voiced fricatives (/ð, ðˤ/) showed higher COG values in Class III individuals (M = 373, SD = 340.589) than in controls (M = 254, SD = 59.381), $F(2, 620) = 832.504$, $p < 0.001$, suggesting articulatory instability.

Repeated measures analysis revealed no significant effect of vowel environment on frication duration or COG ($p > 0.05$). Class III participants consistently exhibited longer frication durations and lower COG averages than controls across all vowel contexts.

These findings highlight the articulatory challenges caused by Class III malocclusion, particularly its effects on airflow, tongue positioning, and fricative production. The results emphasize the need for interdisciplinary collaboration between orthodontists and speech-language pathologists to address speech distortions effectively in individuals with Class III malocclusion.

Workshop: Prosody in Languages of the Middle East

- Talks -

Acoustic correlates of stress in Central Kurdish

Azad Mohammadi,
University of Tehran

This study investigated phonetic stress and its acoustic correlates in the Central Kurdish (CK), focusing on how stress manifested through acoustic parameters like duration, pitch, and intensity in two-syllable and three-syllable words. Stress, characterized by higher pitch, greater intensity, longer duration, and more peripheral vowel quality, makes certain syllables more prominent. Previous studies highlighted various parameters as stress indicators, with duration being the most reliable, although different languages utilize these acoustic correlates differently.

Research on stress-accent languages shows that pitch accents are often confined to syllables bearing primary stress. Unaccented stressed syllables are distinguished by longer duration, greater amplitude, and fuller vowel quality. CK, a major dialect of Kurdish, features an eight-vowel system and a characteristic stress pattern. Primary stress typically falls on the stem-final syllable, though this can shift with different affixes. CK exhibits a binary alternating stress pattern, with primary stress on the final syllable and secondary stress on alternating syllables from right to left. Previous studies indicated that loudness and duration were primary phonetic cues for stress in CK, though these studies were based on auditory perception rather than precise acoustic measurements.

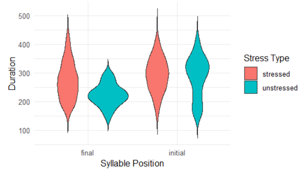
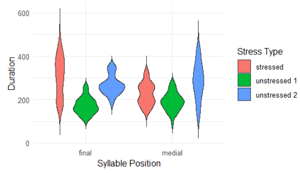
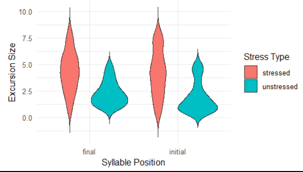
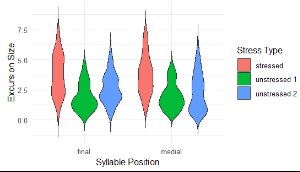
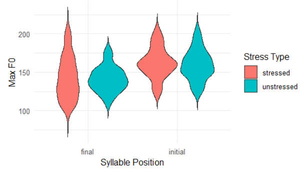
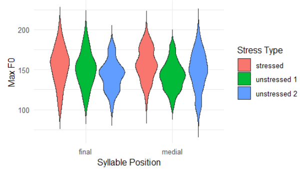
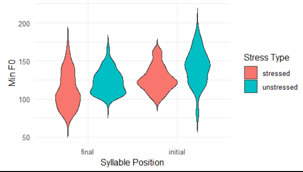
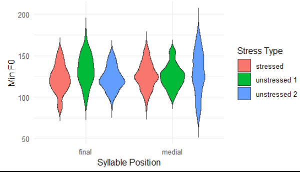
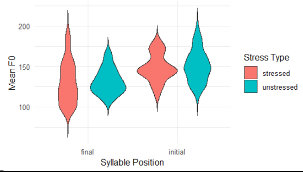
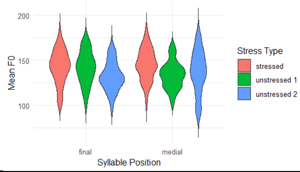
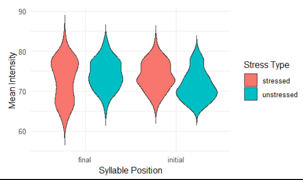
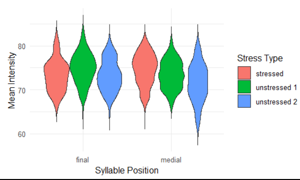
An acoustic study was conducted with ten male native CK speakers from Marivan, Kurdistan Province, Iran, aged 25 to 48, with educational levels ranging from B.A. to Ph.D. Test materials included two-syllable, three-syllable, and four-syllable words in carrier sentences. Each sentence was read twice, yielding 996 tokens. Data were recorded in a quiet environment, with speakers instructed to read naturally at a normal rate. The recorded data were analyzed using Praat with a sampling rate of 22050 Hz in mono wave format. Each target word was manually segmented, labeled by syllable position, and analyzed to extract parameters such as duration, pitch features, and intensity. These parameters were categorized by stress type and syllable position in Excel and analyzed using R software.

In two-syllable words, both stress type and syllable position had a significant effect on syllable duration; also, the interaction between syllable position and stress type was significant. This interaction highlighted the complex interplay between these factors, while in three-syllable words, stress type and syllable position independently affected duration without significant interaction.

Pitch variation, including max F0, min F0, and mean F0, consistently showed higher values in stressed syllables, underscoring their prominence. The effect of stress type on pitch features was significant across the different syllable structures, with syllable position also playing a critical role, especially in two-syllable words. In three-syllable words, a significant interaction between stress type and syllable position was observed for max F0, min F0, and mean F0, indicating a more complex relationship in longer words.

Mean intensity, representing the average loudness, significantly differed between stressed and unstressed syllables. Stress type and syllable position both affected mean intensity, with significant interaction effects in both two-syllable and three-syllable words. This interaction suggested that the influence of stress type on mean intensity varied depending on syllable position.

These findings underscored the importance of considering both stress type and syllable position in the prosody of Central Kurdish. The study provided insights into the prosodic features of Kurdish and their parallels in other languages, contributing to our understanding of linguistic variation and phonetic characteristics across different linguistic contexts. By highlighting the interaction between stress type and syllable position, this research emphasized the nuanced and multifaceted nature of stress in the Kurdish language.

<table>
<tr><th>feature</th><th>two-syllable words</th><th>three-syllable words</th></tr>
<tr><td>Duration</td><td>
</td><td>
</td></tr>
<tr><td rowspan="4">Pitch</td><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Intensity</td><td>
</td><td>
</td></tr>
</table>

Beckman, M. E. (1986). Stress and non-stress accent. Dordrecht: Foris.

Dogil, G. & Williams, B. (1999). The phonetic manifestation of word stress. In H. Hulst (Ed.), Word Prosodic Systems in the Languages of Europe (pp. 273-334). Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110197082.1.273

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America, 27(4), 765–768.

Gordon, M. & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. Linguistics Vanguard, 3(1), 20170007. https://doi.org/10.1515/lingvan-2017-0007

Lehiste, I., and Peterson, G. E. (1959). Vowel amplitude and phoneme stress in American English. Journal of the Acoustical Society of America, 31, 428–435.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. Journal of the Acoustical Society of America, 33, 451–454.

Ortega-Llebaria, M. and Prieto, P. (2010). Acoustic correlates of stress in Central Catalan and Castilian Spanish. Journal of Language and Speech, 54, 73-97.

Sadeghi, V. (2017). Word-level prominence in Persian: An experimental study. Language and Speech, 60(4), 571-596.

Sluijter, A. M., & Van Heuven, V. J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96 (Vol. 2, pp. 630-633). IEEE.

Comparison of declarative and interrogative intonation in Shughni

Sofia Sedunova
Higher School of Economics, Russia

According to the Strong Universalist Hypothesis, pitch rises signal questions and pitch falls signal statements [1, 2], and according to the Nuclear Tone Hypothesis, the distribution of contour types over functions is language-specific, determined by the grammar, and arbitrary [3]. Support for the latter is found in Chickasaw [4]: question Intonational Phrases (IPs) in Chickasaw carry a final L% boundary tone, which contradicts the form-function relation on which the Strong Universalist Hypothesis is built. Moreover, different types of questions may also have different types of pitch patterns, even within the same language. In English, for instance, although polar questions speakers normally use rises, wh-questions may be pronounced with a declarative intonation, namely with a final fall [5]. According to the Functional Hypothesis proposed by Haan (2001), the intonational properties of questions will be more salient in the absence of lexical and/or syntactic markers.

Shughni is an Iranian language (<< Indo-European) spoken by ca. 100,000 people in the Pamir Mountains. Notably, a thorough examination of Shughni intonational patterns is absent from all previous works. According to Edelman & Dodykhudoeva (2009: 792), declarative sentences are marked by a low phrasal tone, while questions are marked by a high boundary tone. On the other hand, Olson (2017: 35–37) notes that content questions begin with the high tone falling during the utterance, and the tone in yes/no questions (marked by the particle =ɔ) has a peak at the verb and then drops at the question marker.

The present study focuses on the intonational contours of three types of constructions: polar questions, wh-questions and statements using the Autosegmental-Metrical framework [9]. The analysis was conducted based on semi-spontaneous speech data collected via various interactive tasks. This exact methodology has been chosen as a better alternative to studying the intonation of natural speech; it was argued that conversational intonation typically has a richer inventory of boundary tones and pitch accents and has more variation in phrasing than reading intonation [10]. Data were collected from 18 native speakers of Shughni. The results showed that the final boundary tone is low (L%) in both questions and statements. Moreover, the typical contour associated with both constructions starts at the high (H*) level. However, polar and wh-questions differ in nuclear pitch accent types: the nuclear pitch accent in polar questions is bitonal with an early rise (L+H*), while in wh-questions it is mostly monotonal high (H* or !H*). These types are phonetically most transparent in question IPs where the final boundary tone is L% [4]. In statements, the nuclear pitch accent is L*, indicating that statements are characteristically marked by a low nuclear pitch accent that falls on the final syllable of an IP (see Figure 3). The placement of the nuclear pitch in wh-questions depends on the positioning of the wh-word (see Figure 2). In polar questions, the word to which the question marker is added receives the final pitch accent, and the most prominent word in an IP receives the nuclear pitch accent (see Figure 1). The question particle itself remains unstressed and serves as a point at which the pitch falls to the final low boundary tone originates. The lack of significant differences between the two question constructions is due to the presence of wh-words in wh-questions and the question particle in polar questions, respectively. That is to say, the usage of semantic and morphological cues makes the prosodic correlates in questions less efficient.

Taken together, the results indicate that in Shughni, boundary tone does not help to distinguish between the two question constructions and the statements, which is not surprising considering the additional question markers present in both question constructions. The question types are distinguished by nuclear accent types and their placement. And a statement sentence is pronounced with a low nuclear pitch accent that marks the final word of an

Intonational Phrase. Evidently the results of present work contradict the previous analyses of intonation patterns in Shughni. In particular, no "slowly rising and immediately falling intonation" in declarative constructions, as predicted by Olson (2017: 35–37), was observed. A further step, which must be left for future work, would be to investigate the intonational contours of another question construction, especially those that miss any question markers. That would allow one to determine whether there is a specific interrogative intonational pattern that marks the question constructions along with the morphological elements.

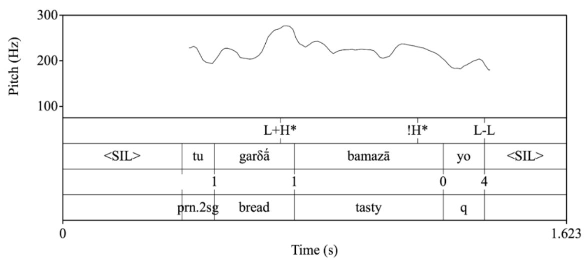


Figure 1: *Pitch contour of the utterance 'Tu garδǻ ba-mazā=yo?' ('Is your bread tasty?'); polar question.*

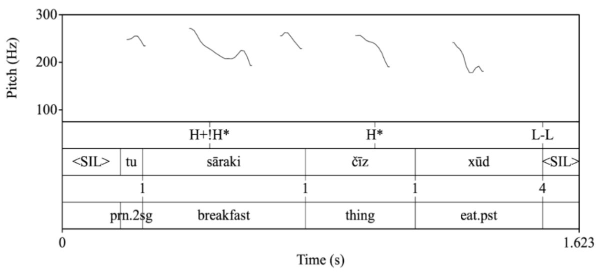


Figure 2: *Pitch contour of the utterance 'Tu sāraki čīz xūd?' ('What did you eat for breakfast?'); wh-question.*

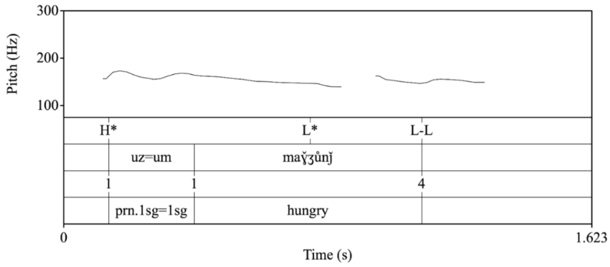


Figure 3: *Pitch contour of the utterance 'Uz=um maγ̌ʒůnǰ' ('I am hungry'); statement.*

[1] Hermann, E. *Probleme der Frage,* VandenHoeck & Ruprecht, Göttingen, 2 vols., 1942.

[2] Bolinger, D. "Intonation across languages". In Greenberg, J. (ed.) *Universals of Human Language. Vol II: Phonology*, Palo Alto, CA: Stanford University Press. 471-524, 1978.

[3] Ladd, D.R. "On Intonational Universals". In Myers, T. et al. (eds). *The Cognitive Representation of Speech*, North Holland Publishing, Amsterdam, 389-397, 1981.

[4] Gordon, M.K."The intonational structure of Chickasaw." *ICPhS* 14, 1993-1996, 1999.

[5] Lindsey, G. Intonation and interrogation: Tonal structure and the expression of a pragmatic function in English and other languages, UCLA, 1985.

[6] Haan, J. Speaking of questions: An exploration of Dutch question intonation. Utrecht: LOT, 2001.

[7] Edelman D. (Joy) I., Dodykhudoeva L. R.. Shughni. Windfuhr G. (ed.) The Iranian languages. London and New York: Routledge, 2009. P. 787–824.

[8] Olson K. Shughni Phonology Statement. SIL International, 2017.

