5th Phonetics and Phonology in Europe conference

June 2 - 4, 2023
Radboud University, Nijmegen

Book of abstracts

https://pape-conference.org

PaPE

NVFW Nederlandse Vereniging voor Fonetische Wetenschappen

CLS | Centre for Language Studies
Radboud University

lab phon association for laboratory phonology

Contents

WELCOME

We are very happy to welcome you to the 5th *Phonetics and Phonology in Europe* conference (**PaPE 2023**), the first PaPE to be held in person since the start of the COVID-19 pandemic. PaPE 2023 is organized by staff at the **Centre of Language Studies** (CLS) of **Radboud University** in **Nijmegen** and follows the successful online organization of PaPE in Barcelona (2021), and three previous conferences, in Lecce (2019), Cologne (2017), and Cambridge (2015).

In terms of coverage, similar to its predecessors, PaPE 2023 is an interdisciplinary forum for research on speech, covering formal as well as empirical approaches. The **theme** of **PaPE 2023** is **"