[9] Pierrehumbert J. "The phonology and phonetics of English intonation," PhD dissertation, M. I. T., 1980.

[10] Jun S. A., Fletcher J. Methodology of studying intonation: From data collection to data analysis. Jun S. (ed.) Prosodic typology II. 2014. P. 493-519.

Prosodic phrasing and its relation to syntax in Farsi

Samaneh Darighgoftar[1], Frank Kügler[1], Gerrit Kentner[1,2]
[1]Goethe-University Frankfurt, [2]Humboldt-University Berlin

This paper reports on the syntax-prosody mapping in Farsi, with a focus on prosodic grouping and its relation to syntactic constituent structure. In various languages, prosody has been shown to have the potential to disambiguate syntactically ambiguous structures. However the disambiguating potential of prosody is by no means universal (Kentner & Féry 2010) [1], as the phonological means that languages have at their disposal to map syntax onto prosody differ widely across languages (Féry 2017) [2].

In Farsi, morphological words bear prosodic prominence on the final syllable (Rahmani et al. 2018) [3]; words are grouped into accentual phrases (AP) usually exhibiting a rising L-H pitch contour (Sadat-Tehrani 2007) [4] with a high boundary tone H% on the right edge when in non-final position, and a falling contour in sentence-final position (at least in declarative sentences). The number of words grouped into APs depends on their syntactic function, with (lexical) arguments or modifiers projecting their own AP, i.e., APs may contain several morphological words, in which case the word-level prominences of non-AP-final words are prosodically demoted (Rahmani 2018) [5]. The scaling of the L-H patterns has been shown to be sensitive to information structure: most importantly, in post-focal position, the rising contours for the APs are significantly compressed (post-focal compression; Rahmani et al. 2018). Other than in intonation languages like English or German, however, no clear effect is observed on the focused constituent itself. In this respect, Farsi resembles South Asian languages like Hindi (Khan 2016, Patil et al. 2008) [6,7]. Here, we examine whether and how Farsi prosody reacts to syntactic constituent structure.

To this end, we compare ambiguous coordinate structures with three and four names (Table 1) (experiment modelled on the example of Wagner 2005 [8], and Kentner & Féry 2013 [9]). The structures were presented in written form with parentheses around grouped constituents for disambiguation. 11 native speakers of Farsi (4f, 7m) were asked to read out the coordinate structures in different grouping conditions. All sentences were read twice, once with the coordinator "-o" (used in more familiar registers), once with the more formal coordinator "va". These coordinators differ in terms of their association with either the left conjunct (enclitic "-o") or right conjunct (proclitic "va"). Pitch contours and duration values were extracted with praat (Boersma & Weenink 2025) [10] and compared among the different coordination structures. The results (Fig 1) show that the prosodic parameters clearly react to the different syntactic groupings. Most conspicuously, names at the end of groupings are lengthened and show a higher pitch excursion of the rising L-H contour relative to the baseline without grouping. Conversely, names at the beginning of groupings are significantly shortened and show a compressed pitch range. Moreover, names outside of groupings are slightly lengthened relative to the baseline. The results are largely in line with predictions from, and interpreted in terms of, the Proximity/Similarity model proposed in Kentner & Féry (2013) [9] for German and English. Even though the prosodic system of Farsi differs significantly from these intonation languages when it comes to the expression of information structure, the prosodic rendering of syntactic constituent structure in the different name groupings is in many respects comparable.

1	N1 N2 N3	Mari and Mali and Nura	ماری و ملی و نورا
2	(N1 N2) N3	(Mari and Mali) and Nura	(ماری و ملی) و نورا
3	N1 (N2 N3)	Mari and (Mali and Nura)	ماری و (ملی و نورا)
4	N1 N2 N3 N4	Mari and Mali and Nura and Nilu	ماری و ملی و نورا و نیلو
5	(N1 N2) N3 N4	(Mari and Mali) and Nura and Nilu	(ماری و ملی) و نورا و نیلو
6	(N1 N2 N3) N4	(Mari and Mali and Nura) and Nilu	(ماری و ملی و نورا) و نیلو
7	N1 N2 (N3 N4)	Mari and Mali and (Nura and Nilu)	ماری و ملی و (نورا و نیلو)
8	N1 (N2 N3 N4)	Mari and (Mali and Nura and Nilu)	ماری و (ملی و نورا و نیلو)
9	(N1 N2) (N3 N4)	(Mari and Mali) and (Nura and Nilu)	(ماری و ملی) و (نورا و نیلو)
10	N1 (N2 N3) N4	Mari and (Mali and Nura) and Nilu	ماری و (ملی و نورا) و نیلو

Table 1. *Experimental conditions as presented in Farsi script (right) with grouping indicated by parentheses. English translations in the middle.*

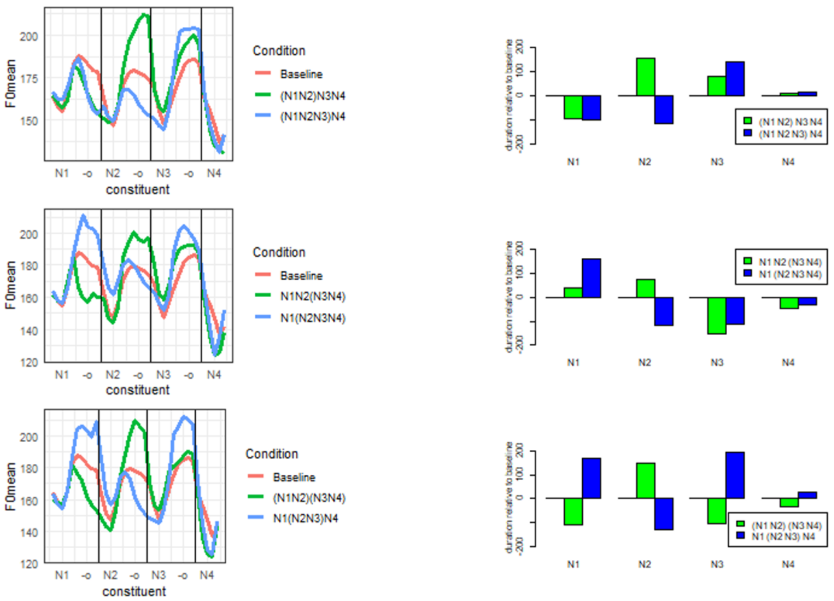


Figure 1. *Left panel: mean pitch contours for left-branching (top), right-branching (middle), and symmetrically branching structures (bottom). Right panel: duration of the four names in left-branching (top), right-branching (middle), and symmetrically branching structures (bottom) relative to the duration of the baseline (set to zero for each name).*

[1] Féry, C. & Kentner, G. (2010). The prosody of embedded coordinations in German and Hindi. *Proc. of Speech Prosody* (Chicago), 100014.

[2] Féry, C. (2017). *Intonation and prosodic structure*. Cambridge University Press.

[3] Rahmani, H., Rietveld, T and Gussenhoven, C. (2018). Post-focal and factive deaccentuation in Persian. *Glossa: a journal of general linguistics* 3(1): 13.

[4] Sadat-Tehrani, N. (2007). *The intonational grammar of Persian*. University of Manitoba dissertation.

[5] Rahmani, H. (2018). Persian 'word stress' is a syntax-driven tone. *Proceedings of Tonal Aspects of Language* (Berlin).

[6] Khan, D. (2016). The intonation of South Asian languages. *Proceedings of Formal Approaches to South Asian Languages* (Amherst).

[7] Patil, U. et al. (2008). Focus, word order, and intention in Hindi. *Journal of South Asian Linguistics* 1.

[8] Wagner, M. (2005). *Prosody and recursion*. MIT dissertation.

[9] Kentner, G. & Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review* 30(2), 277-311.

[10] Boersma, P., & Weenink, D. 2008. Praat: doing phonetics by computer. Computer program.

When sentence prosody overrides word prosody: calling contour in Persian

Cong Zhang,
Newcastle University

One of the core issues in prosody research is the interaction between word and sentence prosody [1]. In tonal language research, extensive research has explored whether lexical tones can be overridden by intonational tunes. Different types of intonational boundary tones have been identified: some modify lexical tones without neutralising tonal contrasts (e.g. floating H boundary tone in Tianjin Mandarin's syntactically unmarked polar question intonation [2], [3]); some add to the lexical tones (e.g. additional L boundary tone in Tianjin Mandarin calling contour [2], [4]); and some almost neutralise tonal contrasts (e.g. additional H boundary tone in Cantonese question intonation [5]). However, no intonational tune fully overrides the lexical prosody. In Persian, word prosody has been described using various terms, including word stress [6], word prominence [7], and pitch accent [8].Despite these terminological differences, most studies agree that prominence typically falls on the final syllables of most words, primarily manifested through pitch. This then raises the interesting question: can sentence prosody override lexical prosody in Persian?

Calling contour, also known as chanted calls, vocatives, or vocative chant, is an intonational tune that has received considerable cross-linguistic attention. Many languages (see [2, p. 180] for review) have a typical intonation contour as in (1), though transcription varies: some transcribe the flat lasting high as a H(Sust)% boundary tone, e.g. [9], others as !H-H%, and some as a !H% alone.

In English, the canonical calling tune consists of an initial L, followed by a H* associated with the stressed syllable, and a final lower !H [10]. When a name has an initial stress, the L is deleted, as shown in (2). In Bengali, for names longer than four syllables, the pitch downstepping can begin at various locations, including the antepenultimate, penultimate, or final syllable [11]. In Dutch, the fall can even include several levels of distinct downward steps [10]. Calling contour in Japanese, a pitch accent language, does not override the pitch accents from what has been reported in [12].

(1) 

(2)

L	H*	!H		H*	!H
Re	be	cca		Jon	ny

Persian calling contour has been mentioned in a number of sources describing Persian word prosody. [13] and [14] present comprehensive phonetic and phonological analyses respectively. However, [13] was published in Persian so it is not accessible to the wider research community; [14] studied the calling contour with a quadri-syllabic name, concluding that it has an intonational pitch accent L+H*, an AP high boundary tone h, and a L% boundary tone. The h AP tone in Persian is the equivalent of a sustained H tone or a !H in some other languages.

This study aims to (1) investigate the interaction between word and sentence prosody in Persian and (2) provide further acoustic evidence for the tune-text association in the Persian calling contour. Audios from a male native Persian speaker, recorded using the Awesome Voice Recorder [15] at a 44.1k Hz sampling rate and a 256 kbps bit rate, were used for the illustrations. Figures 1 clearly shows that all names with a statement tune in the left panes carry a word-final pitch accent. However, in the right panes, the calling contours consistently exhibit an f0 peak at the end of the first syllable regardless of how many syllables there are, indicating a significant shift in f0 contours. Notably, no trace of word-level prosody is preserved in the calling contours.

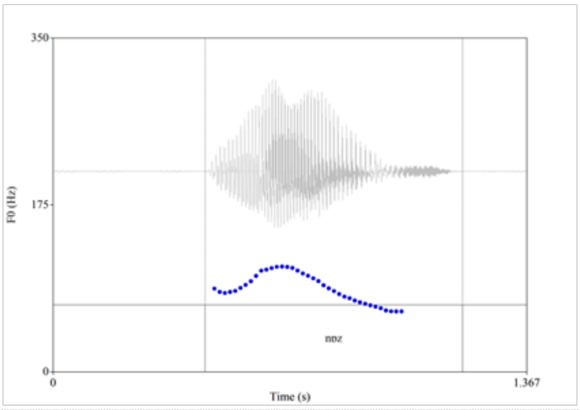
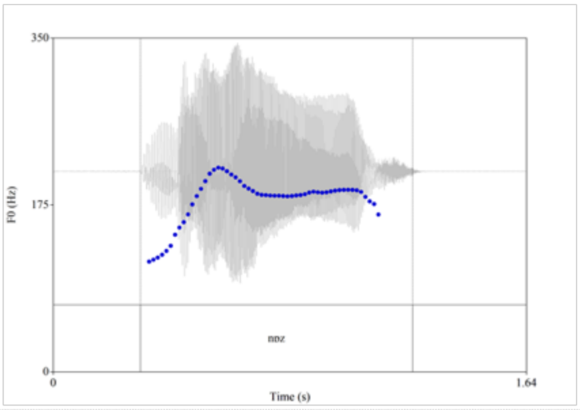
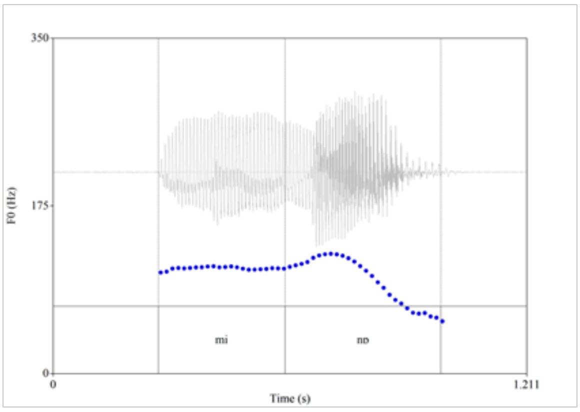
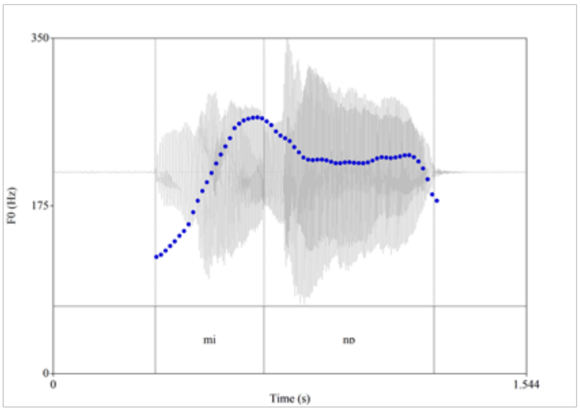
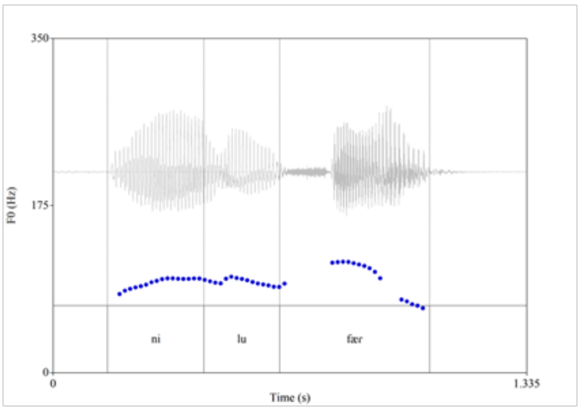
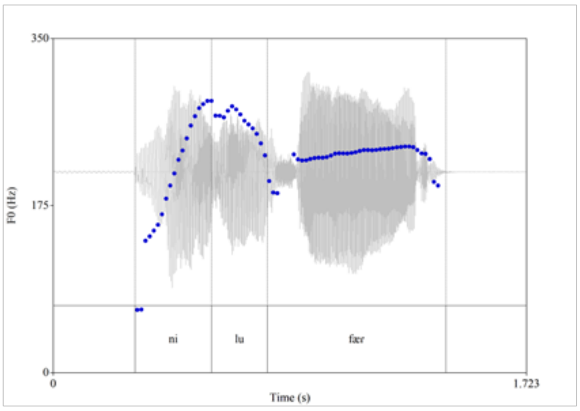
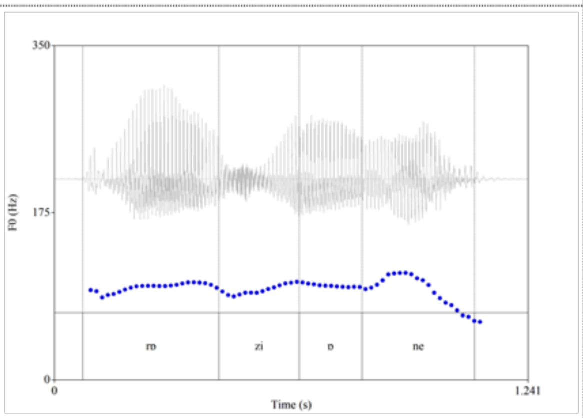
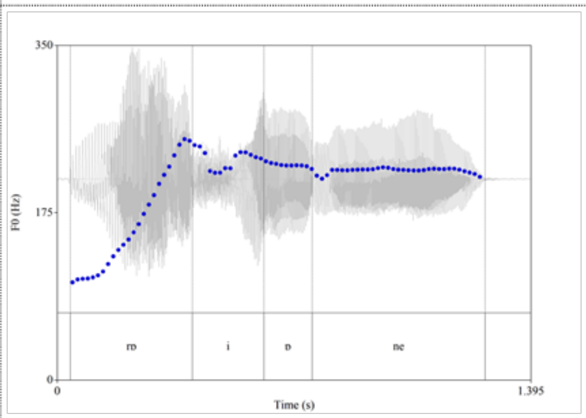
Mono syllabic		
Di syllabic		
Tri syllabic		
Quadri syllabic		

Figure 1: *statement tune (left) and calling tune (right) with different numbers of syllables in Persian*

[1] C. Zhang, 'Stacking and unstacking prosodies: The production and perception of sentence prosody in a tonal language.', in *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, Aug. 2019.

[2] C. Zhang, 'Tianjin Mandarin Tones and Tunes', Doctoral thesis, University of Oxford, 2018. https://doi.org/10.5287/ora-jnnqaven0

[3] C. Zhang and A. Lahiri, 'Floating Boundary Tone: Production and Perception of Syntactically Unmarked Polar Question in Tianjin Mandarin', in prep, doi: https://doi.org/10.31234/osf.io/rd7gw.

[4] C. Zhang, 'Chanted Call Tune in Tianjin Mandarin: Disyllabic Calls', in *Proceedings of Speech Prosody 2018*, Poznań, Poland, Jun. 2018, pp. 522–526.

[5] J. K.-Y. Ma, V. Ciocca, and T. L. Whitehill, 'Effect of intonation on Cantonese lexical tones', *J. Acoust. Soc. Am.*, vol. 120, no. 6, pp. 3978–3987, 2006.

[6] C. A. Ferguson, 'Word Stress in Persian', *Language*, vol. 33, no. 2, pp. 123–135, 1957.

[7] V. Sadeghi, 'Word-level prominence in Persian: An Experimental Study', *Lang. Speech*, vol. 60, no. 4, pp. 571–596, 2017.

[8] V. Abolhasanizadeh, M. Bijankhan, and C. Gussenhoven, 'The Persian pitch accent and its retention after the focus', *Lingua*, vol. 122, no. 13, pp. 1380–1394, Oct. 2012.

[9] A. Rialland and S. Robert, 'The Intonational System of Wolof', *Linguistics*, vol. 39, no. 375, pp. 893–940, 2001.

[10] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, 2008.

[11] B. Hayes and A. Lahiri, 'Durationally specified intonation in English and Bengali', in *Music, Language, Speech and Brain. Wenner-Gren Center International Symposium Series*, J. Sundberg et al. Eds.,1991, pp. 78–91.

[12] I. Abe, 'Intonation in Japanese', in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst and A. Di Cristo, Eds., Cambridge, UK: Cambridge University Press, 1998, pp. 363–378.

[13] M. Eslami and N Bamanian 'Intonation of vocatives in Persian', *Lang. Linguist.*, vol. 16, no. 31, pp. 21–44, 2020.

[14] N. Sadat-Tehrani, 'The Intonational Grammar of Persian', Doctoral Thesis, University of Manitoba, 2007.

[15] C. Zhang, K. Jepson, and Y.-Y. Chuang, 'Investigating differences in lab-quality and remote recording methods with dynamic acoustic measures', *Lab. Phonol.*, vol. 24, no. 1, 2024.

The intonation of polar questions in Syrian Arabic

Malek Al Hasan, Shakuntala Mahanta
Indian Institute of Technology Guwahati

Cross-linguistically, it has been widely agreed that polar or "yes/no" questions are intonationally cued by a high/rising pitch in the utterance, most commonly on the terminal part of the pitch contour [1], [2], [3]. This is the general observation that is considered in the grammatical function of intonation. However, there might also be some other features that are not observed in other languages or dialects of a similar language. In this paper, we explore the intonation of polar questions in Syrian Arabic (SyrA) by looking at the distribution of the rise on the final word or on an earlier position.

[4] argued that polar questions in SyrA are characterized by initial low tone L and a rise in pitch movement to a higher-level tone and that the nuclear accent, in many cases, is on the verb, not on the last content word. In order to investigate this further, an experiment was designed that systematically shifted the position of the focused word in the utterance and, therefore, of the intonation nucleus.

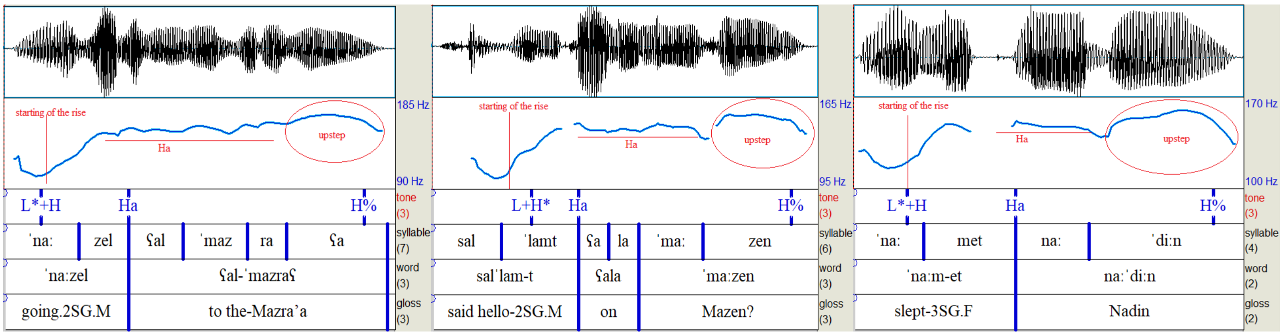
Following [5], we established appropriate contexts by embedding the test sentences in short dialogues. 18 question utterances of short polar questions (one Accentual Phrase AP each) representing two focus conditions (early nucleus & late nucleus) were recorded by 8 (4 males and 4 females) native speakers of SyrA. The location of stress in the final word was systematically varied to have three different stress positions (penult, antepenult & ultima) to figure out whether the final upstep in the F0 contour is aligned with the stressed syllable of the final word (prominence-lending) or with the utterance-final syllable regardless of the stress position in the word (edge-marking).

The analysis of the data reveals the occurrence of two different pitch contours based on differences in the placement of the focused word in the question. The general shape of the two contours would be the ones stylized in Fig. 4. In both early- and late-nucleus utterances, the rise in the pitch always starts from the stressed syllable of the focused word, regardless of the stress position of the final word. All questions were uttered with one accentual phrase (AP), a bitonal pitch accent, a high AP phrase accent (**Ha**), and a high boundary tone H%. The shape of the nuclear pitch accent and the alignment of the starred tone differ according to the position of the stressed syllable in the word. A schematic representation of pitch accent alignment with the stressed syllable of the focused word is presented in Fig. 3.

One main difference between the two focus conditions is that the rise in the pitch was observed to be ***gradual*** in late-nucleus utterances, unlike the ***steep*** rise that occurred in early-nucleus utterances. More importantly, the utterance in Fig. 2c is uttered with a late nucleus and the stress in the final word (the focused item) is on the final syllable (the ultima), resulting in a ***tonal crowding***.

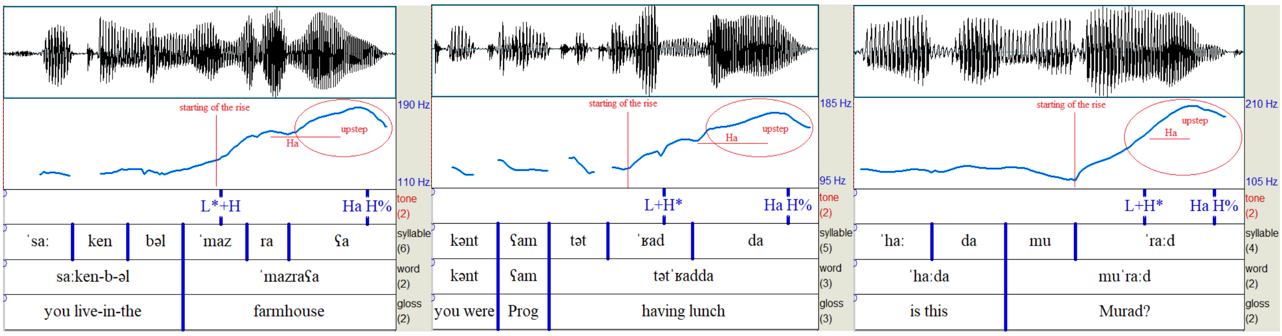
The placement of the final stress (ultima) in late-nucleus questions (Fig. 2c) is realized on the same syllable [*ˈraːd*] of the focused word [*muˈraːd*]. We have three targets here: (1) the nuclear pitch accent (**L+H***), (2) the AP phrase accent (**Ha**), and (3) the IP boundary tone (**H%**), all competing for alignment with the same segmental material. This results in ***tonal crowding*** since all three tones (**L+H***, **Ha** & **H%**) are realized within the same syllable. Thus, the whole rise pitch movement is realized on this final syllable. The nuclear pitch accent is usually a delayed peak (rise) followed by a complex boundary tone analyzed as L+H* Ha L% tune.

The fact that the *step-up* in pitch on the ultimate syllable of the question can appear on an unstressed syllable supports the analysis that this rise is a non-prominence lending pitch movement, i.e., not a pitch accent, but a boundary event occurring at the right edge of the question. The nucleus is more prominent than the others, and it can have any of the two pitch accents (**L*+H**) or (**L+H***). The results indicate that the intonation of polar questions is intricately linked to the position of the nucleus and the stress patterns of the final word, revealing a complex interplay between focus, stress, and intonation. The findings align with Lebanese Arabic [6] and Egyptian Arabic [7], but notable differences are Moroccan Arabic [8], which stands out with a clear rise-fall nuclear contour, and Tunisian [8] with a rise-plateau contour.



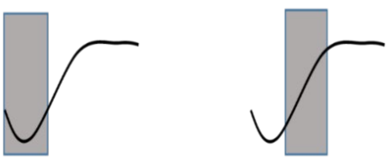
a. antepenult final stress b. penult final stress c. ultima final stress

Figure 1. *Early nucleus polar questions with three stress positions of the final word.*



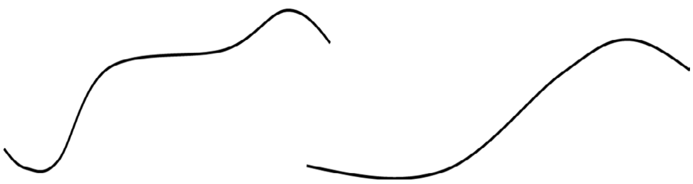
a. antepenult final stress b. penult final stress c. ultima final stress

Figure 2. *Late nucleus polar questions with three stress positions of the final word.*



a. **L*+H** b. **L+H***

Figure 3. *Schematic representation of pitch Accents alignment with the stressed syllable of the focused word in early nucleus questions*



a. Early nucleus b. Late nucleus

Figure 4. *Stylization of two different F0 contours by position of the focused word in the utterance*

[1] Bolinger, D. L. 1978. Yes-no questions are not alternative questions. In *H. Hiz (ed.), Questions*. (pp. 87-105). Dordrecht: Reidel.
[2] Ladd, D. R. 1980. *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.
[3] Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
[4] Al Hassan, M., & Mahanta, S. 2022. The Intonational Phonology of Syrian Arabic: A Preliminary Analysis. *Proceedings of Speech Prosody 2022*, 190-194. Lisbon, Portugal.
[5] Arvaniti, A., D. Robert Ladd and Ineke Mennen. (2006). Phonetic effects of focus and 'tonal crowding' in intonation: Evidence from Greek polar questions. *Speech Communication* 48, no. 6: 667-696.
[6] Chahal, D. 2001. Modeling the intonation of Lebanese Arabic using the autosegmental-metrical framework: a comparison with English. Doctoral dissertation, University of Melbourne.
[7] Hellmuth, S. 2006. *Intonational Pitch Accent distribution in Egyptian Arabic*. PhD dissertation, SOAS, University of London.
[8] Hellmuth, S. 2018b. Variation in polar interrogative contours within and between Arabic dialects. *Proceedings of Speech Prosody 2018*. https://doi.org/10.21437/SpeechProsody.2018-200.

Acquisition of Levantine Arabic word stress in 3- to 5-year-olds

Summer Abdalla, Angeliki Athanasopoulou
University of Calgary

Children developing fixed-stress languages are already insensitive to stress in perception by 9 months [1], while the shape of the word and the location of stress in children's initial productions (2-3 years) depend on the type of stress system they develop as well as the maturity of their phonological domains [2-4]. Arabic is a quantity-sensitive stress system with predictable stress (on the rightmost trochaic foot) but variable stress location (depends on the moraic structure of a word) [5]. Adult speakers of the language are not as sensitive to stress as speakers of unpredictable stress languages [6-7], but we have very limited information about L1 acquisition of Arabic, as the research has primarily focused on L2 acquisition (with the exception of a few production studies [8-9]). In our study, we test the perception and production of nonce and real words in children learning Levantine Arabic as their L1 in the diaspora (Canada) and in native adult speakers of Arabic to explore which components of the Arabic stress system develop in 3 – 5-year-olds.

We tested 19 children and 33 adults in a word-learning task where participants help a character practice their Arabic words, some of which they know (20 real words) and some they have not heard before (10 nonce words). Their task is to listen to the word the character says (3 repetitions) and indicate if they produced them OK, and then to produce the word themselves. We constructed nonce trisyllabic words that should be stressed on the first, second, or third syllable. We recorded these words produced with the correct stress, and with incorrect stress on the other two syllables. We also used real words as control and filler items and those were correct [sa.ˈ**fii**.ni] 'ship', with wrong stress [ba.ˈtaa.**taa**] 'potato', or with a segment error [ˈsa.ma.**t/ki**] 'fish'.

PERCEPTION: Adults and children were not sensitive to differences in the stress location for the nonce words (adults: d' = -0.367; children: d' = -0.197) while for real words children were also not sensitive (d' = 0.745), but the adults were (d' = 1.0727). In general, both groups judged nonce and real words with stress errors as correct with equally high rates, but the effect was not the same for all the stress locations and word forms.

PRODUCTION: Adults and children were highly accurate in producing stress for the real words. In nonce words, we found lower accuracy in both groups that was not symmetrical across all the stress locations. Interestingly, children often modified the word form to fit the stress location they heard in the stimulus instead of moving the stress to the correct location. E.g., the nonce word [dabatabt] was produced as [ˈtab.ba.tab] and not as [da.ba.ˈtabt]; the nonce word [ditabnuk] was produced as [ˈtiː.dab.nuk] and not as [di.ˈtab.nuk]. Words of the form CV(C).CVC.CVC were the hardest for the children, as they were only produced correctly if the stimuli had the correct penultimate stress. This possibly indicates that words with multiple heavy syllables are still difficult to handle. Words CV.CV.CV were highly accurate if the stimuli were stressed on the antepenult (correct) or penult. For both word forms, final stress in the stimuli led to incorrect productions of final stress with no modification of the word form to fit with the stress. This might indicate that the distinction between stress and final lengthening is interfering with children's production. Finally, CV.CV.CVCC were the easiest as children produced them with correct (final) stress. This is the only structure where we see the deletion of the coda when the stimuli are stressed on the antepenult, which points to some knowledge about the role of the coda in the moraic structure of the word and the extrametricality of the word-final coda.

Overall, we found that children who develop Arabic as their L1 demonstrate an understanding of the role of mora and syllable structure in their stress system, in line with the developmental patterns we see in other languages [2-4].

[1] Bhatara, A., Boll-Avetisyan, N., Höhle, B., & Nazzi, T. (2018). Early sensitivity and acquisition of prosodic patterns at the lexical level. In P. Prieto and N. Esteve-Gibert (eds.), *The Development of Prosody in First Language Acquisition* (pp. 37-57). John Benjamins.

[2] Kehoe, M. (2018). Prosodic phonology in acquisition. In P. Prieto and N. Esteve-Gibert (eds.), *The Development of Prosody in First Language Acquisition* (pp. 165-184). John Benjamins.

[3] Demuth, K. (2018). Understanding the development of prosodic words. In P. Prieto and N. Esteve-Gibert (eds.), *The Development of Prosody in First Language Acquisition* (pp. 207-224). John Benjamins.

[4] Fikkert, P. (1994). *On the acquisition of prosodic structure*. ICG Printing.

[5] Hayes, B. (1995). Metrical stress theory: Principles and case studies. UChicago Press.

[6] Segal, O., & Kishon-Rabin, L. (2019). Influence of the native language on sensitivity to lexical stress: evidence from native Arabic and Hebrew speakers. *Studies in Second Language Acquisition, 41*(1), 151-178.

[7] Albadar, I., & Vogel, I. (2023). Are Arabic listeners "stress deaf" to their own L2 pronunciation? In M. Azaz (ed.), *Perspectives on Arabic Linguistics XXXIV* (pp. 79-102). John Benjamins.

[8] Daana, H. A. (2009). The development of consonant clusters, stress and plural nouns in Jordanian Arabic child language. Doctoral dissertation, The University of Essex.

[9] Huneety, A., Mashaqba, B., Alhala, M. A., Guba, M. N. A., & Al-Shdifat, K. G. (2023). Acquisition of stress in the speech of Ammani Arabic-speaking children. *Ampersand, 10*, 100094.

- **Posters** -

Some observations on prosody-driven phonological phenomena in the Jibbali/Shehret dialect of al-Hallaniya (Kuria Muria archipelago)

Giuliano Castagna
Beijing Normal University

Jibbali/Shehret is one of the six pre-literate and endangered Modern South Arabian languages (MSAL), a subgroup of Semitic, the others being Mehri, Soqotri, Harsusi, Hobyot and Bathari. Jibbali/Shehret is spoken in the coastal towns and adjacent mountainous areas of Dhofar, in southwest Oman, as well as on the island of al-Hallaniya (Rubin 2014: 3).

A preliminary study (Castagna 2018) in the insular variety of this language, based on a number of recordings from the early 1980s and 2017, suggests that utterance stress has a role in the occurrence/non-occurrence of some phonological phenomena which can be thus summarised:

1) Non etymologically motivated gemination of consonants

2) /il/ > [il:]

3) /ḳ/ > [g]

Firstly, some consonants tend to become geminate when in a prosodically strong position within an utterance, that is, where speakers resort to utterance stress to convey emphasis. Secondly, by the same token, the cluster /il/, and probably also other clusters involving a high vowel in the adjacency of /l/, does not undergo the well documented shift /l/ > [ɮ] (Johnstone 1981: xiv; Rubin 2014: 26). Rather, /l/ becomes geminate. Conversely, a weak prosodic position within an utterance causes the so-called "emphatic" velar stop /ḳ/, normally realised as [k'], to shift to a voiced velar plosive [g]:

1) gaḥāt aḥ-ḥoggólt
come.ashore.PRF.3.F.SG DEF-Ring.F.SG. (Castagna 2018: 119)

2) ber rtɔ̄f-ɛ́n n-kɔs sillɔ́b
already arrange.PRF-1.CPL 1.CPL-find.IMPV rabbit.fish.M.SG. (Castagna 2018: 134)

3) ṣod ʕag dərɛ́m šɛ
fish.M.SG. in barrel.M.SG. PR.3.M.SG.

In light of the above, this presentation aims to illustrate data from al-Hallaniya Jibbali/Shehret, contrast it with the state of affairs in mainland dialects of the same language, and provide some general typological remarks.

Castagna, G. 2018. A Sketch of the Kuria Muria Language Variety and Other Aspects of Modern South Arabian. Doctoral dissertation. Leeds: University of Leeds.
Johnstone, Thomas Muir. 1981. Jibbāli Lexicon. London: Oxford University Press.
Rubin, Aaron D. 2014. The Jibbali (Shaḥri) Language of Oman. Grammar and Texts. Leiden: Brill.

Adding an Arabic-English corpus of recordings for prosodic analysis to the OMProDat database

Mohammad A. S. Bani Younes[1], Daniel Hirst[2], Wael Zuraiq[3], Rasheed Al-Jarrah[4], Bara'ah AlAbabneh[1]
[1]Al al-Bayt University, [2]Laboratoire Parole et Language, CNRS & Aix-Marseille University, Aix-en-Provence, [3]Hashemite University, [4]Yarmouk University

In the past, the prosody of spoken Arabic has gained little attention, as most previous studies were impressionistic, usually based on the British tradition of intonation analysis (see Bani Younes, 2020). Recently, there has been a growing body of research on different aspects of Arabic prosody using modern software, such as Praat (e.g., AlZaidi, 2019; Alzaidi, et al., 2023; Kelly, 2023; Bani Younes, et al (in press)). Despite this progress, there remains a scarcity of studies comparing Arabic prosody with that of English, whether L1 English or L2 English. One possible reason for this gap might be the lack of a bilingual spoken corpus containing the same data read by the same speakers.

One existing corpus that contains the same texts read by the same speakers in L1 and L2 is the open multilingual prosodic database (OMProDat, Hirst, et al 2013), which currently contains the same recorded texts in seven languages (English (UK), French, Korean, Mandarin Chinese, Finnish, Cantonese, and Persian), and also contains recordings of L2 English read by native speakers of French, Mandarin Chinese and Persian (Taheri-Ardali & Hirst 2022). The corpus uses the same texts originally used in the Eurom1 corpus, developed as a resource in the European SAM project (Chan et al 1995).

However, the Arabic language had not yet been included in this database. In this paper we describe the contribution of Arabic (L1) and English (L2) recordings from Jordanian speakers to the OMProDat corpus, expanding the OMProDat database to include Arabic as the eighth language in this corpus. Thus, four of these languages will have the same texts read by native speakers in their L1 and L2 languages, enabling direct prosodic comparisons across all or some of the languages. In particular, we intend to compare the prosody of recordings of the same texts between L1 Arabic, L2 English (by Arab speakers), and L1 English (by native speakers of British English).

Consequently, ten Jordanian Arabic speakers (5 female & 5 male) recorded 40 English 5-sentence paragraphs (and their Arabic translation). The five sentences in each paragraph are thematically connected. Participants were first briefly informed of the corpus purposes, and then they read and signed the information sheet and the consent form. The English paragraphs were translated into Arabic, adapting the texts to Arabic culture by changing some terms unfamiliar to Arabs. Then, speakers were asked to read the 80 paragraphs (40 English & 40 Arabic). The English and Arabic texts were recorded separately on long sound files with Zoom H8 (sampling rate of 44100Hz 16 bit) and Shure WH20 headset microphone.

Then, TextGrids in Praat were created for each long sound file. Each of the 40 texts in each TextGrid was labelled with the text code (e.g., t01, t02, etc.). We ended up with L1 Arabic and L2 English recordings and TextGrids for the same texts read by the same speakers. This corpus (the recordings, TextGrids, and transcripts) is now free to download by researchers interested in studying the intonation, prosody, or any linguistic aspect of L1 Arabic and L2 English spoken by Arabs.

All collected data were transcribed twice (in orthography and in IPA). An example of the five-sentence paragraphs is provided in (1) below:

(1) t02

ʕindiː muʃkilih maʕ filtar ʔilmajj/mustawal maijj ʕaːliː kθiːr w fajidˤ w binaggitˤ/ btigdar tibʕaθliː fanniː ʔiθalaθaːʔ ʔisˤsˤubiħ law samaħat/ huwwal joomil waħiːd ʔilliː bagdar ʔaʕmaluh haðal ʔisbuːʕ/bakuːn ʃaːkir ʔilkuː law ʔakadtuːli ʔisˤsˤijjaːnih xatˤtˤiːjjan

‘I have a problem with my water softener. The water-level is too high and the overflow keeps dripping. Could you arrange to send an engineer on Tuesday morning please? It's the only day I can manage this week. I'd be grateful if you could confirm the arrangement in writing’.

Currently, the only corpus available for Arabic spoken dialects is IVAr (Hellmuth & Almbark, 2017). Our new corpus is different from IVAr in that it is a new contribution because it might be the first corpus containing data from L1 Arabic with their corresponding L2 English data, which is badly needed in this field. In the full paper, we will provide additional details about the freely available corpus and how to access it. 2

Alzaidi, M. S., Xu, Y, & Xu, A. (2019). Prosodic encoding of focus in Hijazi Arabic. *Speech Communication* 106:127-149.

Alzaidi, M. S. A., Xu, Y., Xu, A., & Szreder, M. (2023). Analysis and computational modelling of Emirati Arabic intonation–A preliminary study. *Journal of Phonetics*, *98*, 101236.

Bani Younes, M. A. S. (2020). *The Role of Phonology in the Disambiguation of Disjunctive Questions* (Doctoral dissertation). University of York, UK.

Bani Younes, M., Colantoni, L., AlZaidi, M. (in press). Prosodic Disambiguation of Disjunctive Declaratives and Disjunctive Questions in Jordanian Arabic. Journal of the International Phonetic Association.

Chan, D., Fourcin, A., Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G., ... & Zeiliger, J. (1995, September). EUROM-a spoken language resource for the EU. In *Proceedings of the 4th European Conference on Speech Communication and Technology*, Madrid, Spain, 1995, pp. 867-870.

Hellmuth, S. & Almbark, R. (2017). *Intonational Variation in Arabic Corpus*. UK Data Archive, Colchester, Essex.

Hirst, D., Bigi, B., Cho, H., Ding, H., Herment, S., & Wang, T. (2013, August). Building OMProDat: an open multilingual prosodic database. in *Proceedings of TRASP, Tools and Resources for the Analysis of Speech Prosody* [satellite workshop of Interspeech], Aix-en-Provence, France, 2013, pp. 11-14.

Kelly, N. E. (2023). Phrase-final intonation adjustment in Lebanese Arabic. *Journal of the International Phonetic Association*, 53(1), 129–158. doi:10.1017/S0025100321000037

Taheri-Ardali, Mortaza & Hirst, Daniel 2022. Building a Persian-English OmProDat database read by Persian speakers. *Proceedings of the 11th International Conference on Speech Prosody,,* Lisbon, Portugal pp *440-444,* doi: 10.21437/SpeechProsody.2022-90.

Extended rhyme in the Shahnameh

Kirill Fessenko,

Freie Universität Berlin, Institut für Iranistik

Persian poetic meter is described as quantitative, using syllable length as the building block for its verse patterns. However, besides syllable length, Persian is characterised by a stress system, which is independent of its syllable length (Hosseini 2004:26). It is also understood that although stress is determined lexically, the grouping of lexical items within a sentence affects the placement of syllabic stress (Ferguson 1957:132). An insight we can add to this discussion comes from the nature of rhymes in Ferdowsi's Shahnameh.

Typically, rhyming selects for identity in the nucleus and coda (but not the onset) of the last syllable of a word:

I'm a little teapot, short and stout,
Here is my handle, here is my spout

However, when a word is polysyllabic and stress doesn't fall on the ultimate syllable, rhyming has to start on the nucleus of the last stressed syllable, with identity in all following material.

There was a young man of Nantúcket.
Who went down a well in a búcket;

We can recognise this behaviour in the Shahnameh for verses where the unstressedness of the last syllable is non-controversial, since that syllable is enclitic:

hamē ātaš afrūzad az gavhár-aš
PROG fire shone from jewel=its

hamē maɣz-i pīlān besāwad sár-aš
PROG brain-ez. heroes rubbed head=its

'The fire from its (i.e., his mace's) jewel was shining; its head was rubbing the brain of the (enemy) heroes.'

Sh. 9.81

However, having established based on these examples that, like in English, stress placement can pull Classical Persian rhyme further back in a line-final phonological phrase, we will show that much larger phrasal units than clitic groups are involved in the process. For example, in the lines below we see that the finite verb can enter into a phrasal grouping with a preceding nominal (to which clitics are attached), its accent subordinated to that of the nominal:

k-aš andēše-ī gấh-i ū āmadī,
so-that=to-him thought-ez. throne-ez. his came

v-agar-š ārəzō jấh-i ū āmadī!
and-if=to-him greed power-ez. his would-come

'So that thought of his throne came to him, and (of) whether his greed and power would come to him.'

Sh. 9.44

These extended rhymes then reveal to us some aspects of Classical Persian phrasal stress, which overrides the stress of some words when they are grouped in a phrase with others. Interesting questions concerning the phonology-syntax interface arise with regards to the conditions on phrasal stress, particularly when finite verbs are involved. Previous work on Persian meter (Elwell-Sutton 1976, Hayes 1979, Heny 1981) has not engaged with the important role that rhyme patterns can play in determining prosodic constituency and phrasal stress patterns in the language. This paper represents the first steps in leveraging this new data source.

Elwell-Sutton, Laurence P. 1976. The Persian Metres. Richmond: Kingprint ltd.

Ferguson, Charles A. "Word Stress in Persian." Language, vol. 33, no. 2, 1957, pp. 123–35.

Hayes, Bruce. 1979. The rhythmic structure of Persian verse. Edebiyat 4, 193-242.

Heny, Jeannine Marie. 1981. Rhythmic Elements in Persian Poetry. Ph.D. Dissertation, University of Pennsylvania.

Hosseini, Seyed Ayat. 2014. The Phonology and Phonetics of Prominence in Persian. Ph.D. Dissertation, University of Tokyo.

The interaction between prosodic features and voice quality metrics in Syrian Arabic: a pilot study

Firas Shbeeb
Department of Linguistics, Eötvös Loránd University

This pilot study investigates the relationship between prosodic features and voice quality metrics in Syrian Arabic. The aim is to address a gap in current phonetic research by examining how intonational patterns—specifically boundary tones and prosodic prominence—affect acoustic measures such as jitter, shimmer, and harmonics-to-noise ratio (HNR). Syrian Arabic, a non-tonal language with unique intonational characteristics, provides a valuable context to test whether established prosodic–phonatory interactions observed in languages like English also apply to Arabic dialects. The originality of this study lies in its focus on Syrian Arabic and its systematic exploration of how phrase position and pitch accent types influence voice quality.

The study is justified by the critical role that prosody plays in conveying grammatical, emotional, and lexical details. Prior research (e.g., Silverman, 2003; Epstein, 2003) has shown that voice quality variations serve as reliable cues to prosodic structure. However, most findings are based on well-studied languages, leaving a gap in our understanding of these relationships in less explored dialects such as Syrian Arabic. Thus, this research aims to determine whether variations in prosodic features lead to measurable changes in acoustic parameters, offering insights into the interplay between speech rhythm and phonation.

The theoretical framework is built upon established models of phonation and prosody (Ladefoged, 1971; Chávez Peón, 2010; Keating et. al., 2015) and incorporates modern acoustic analysis tools, including Praat (Boersma & Weenink, 2024) and the TOBI transcription system (Beckman & Ayers, 1997). The methodology involved recording a single native Syrian Arabic speaker producing 150 sentences with target words placed in various positions (beginning, middle, and end) and across different sentence types (declarative, exclamatory, wh-questions, and yes/no questions). Acoustic measures were extracted and analyzed using descriptive statistics and multiple linear regression in R.

Three main hypotheses guide the study: (1) Syrian Arabic employs a common boundary tone pattern, with falling tones in declarative and exclamatory sentences and rising tones in wh-questions; (2) prosodically prominent words—marked by specific pitch accents—exhibit lower jitter and shimmer due to more stable phonation; and (3) words at phrase boundaries, particularly at the end, display lower HNR, indicating increased creaky phonation. Preliminary findings indicate that jitter is significantly higher at the beginning of phrases, supporting the idea that phrase onsets are more prosodically marked, while differences in shimmer and HNR, although observable, did not reach statistical significance. Descriptive statistics and multiple linear regression analysis also revealed that falling boundary tones (L-L%) predominate in exclamations and statements, whereas wh-questions and yes/no questions exhibit a distinctive H-L% pattern

Long-term objectives involve expanding the study to include a larger speaker pool and a broader range of sentence types, thereby refining models of prosody–phonation interaction. This pilot study lays the groundwork for future research into the acoustic details of intonation and voice quality in Arabic dialects.

Keywords: Prosody, Syrian Arabic, non-modal phonation, intonation, phonatory instability.

[1] Beckman, M., & Ayers, G. (1997). Guidelines for ToBI labelling (Version 3) [Manuscript]. The Ohio State University.

[2] Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* [Computer program] (Version 6.4.05). https://www.praat.org
[3] Chávez-Peón, M. E. (2010). The interaction of metrical structure, tone, and phonation types in Quiaviní Zapotec [Doctoral thesis].
[4] Epstein, M. A. (2003). Voice quality and prosody in English. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2405–2408).
[5] Keating, P. A., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. M. Scobbie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. University of Glasgow.
[6] Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. University of Chicago Press.
[7] Silverman, D. (2003). Pitch discrimination during breathy versus modal phonation. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Laboratory Phonology VI*. Cambridge University Press.

The correlations between different intonational patterns associated with different sentence types within Saudi Dialects: A Cross-dialectal corpus-based Approach

Aljawharah Alzamil,
Independent Researcher

Intonational patterns have been shown to vary across Arabic dialects on a regional level [1-2]. The previous papers showed parallel variation within a single Arabic-speaking country, specifically in Saudi Arabic dialects, based on the first investigations of inter-dialectal intonational variation in yes/no questions (ynqs) [3] and focus utterances [4, 5], in speech production corpus from three urban dialects (Najdi, Hijazi, and Jizani). The results revealed variation in the intonational patterns associated with ynqs and focus with dialectal differences between the Jizani dialect on the one hand, and the Najdi and Hijazi dialects on the other. It also showed that Jizani speakers use non-prosodic syntactic strategies to encode focus more than others, which suggests a possible cross-dialect trade-off between prosodic and non-prosodic strategies in Saudi dialects arguing that the three patterns found in the results are not independent of each other, but due to a trade-off in the allocation available of prosodic strategies for marking of ynqs and marking of focus. Alongside these dialectal differences, this paper aims to explore prosodic description observed in wh-questions using scripted and unscripted tasks of speech corpus of three Saudi dialects.

The analysis exploits a subset of the overall corpus, using scripted dialogue [*sd*] data with comparison to tokens from reading a story [*sto*], retelling a story from memory (*ret*), and free conversation [*fco*]. Data were elicited using instruments adapted from Hellmuth & Almbark [6]. Data were collected with 72 speakers of three urban Saudi varieties: Najdi, Hijazi and Jizani, stratified by gender and age (6M/6F age 18-30 + 6M/6F aged 40-60). The qualitative analysis of prosodic patterns is based on approx. 870 tokens.

The preliminary results show variation in observed prosodic marking patterns between dialects, and expectedly by age and gender. These patterns are in alignment with previously described wh-questions prosody in other Arabic dialects in which the wh-word is rise-fall [7-9] as well as Saudi dialects [10-11]. The difference is in the nuclear contour at the end of wh-word. The question word plays an important role regardless of its position within the sentence, Hijazi dialect showed rise in some cases (Fig.1), unlike Najdi and Jizani, which appear to show fall and rise-fall (Fig. 2-3).

These results along with the previously reported results are not independent, but due to a trade-off in the allocation of available prosodic and non-prosodic strategies for marking different types of sentences. The trade-off is represented by Jizani speakers in the use of a rise-fall in ynqs, which is related to use of less prosodic and more syntactic cues to mark focus to convey meaning, and expected to correlate with wh-questions in the full analysis which is expected to cause a misleading interpretation between wh-questions and focus utterances.

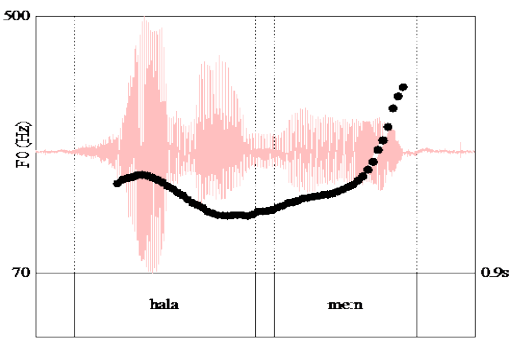


Fig 1. *Rise intonation contour in a sample wh-question from a Hijazi speaker [sahu-sd1-fill-whq1-fo1], 'Hello, who is it?'*

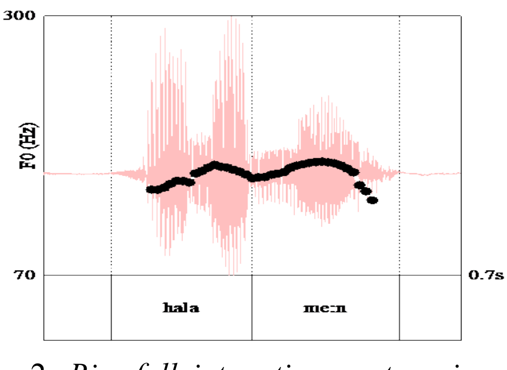


Fig 2. *Rise-fall intonation contour in a sample wh-question from a Jizani speaker [saju-sd1-fill-whq1-mo2], 'Hello, who is it?'*

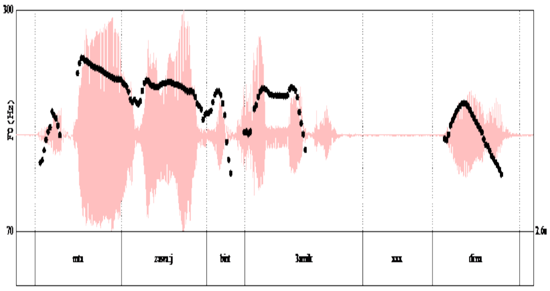


Fig 3. *Rise-fall intonation contour in a sample wh-question from a Jizani speaker [saju-sd1-whq1-mo1], 'What is the name of her Yemeni's husband?'*

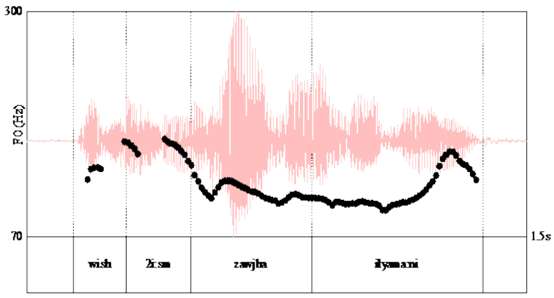


Fig 4. *Rise-fall intonation contour in a sample wh-question from a Najdi speaker [sanu-sd2-whq3-mo6], 'When is the wedding of your cousin Dina?'*

[1] Chahal, D. (2006). The Intonation of Arabic. In K, Versteegh (Ed.), Encyclopaedia of Arabic Language and Linguistics (Vol. 2). The Netherlands: Brill Academic.
[2] El Zarka, D. (2017). Arabic intonation. Oxford: Oxford Handbooks Online.
[3] Hellmuth, S., & Almbark, R. (2019). Intonational Variation in Arabic Corpus. Retrieved from: https://reshare.ukdataservice.ac.u
[4] Alzamil, A., Hellmuth, S. (2021) The prosodic realisation of focus in Saudi Arabic dialects in comparative perspective. *Proc. 1st International Conference on Tone and Intonation* (TAI), 195-199, doi: 10.21437/TAI.2021-40k/852878/
[5] Alzamil A and Hellmuth S. (Manuscript). 'Intonational Variation in Saudi Dialects: A Cross-dialectal corpus-based approach on Yes/No-Questions'.
[6] Alzamil A and Hellmuth S. (In Manuscript). 'The possible mapping between Saudi Arabic dialectal differences and prosody-syntax trade-off based on a speech corpus approach'.
[7] Alsmadi, K., Yunus, K., & Almadani, Y. (2019). THE INTONATION OF CONDITIONAL AND INTERROGATIVE PROMISES IN JORDANIAN ARABIC. International Journal of Humanities, Philosophy, and Language, V.2, Issue. 8, pp. 217-230.
[8] Hasan, M.A., & Mahanta, S. (2022). The Intonational Phonology of Syrian Arabic: A Preliminary Analysis. *Speech Prosody 2022*.
[9] Mohammed Salem, N., & Pillai, S. (2019). An acoustic analysis of the rhythm of Yemeni Arabic. Journal of Modern Languages, 29, 1–22.
[10] Semsem, M. (2018). The use of word order and determiners to mark information status in adult and child Saudi Arabic. [Master thesis, California State University, Fresno].
[11] Moussa, H. (2019). Jeddah Arabic Intonation: An Autosegmental-metrical Approach. [PhD thesis, Newcastle University].

Prosodic marking of focus in Palestinian Arabic: a phonological analysis

Ibrahim Abu-Elrob
Ph.D. student in Linguistics, CUNY Graduate Center

This study presents a phonological analysis of focus marking in Palestinian Arabic (PA). The goal is to investigate how prosodic cues signal focus in PA and to outline the language's inventory of prominence and edge-marking tones, positioning PA within the broader prosodic typology of Arabic dialects. The study is based on a fieldwork experiment using a Question-Answer paradigm, in which four native speakers of PA from the West Bank were recorded producing sentences under three focus contrasts and in three sentence positions. Results suggest that focus in PA is marked by prosodic prominence, primarily through the presence of pitch accents, but without distinct pitch accents tied to specific focus-related contrasts. Instead, two other strategies—categorical de-accenting and phrase break insertion—are employed to signal the discourse meanings of narrow and contrastive focus. The data also indicate that PA features right-headed accents, a large inventory of pitch accent types, and two prosodic units higher than the Prosodic Word (PW): the Intonational Phrase (IP) and the Intermediate Phrase (ip). Overall, the study finds that PA's intonational patterns align closely with those of Levantine Arabic dialects, particularly Lebanese and Syrian Arabic.

Workshop: Multimodal marking of information and discourse structure

Signalling information status in face-to-face dialogue

Janne Lorenzen, Stefan Baumann, Nadia Pelageina, Martine Grice
University of Cologne, Germany

In West-Germanic languages, information status (IS, i.e., the degree of givenness of a word or referent) is known to be marked via prosodic prominence, in that *new* entities are usually more prominent than *accessible* entities, and those are in turn more prominent than *given* entities (e.g., Baumann & Riester 2013). Nevertheless, IS can be overridden by *contrastive focus*, which tends to receive the highest prominence in an utterance (Katz & Selkirk 2011). Importantly, many studies providing evidence for these relations are based on scripted speech, where the information status is indicated by a carefully controlled context. In the present study, we consider unscripted speech collected in face-to-face dialogues in which the context is highly variable. Spontaneous speech produced in such naturalistic settings is often accompanied by co-speech gestures, which may serve as an additional cue to IS (e.g., Im & Baumann 2020). We thus aim to contribute new insights to two research questions: 1) How do speakers mark IS (& contrast) in face-to-face dialogue? and 2) What is the role of gesture?

To address these questions, we make use of a corpus of spontaneous face-to-face dialogues in German (Spaniol et al., 2023), in which two participants sit in the same room facing each other and perform three tasks, each lasting 10 minutes: an introductory conversation, a tangram task, and a discussion of the tangram task. We focus on the discussion, in which a degree of common ground has already been established between the participants.

After transcribing the dialogues, we annotate three aspects of the linguistic material: spoken prosody, gesture, and IS/contrast. All annotations are conducted independently of each other. For spoken prosody, we annotate pitch accent types following the DIMA guidelines (Kügler et al. 2022). For gesture annotation, we focus on manual gestures and annotate gesture strokes, apices, and semantic gesture types following M3D (Rohrer et al. 2023). Lexical and referential IS is identified via the RefLex scheme (Baumann & Riester 2013). At the lexical level, we distinguish between *l-given* (previously mentioned) and *l-new* (first mentioned) as well as *l-accessible* words, which can be inferred through some meaning relation to a previous word. At the referential level, we consider *r-given* and *r-new referents*, as well as three in-between categories: *r-given-sit* for uniquely identifiable entities in the text-external context, *r-bridging* for referents that are dependent on a previously introduced scenario, and *r-unused* for discourse-new referents that are marked with a definite article. In addition, we annotate contrastive focus when an expression has identifiable alternatives in the context (Rooth 1992). Annotation is ongoing. Although our preliminary analysis is based on one (female) speaker, results from more speakers will be presented at the workshop.

Contrast and referential IS are marked by pitch accents in the expected direction (more pitch accents in *contrast* and *r-new/r-bridging/r-unused*, fewer in *r-given/r-given-sit*, Fig.1). Moreover, the ratio of combined accents and gestures increases in the same direction (Fig.2). Lexical IS is not marked as expected (Fig.1). In particular, a considerable number of *l-given* items (69%) are accented, although 30% of accents on *l-given* items are lower in prominence (L* and !H*). 17% of *l-given* items are accompanied by a gesture, compared to 22% of *l-accessible* and 21% of *l-new* items (Fig.2). Gestures most frequently occur in conjunction with a pitch accent, although around 14% of *r-bridging* items are accompanied by a gesture with no pitch accent. These preliminary results suggest that IS is marked differently in face-to-face interaction than in read or scripted speech. Referentiality appears to be more reliably marked, both by accentuation and by gesture. However, there are 30% of cases where *r-new* items are neither pitch accented nor accompanied by a gesture, indicating that either the categorisation using RefLex needs adapting to spontaneous communication or that the mapping of these categories onto intonation and gesture is more complex than in scripted speech.

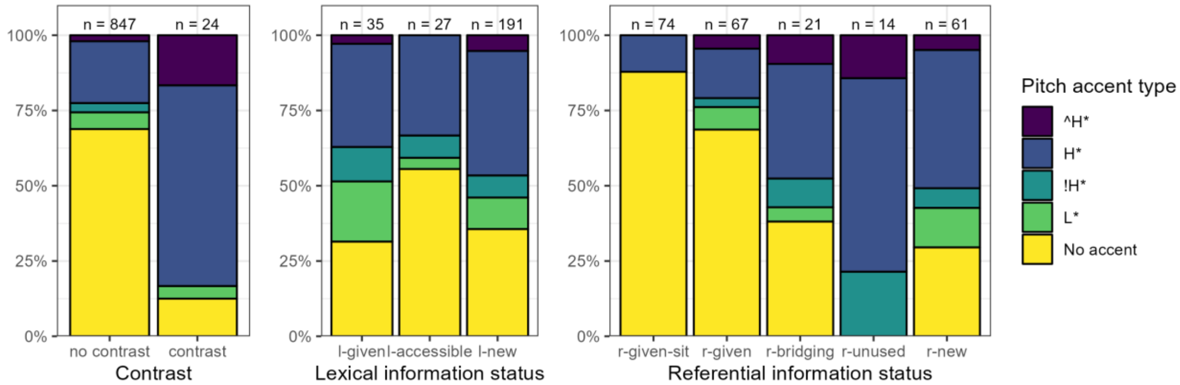


Figure 1: *Marking of contrastive focus as well as lexical and referential information status via (DIMA) pitch accent type. Bar plots show percentages of words associated with pitch accent type. Raw numbers of occurrences per annotation category are annotated at the top of each bar.*

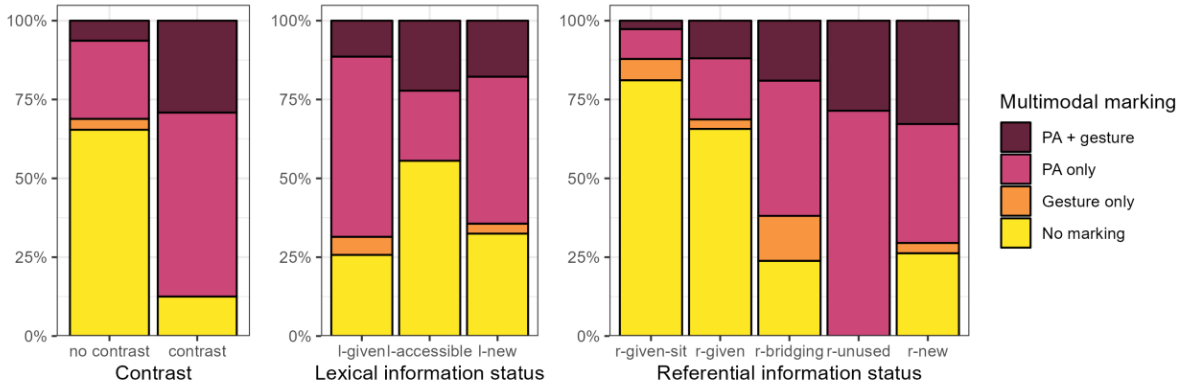


Figure 2: *Multimodal marking of contrastive focus as well as lexical and referential information status. Bar plots show percentages of words associated with pitch accent (PA) and gesture, PA only, gesture only, or without any marking.*

Baumann, S. & Riester, A. (2013). Coreference, Lexical Givenness and Prosody in German. In: J. Hartmann, J. Radó & S. Winkler (eds.). *Lingua* 136, 16-37.

Im, S. & Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings Ling Soc Amer* 5(1), 685–697.

Katz, J. & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4), 771-816.

Kügler, F., S. Baumann & Röhr, C. T. (2022). Deutsche Intonation, Modellierung und Annotation (DIMA) – Richtlinien zur prosodischen Annotation des Deutschen. In: Schwarze, C. & S. Grawunder (eds.): *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion*. Tübingen: Narr. 23–54.

Rohrer, P. L., Tütüncübasi, U., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren, A, Shattuck-Hufnagel, S., & Prieto, P. (2023). The MultiModal MultiDimensional (M3D) labeling system. https://doi.org/10.17605/OSF.IO/ANKDX

Rooth, M. (1992). A Theory of Focus Interpretation. *Natural Lang. Semantics* 1(1), 75-116.

Spaniol, M., Janz, A., Wehrle, S., Vogeley, K. & Grice, M. (2023). Multimodal signalling: the interplay of oral and visual feedback in conversation. *Proceedings of ICPhS 2023*, Prague, Czech Republic.

Are your eyebrows questioning or focusing? Evidence from spoken European Portuguese

Marisa Cruz, Sónia Frota
University of Lisbon

Eyebrow movements have been described as a question marker in the spoken modality of several languages ([1-4]), while also playing a relevant role in the perception of focus ([5-6]). Similarly, in the spoken modality of the Standard European Portuguese (SEP), eyebrow raising, together with a falling head movement, has been shown to be the dominant visual cue for both neutral yes-no questions and narrow focused statements ([7]). However, to our knowledge, the time-alignment of this visual cue with intonational cues is still barely known. [8] observed that eyebrow peaks mainly align with the head of non-nuclear Prosodic Words (PW) in neutral yes-no questions, but they did not describe eyebrow movement over time, or its time-alignment in narrow focused statements. In the current study, our main goal is to inspect the realizational difference (if any) of eyebrow raising in conveying interrogativity or contrastive focus by analysing its kinematics over time for each sentence type/pragmatic meaning.

Audiovisual data from 5 female native speakers of SEP was used. Speakers were videotaped within the InAPoP project [9], while performing a Discourse Completion Task [10] adapted for EP. For the analysis, 33 neutral yes-no questions and 8 narrow focused statements were selected, all (i) formed by more than one PW, (ii) produced with the same visual cues – eyebrow raising together with head falling movement –, and (iii) with the prosodic nucleus at the end of the utterance (only late focused statements were considered). The unbalanced sample is due to constrains of the DCT corpus, as only one utterance (out of two) targeting narrow focused statements is formed by more than one PW. A kinematic analysis of the eyebrows' vertical displacement (pixels - px) along the time series (ms) was conducted using *Kinovea* [11]. Kinematics were automatically extracted from 100 datapoints time-normalized across utterances, thus resulting in a total of 4100 measurements for analysis.

As shown in Figure 1, the eyebrows start moving upwards from the very beginning of the production of narrow focused statements, attaining the highest peak at around 150ms. Differently, in neutral yes-no questions, the eyebrows start moving upwards at around 150ms, attaining the highest peak at around 400ms. Interestingly, within the time-window 600-1000ms, this visual cue exhibits different behaviours, as it gets stable in neutral yes-no questions until the end of the production, whereas it gradually moves upwards in narrow focused statements. A Generalized Additive Mixed Model (GAMM), with participant as random factor and sentence type/pragmatic meaning as predictor (neutral yes-no question, narrow focused statement) shows that the eyebrow raising movement is significantly different between narrow focused statements and neutral yes-no questions (β=9.78, SE=0.47, t=20.98, p<.001). Figure 2 shows that the eyebrow raising movement differs between sentence types/pragmatic meanings in two specific time-windows – 150-200ms (the beginning of the production) and 700-1000ms (the end of the production). For neutral yes-no questions, these results seem to be consistent with previous findings showing that eyebrow peaks mainly align with the head of non-nuclear PWs ([8]). For narrow focused statements, we may conclude that the eyebrow raising movement signals the pragmatic meaning right from the beginning of the production, corresponding to the highest peak of this visual cue, followed by a raising movement at the end of the utterance, time-aligned with the prominent PW (i.e., the one bearing the focus pitch accent). These findings, together with the fact that the eyebrow raising amplitude in both sentence types/pragmatic meanings is similar in the time-window 200-700ms (Figures 1 and 2), and the fact that the head amplitude distinguishes between statements and yes-no questions ([12]), seem to suggest that eyebrow raising movement is more relevant for conveying focus than interrogativity in the prosodic grammar of the spoken modality of EP.

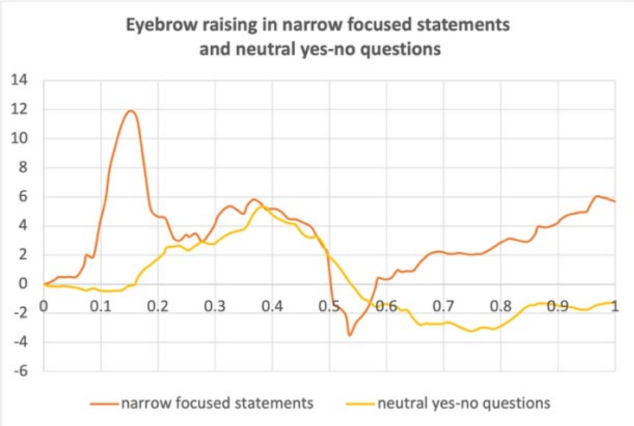


Figure 1. *Eyebrow raising movement tracked (in pixels) along the normalized time series for narrow focused statements (orange line) and neutral yes-no questions (yellow line). This figure illustrates aggregated data.*

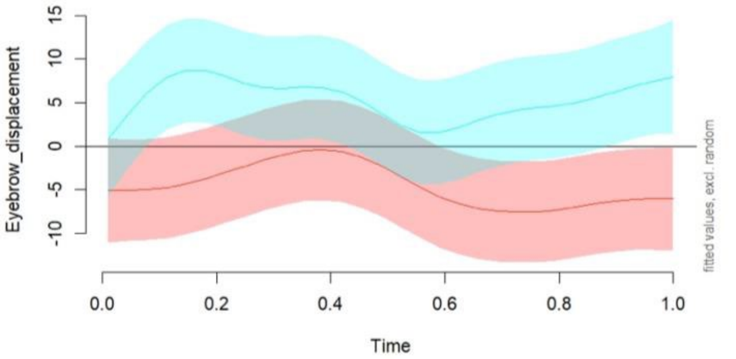


Figure 2. *Non-linear smooths (fitted values) for narrow focused statements (blue) and neutral yes-no questions (red). 95% Confidence Intervals are shown by shaded bands.*

[1] Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F. & Espesser, R. 1996. About the relationship between eyebrow movements and F0 variations. *Proceedings of the ICSLP*, vol. 4, Philadelphia, 2175–2178.

[2] Purson, A., Santi, S., Bertrand, R., Guaïtella, I., Boyer, J. & Cavé, C. 1999. The relationships between voice and gesture: Eyebrow movements and questioning. *Proceedings of ECST 6*, Budapest, Hungary, 1735–1739.

[3] Crespo-Sendra, V., Kaland, C., Swerts, M. & Prieto, P. 2013. Perceiving incredulity: The role of intonation and facial gestures. *Journal of Pragmatics, 47*, 1–13.

[4] Borràs-Comes, J., Kaland, C., Prieto, P. & Swerts, M. 2014. Audiovisual correlates of interrogativity: A comparative analysis of Catalan and Dutch. *J of Nonverbal Behav, 38*, 53–66.

[5] Krahmer, E., Ruttkay, Z., Swerts, M. & Wesselink, W. 2002. Pitch, eyebrows and the perception of focus. In B. Bel (Ed.), *Proceedings of Speech Prosody 2002*, 443–446.

[6] Borràs-Comes, J. & Prieto, P. 2011. 'Seeing tunes'. The role of visual gestures in tune interpretation. *Laboratory Phonology, 2*(2), 355–380.

[7] Cruz, M., Swerts, M. & Frota, S. 2015. Variation in tone and gesture within language. *Proceedings of the 18th ICPhS*, University of Glasgow, UK. Paper number 452.

[8] Cruz, M. & Frota, S. 2023. Prosodic domains in the head of the signer. Poster presented at the *MMSYM 2023* conference (Barcelona).

[9] Frota, S. (Coord.). 2012-2015. *InAPoP – Interactive Atlas of the Prosody of Portuguese*. Funded by FCT – Fundação para a Ciência e a Tecnologia (PTDC/CLE-LIN/119787/2010).

[10] Félix-Brasdefer, J. C. 2010. Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In E. Usó Juán & A. Martinéz-Flor (Eds.), *Speech act*

performance: Theoretical, Empirical, and Methodological Issues (pp. 41-56). John Benjamins.
[11] *Kinovea* 0.9.4. https://www.kinovea.org/
[12] Cruz, M. & Frota, S. 2024. Four seasons in one head. The prosodic phrasing of enumerations in Portuguese Sign Language. Talk presented at the *MMSYM 2024* conference (Frankfurt).

How does gesture mark information structure and prosodic phrases?

Olcay Türk
AG Phonetik, Bielefeld University

This study investigates how manual gestures are synchronised with prosody and information structure using Turkish natural speech data [1, 2]. The link between gesture and prosody has long been studied in many languages from different perspectives, such as temporal coordination. One similarity between prosody and gesture is that both can be conceptualised to have a hierarchical phrasal structure [3]. At the lowest level, gestural landmarks that are assumed to be prominent (i.e. apexes) have been shown to be synchronised with prominent tonal events (i.e. pitch accents, although there is also evidence for synchronisation with boundary tones, see [1], [9]). However, more studies are required on the synchronisation of whole phrases (i.e., prosodic phrases and gesture phrases) at higher levels of the hierarchy [4, 5]. Moreover, information structural categories such as topic and focus, which are closely related to prosody, also offer themselves as potential anchors for gesture phrases, as evidenced in studies in the recent years [6, 7], requiring further investigations of gesture synchronisation.

The present study analysed the synchronisation of gesture phrases with prosodic phrases (i.e., intermediate and intonational phrases) and information structural categories (i.e., topic and focus) in 3 hours of video recounts of 4 speakers in Turkish [1, 2]. In total, the data contained 515 representational gestures, 1363 intermediate phrases (648 pre-nuclear, 491 nuclear, 224 post-nuclear), 675 intonational phrases, and 1060 information structural units (387 topics, 540 foci, 133 backgrounds). Statistical analysis of synchronisation consisted of calculating time differences between the onsets and offsets of these units, which were then fitted to mixed effects linear regression models to see whether these time differences were influenced by the following factors: intermediate phrase type, gesture type, IS unit type, and contrast (with participant as the random effect). Finally, synchronisation was statistically determined using equivalence tests measuring whether phrases started and ended within an average syllable duration (160 ms) given the effects observed through mixed-effect regressions [8].

The results showed that there was no one-to-one synchronisation with intermediate or intonational phrases (i.e., a single gesture phrase was not synchronised with a single prosodic phrase), but there was close synchronisation with onsets and offsets of intermediate phrases, suggesting that gesture phrases were synchronised with a series of intermediate phrases that were typically in the preverbal region associated with focus in Turkish. A subsequent analysis of synchronisation with information structure showed that topics and foci were tightly synchronised with stroke + post-hold combinations, which were referred to as apical areas. Interestingly, iconic and metaphoric gestures were more likely to accompany foci, and deictics were more likely to accompany topics, suggesting a further sensitivity to the discursive functions of information structural categories.

Overall, the results confirm a synchronisation of gesture and prosody at the phrasal level and provide evidence that there is a direct link between gesture and IS, where the anchoring location of a gesture as well as its type can be predicted by the organisation of information within an utterance.

[1] Türk, O. (2020). Gesture, Prosody and Information Structure Synchronisation in Turkish. https://doi.org/10.26686/wgtn.17149058.v1

[2] Türk, O., & Calhoun, S. (2024). Phrasal Synchronization of Gesture With Prosody and Information Structure. *Language and Speech*, *67*(3), 702-743. https://doi.org/10.1177/00238309231185308

[3] McNeill, D., & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), Language and Gesture, 141-161. Cambridge: Cambridge University Press.

[4] Karpiński, M., Jarmolowicz-Nowikow, E., & Malisz, Z. (2009). Aspects of gestural and prosodic structure of multimodal utterances in Polish task-oriented dialogues. *Speech and Language Technology, 11*, 113–122.

[5] Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S. (2004). The timing of speech-accompanying gestures with respect to prosody. In *Proceedings of From Sound to Sense (Vol. 50, pp. 97–102)*. Cambridge: MIT.

[6] Kügler, F. & Gregori A. (2023). Iconic Gestures in Focus – Synchronization of Prosody and Gestures in Prominence. In: *Skarnitzl, R. & Volín, J. (eds) Proceedings of 20th ICPhS 2023, Prague, Czech Republic August 2023. (pp. 4125-4129)* (ID 232). Guarant International.

[7] Gregori, A., Sánchez-Ramón, P.G., Prieto, P., Kügler, F. (2024) Prosodic and gestural marking of focus types in Catalan and German. In *Proceedings of Speech Prosody 2024. (891-895)*.

[8] Lakens D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362.

[9] Rohrer, P. L., Delais-Roussarie, E., & Prieto, P. (2023). Visualizing prosodic structure: Manual gestures as highlighters of prosodic heads and edges in English academic discourses. *Lingua*, *293*, 103583.

Continuity of head movement patterns in the encoding of focus structures

Lena Pagel[1], Simon Roessig[2], Doris Mucke[1]
[1]University of Cologne, [2]University of York

Prosodic focus is encoded by a wide range of phonetic cues as well as co-speech body movements. Focused items are often accompanied by, e.g., head nods, eyebrow raises or manual gestures, which can be called visual prosody [1-4]. However, the stream of co-speech body motion is continuous and does not necessarily consist of distinct events, such as clearly delimitable head nods. Therefore, the current study explores the continuous head movement patterns that speakers produce in coordination with speech, to investigate how fine-grained visual prosody unfolds over time in monologic vs. dialogic speech. 30 native speakers of German were recorded with electromagnetic articulography (EMA) for the kinematic tracking of head motion. They played a card game, first in a solo condition in front of a screen, then in a dialogue condition face to face with an interlocutor. In response to questions, they produced utterances with controlled lexical material and a consistent syntactic structure, as exemplified in (1). Each utterance contained two target words: One ITEM (e.g.'die Bohne') and one CITY (e.g.'Manila'). Depending on the preceding question, the utterance could be produced in two focus conditions: In the condition [BA-CO], the ITEM was in the background and the CITY was in corrective focus; in the condition [CO-BA], the ITEM was in corrective focus and the CITY was in the background. Speakers' vertical head trajectories (i.e. up-down motion) were captured during the production of the utterances via EMA.

(1) Du hast die Bohne aus Manila auf der Hand.
ITEM CITY
You are holding the bean from Manila in your hand.

The results reveal a large degree of interindividual variation between the 30 speakers in their trajectories of co-speech vertical head motion in both the solo and dialogue conditions. However, interestingly, the majority show up-down head movements associated with the verbal productions of ITEM and CITY, which resemble two beat gestures of varying sizes. The two focus conditions [BA-CO] and [CO-BA] are visually differentiated by several speakers, who enhance this beat-like movement in corrective words compared to words in the background. Figure 1 illustrates head motion trajectories for three exemplary speakers; Figure 2 shows the maximum head position during the two target words. Head motion is negligible in speaker 1 and the two focus conditions are not differentiated. Speakers 2 and 3, however, produce differential motion patterns between focus conditions: Target words in corrective focus (i.e. the ITEM in [CO-BA] and the CITY in [BA-CO]) are associated with a more pronounced up-down movement than those in the background. While the degree of focus differentiation is consistent across solo and dialogue conditions for speakers 2, we see that speaker 3 increases the degree of between-focus differentiation when engaging in a dialogue as compared to the solo condition, which is also observable in several other speakers in the data set. The present results demonstrate the potential of studying visual prosody through finegrained kinematic analyses. Our analyses reveal speaker-specific differences in the encoding of focus structure by means of continuous head movements patterns. Motion differences can be observed throughout the entire utterance, with larger movements co-occurring with prominent words and smaller movements co-occurring with non-prominent words, creating consistent patterns of head motion associated with focus structures. Interestingly, many speakers increase the strength of focus differentiation when speaking with an interlocutor compared to speaking individually. This suggests that speakers exploit the communicative function of visual prosody, as they enhance it when engaging in a face-to-face interaction.

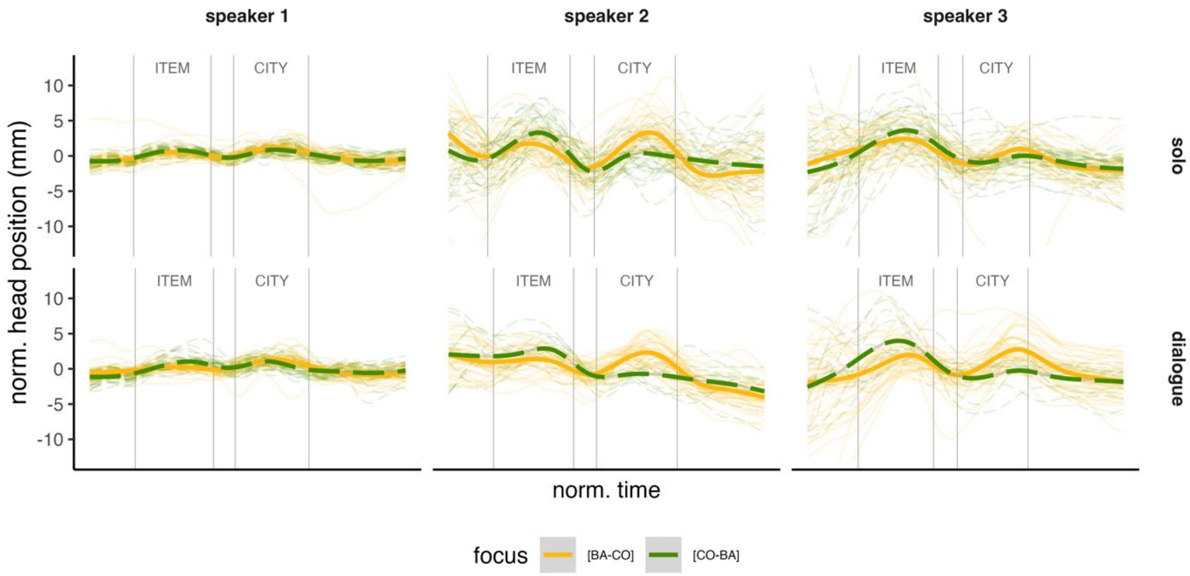


Figure 1. *Vertical (up-down) head motion trajectories during utterance productions for three exemplary speakers in two focus structures and solo and dialogue conditions.*

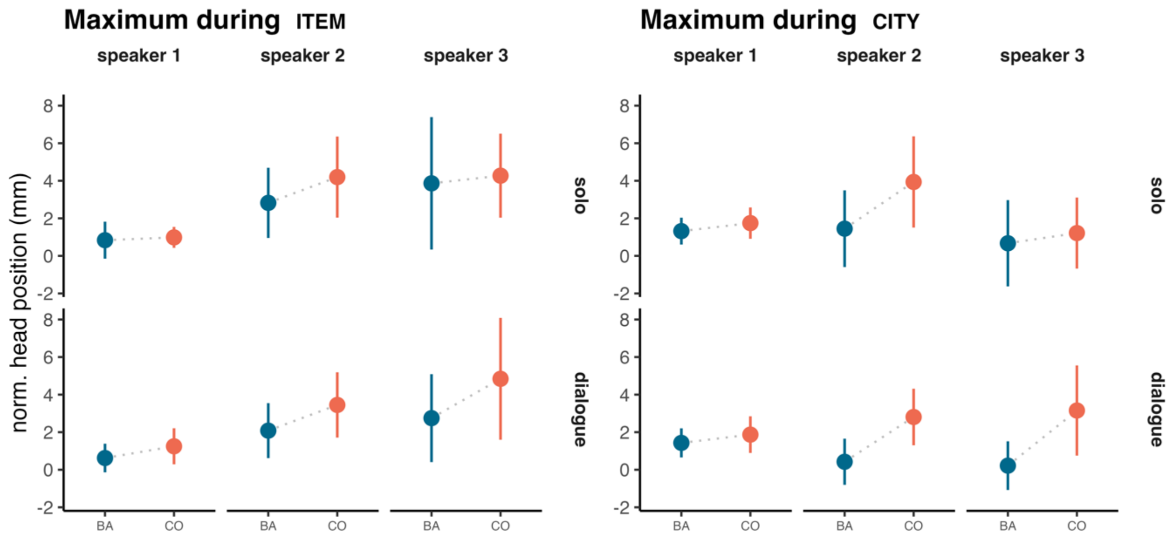


Figure 2. *Maximum vertical head position during productions of target words (ITEM and CITY) for three exemplary speakers in two focus structures and solo and dialogue conditions.*

[1] Ambrazaitis, G., & House, D. 2017. Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Communication* 95, 100–113.

[2] Esteve-Gibert, N., Borras-Comes, J., Asor, E., Swerts, M., & Prieto, P. 2017. The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical* Society of America 141(6), 4727–4739.

[3] Gregori, A., Sanchez-Ramon, P. G., Prieto, P., & Kugler, F. 2024. Prosodic and gestural marking of focus types in Catalan and German. *Proceedings of Speech Prosody* (Leiden, the Netherlands), 891-895.

[4] Krivokapić, J., Tiede, M. K., & Tyrone, M. E. 2017. A Kinematic Study of Prosodic Structure in Articulatory and Manual Gestures: Results from a Novel Method of Data Collection. *Laboratory Phonology* 8(1), 1–26.

The expression of focus types through multimodal prominence in Catalan and German

Paula G. Sánchez-Ramón[1,2], Alina Gregori[2], Frank Kügler[2], Pilar Prieto[3,1]
[1]Universitat Pompeu Fabra, [2]Goethe University Frankfurt, [3]ICREA

Gesture and prosody are known to be integrated [1], in terms of temporal alignment [2, 3] and prominence patterns [4, 5]. In addition, both prosody and gesture have been suggested to be involved in the marking of focus. Different focus types can be distinguished (information focus, contrastive focus, corrective focus, [6, 7]) that have been suggested to be differentiated by prosody and gesture: For prosody, higher levels of prosodic prominence are used to mark contrastive and corrective focus, compared to information focus [8]. For gestures, an increase in gesture production has been found across focus conditions in children, from information to corrective focus [9]. While a few studies separately look into the prosodic or gestural marking of focus, less is known about how prosodic and gestural prominence together contribute to the marking of focus. Therefore, this study is the first to investigate the contribution of (1) prosody, (2) gesture and (3) their interaction to the marking of focus types in Catalan and German. Considering previous findings [4, 5, 8, 9], we hypothesize an increase in (1) prosodic prominence, (2) gesture prominence, and (3) multimodal prominence marking across focus types (information focus < contrastive focus < corrective focus).

We conducted a semi-spontaneous production experiment in Catalan and German that systematically varied elicitation contexts for each focus type, in which participants (n per language = 35) interacted with a digital character. Target focused words (n = 702 in Catalan; n = 682 in German) were annotated for pitch accentuation presence following ToBI [10, 11] and gesture presence (performed by hands, head or eyebrows) following M3D [12]. We also annotated degrees of perceived prosodic (DIMA, [13]) and gestural prominence (M3D) on a scale from 0 to 3 on the target words from a holistic perspective. Inter-rater reliability of each annotated variable was successfully assessed. Generalized linear mixed models were fit to test the effect of focus types on gesture presence. Additionally, cumulative link mixed models were performed to test the effect of focus types on prosodic and gestural prominence as well as on the distribution of gestural prominence given prosodic prominence.

Results for prosody suggest that target words are systematically accented across focus types in both languages. Prosodic prominence increases from information focus to contrastive and corrective focus in both languages. In terms of gestures, gesture presence increases across focus types in both languages, although the effect is stronger in Catalan than in German. Gestural prominence increases across focus types in Catalan, but not in German. The integration of prosody and gesture in relation to focus types is similar in the two languages. In Catalan, a significant effect for prosodic prominence to predict gesture prominence in an increasing relationship is found in contrastive focus, and in German this effect is found in corrective focus.

These results partly confirm our hypotheses by suggesting that multimodal prominence marks some focus types in a cumulative relationship in Catalan and German (in line with [5]). However, due to the different patterns found when assessing prosodic and gesture prominence separately in focus marking, we conclude that gesture is not parasitic on prosody, as prosody and gesture behave differently in the marking of focus types (as suggested by [9]).

[1] D. McNeill, Hand and mind: What gestures reveal about thought. UCP, 1992.

[2] S. Shattuck-Hufnagel, Y. Yasinnik, N. Veilleux, & M. Renwick, “A Method for Studying the Time Alignment of Gestures and Prosody in American English: Hits’ and Pitch Accents in Academic-Lecture-Style Speech,” *NATO Secur. Sci. Ser. E Hum. Soc. Dyn. 18*, 34, 2007.

[3] N. Esteve-Gibert & P. Prieto, “Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements,” *J. Speech Lang. Hear. Res. 56(3)*, 850–864, 2013.
[4] S. Im & S. Baumann, “Probabilistic relation between co-speech gestures, pitch accents and information status,” *Proc. Linguist. Soc. Am. 5(1)*, 685–697, 2020.
[5] G. Ambrazaitis & D. House, “Probing effects of lexical prosody on speech-gesture integration in prominence production by Swedish news presenters,” *Lab. Phonol. 24(1)*, 2022.
[6] M. Krifka, “Basic notions of information structure,” *Acta Linguist. Hung. 55(3–4)*, 243–276, 2008.
[7] S. Repp, “Contrast: Dissecting an elusive information-structural notion and its role in grammar,” 2014.
[8] M. Zimmermann, “Contrastive focus and emphasis,” *Acta Linguist. Hung. 2017 Acta Linguist. Acad. 55(3–4)*, 347–360, 2008.
[9] N. Esteve-Gibert, H. Loevenbruck, M. Dohen, & M. D’Imperio, “Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech,” *Dev. Sci. 25(1)*, e13154, 2022.
[10] P. Prieto Vives, C. Puglesi, J. M. Borràs Comes, E. Arroyo, & J. Blat, “Exploring the contribution of prosody and gesture to the perception of focus using an animated agent,” *J. Phon. 2015 Mar 49 41-54*, 2015.
[11] M. Grice, S. Baumann, & R. Benzmüller, “German intonation in autosegmental-metrical phonology”. In S.-A. Jun (Ed.), *Prosodic typology. The phonology of intonation and phrasing*, 55–83. OUP, 2005.
[12] P. L. Rohrer *et al.*, “The MultiModal MultiDimensional (M3D) labeling system,” 2023.
[13] F. Kügler, S. Baumann, & C. T. Röhr, “Deutsche Intonation, Modellierung und Annotation (DIMA),” *Transkription Annot. Gespr. Sprache Multimodaler Interakt. Konzepte Probl. Lösungen*, 23, 2022.

The interplay between prosody, gesture, and syntax to signal information structure in French

Florence Baills[1], Stefan Baumann[2]
[1]Universitat de Lleida, [2]IfL-Phonetik, University of Cologne

Information Structure (IS) refers to how speakers organize parts of their message to guide listeners' attention and comprehension, distinguishing between what is known (or given) and what is new, inferable or emphasised in the discourse [1]. This structuring occurs across multiple linguistic domains which have traditionally been studied in isolation: prosody (i.e., pitch accent choice and placement), syntax (i.e., word choice and order), and gesture [2, 3, 4, 5]. However, recent research reveals that prosody and gesture work in concert to signal new and focussed information [6, 7] supporting the idea of a multimodal prosody. In French, IS marking has been associated with syntactic constructions, such as fronting, clefts, and dislocations [8], with prosody primarily marking discourse structure at a higher level [9]. Nevertheless, more empirical studies on French highlight the interplay between prosody and syntax in marking focussed constituents [10, 11] and suggest that given and accessible information can be attenuated in spontaneous speech [12, 13]. In addition, a recent study investigating an audio-visual corpus of semi-spontaneous speech in French showed that given information tends to be less marked multimodally (with pitch accents and head gestures) as well as by pitch accents alone compared to inferable and new information, with most of the attenuated given elements likely being pronouns. This seems to indicate that multimodal prosody only plays a limited role for IS marking in French [14]. Using the same corpus, the present study explores further the role of multimodal prosody in the context of non-canonical (NC) syntactic constructions typically used to convey information structure in French.

Our audiovisual corpus comprises 19 video recordings of native French speakers describing their best friends (total phonation time = 19.46 minutes, mean phonation time = 58 seconds). Participants were recorded through their laptop webcams, with clear visibility of their upper body and face. Speech transcriptions were automatically extracted, corrected, and exported as a csv file, where information status (given, inferable, i.e. bridging and unused, and new) plus contrastive focus were annotated using a simplified adaptation of the RefLex system [15]. Audio from the videos was extracted and forced-aligned with the orthographic transcriptions in Praat, where pitch accents were annotated following a broadly phonetic approach [16]. Transcriptions were imported into ELAN for head gesture annotation (with hand gesture being scarce in this corpus) using the M3D scheme [17]. NC syntactic constructions (clefts, pseudoclefts, dislocations, fronting, presentational constructions) were annotated directly in ELAN following the taxonomy by Brunetti et al. [18], marking both full expressions and their lexical "heads", that is, the main content-bearing words within the NC construction that carry the core meaning of the expression.

Preliminary analysis reveals that the frequency of multimodal marking substantially increases for all IS categories (including contrastive focus) in NC constructions compared to the whole corpus (Figure 1). This finding suggests that prosody, gesture, and syntax combine to signal information structure in French. Importantly, even given information (previously mentioned referents and lexical items) can be highlighted in NC syntactic constructions and are generally not prosodically attenuated. Further annotations and analysis of the data will help refine these findings by comparing IS marking in non-canonical vs. canonical constructions (i.e., standard word order) and examining whether multimodal IS marking varies according to the position of the element in the accentual phrase and the intonational phrase.

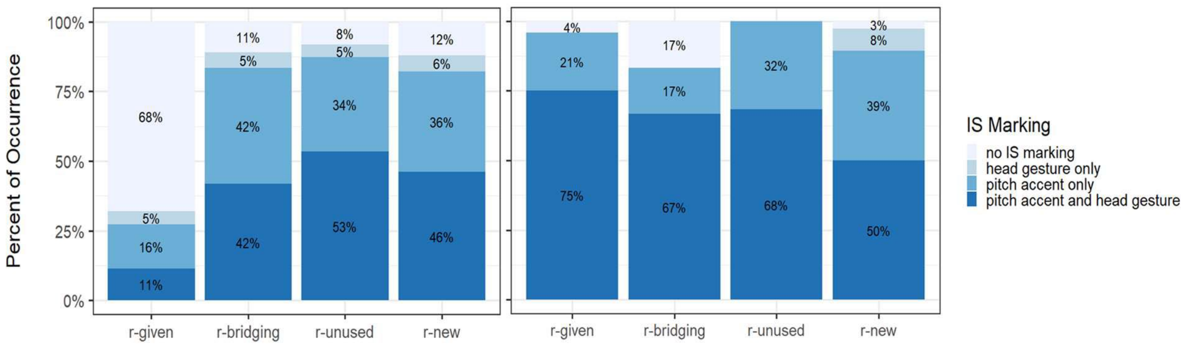


Figure 1. *Percentage of IS marking of referents in the whole corpus (left, N=1096) and on the "heads" of non-canonical syntactic constructions only (right, N=90).*

[1] K. Lambrecht, Information structure and sentence form. Cambridge University Press, 1994.
[2] N. Erteschik-Shir, Information structure: The syntax-discourse interface. Oxford University Press, 2007.
[3] F. Kügler, & S. Calhoun, S. "Prosodic encoding of information structure: A typological perspective," in C. Gussenhoven & A. Chen (eds.), The Oxford handbook of Language prosody. Oxford Academic, p. 454–467, 2020.
[4] S. Debreslioska, A. Özyürek, M. Gullberg, & P. Perniss, "Gestural viewpoint signals referent accessibility," Discourse Processes, 50, p. 431–456, 2013.
[5] C. Ebert, S. Evert, & K. Wilmes, "Focus marking via gestures," in I. Reich, E. Horch, & D. Pauly (eds.), Proceedings of Sinn und Bedeutung 15, p. 193–208, 2011.
[6] S. Im & S. Baumann, "Probabilistic relation between co-speech gestures, pitch accents and information status," Proceedings of the LSA, 5, p. 685–697, 2020.
[7] P. Rohrer, A temporal and pragmatic analysis of gesture-speech association: A corpus-based approach using the novel MultiModal MultiDimensional (M3D) labeling system, PhD dissertation, Universitat Pompeu Fabra, 2022
[8] W. Klein, "The information structure of French," in M. Krifka and R. Musan (eds.), The expression of information structure. De Gruyter, p. 95–126, 2012.
[9] S.-A. Jun, & C. Fougeron, "A phonological model of French intonation," in A. Botinis (Ed.), Intonation: Analysis, modeling and technology. Springer, p. 209–242, 2000.
[10] E. Delais-Roussarie & A. Rialland, "Metrical structure, tonal association and focus in French," in S. Baauw, F. Drijkoningen, & M. Pinto, Romance languages and linguistic theory 2005: Selected papers from 'Going Romance'. John Benjamins, p. 73–98, 2007.
[11] C. Féry, "Focus and phrasing in French," in C. Féry & W. Sternefeld (ed.), Audiatur Vox Sapientiae: A Festschrift for Arnim von Stechow. Akademie Verlag, p. 153–181, 2001.
[12] C. Portes & U. Reyle, "Combining syntax and prosody to signal information structure: the case of French", in Speech Prosody 2022, May 2022, Lisbon, Portugal.
[13] Dohen & H. Loevenbruck, H. "Pre-focal rephrasing, focal enhancement and postfocal deaccentuation in French," Proceedings of Interspeech 2004, p. 785–788, 2004.
[14] Authors, 2024
[15] A. Riester & S. Baumann, "The RefLex Scheme – Annotation Guidelines," SinSpeC, Working Papers of the SFB 732, vol. 14. University of Stuttgart, 2017.
[16] J.I. Hualde & P. Prieto, "Towards an international prosodic alphabet (IPrA)," Laboratory Phonology, 7, 5, 2106.
[17] P. Rohrer, U. Tütüncübasi, I. Vilà-Giménez, … & P. Prieto, "The MultiModal MultiDimensional (M3D) labeling system". doi: 10.17605/OSF.IO/ANKDX

[18] L Brunetti, S. Bott, … & E. Vallduví, “A multilingual annotated corpus for the study of Information Structure 1,” Dritte internationale Konferenz Grammatik und Korpora, 2009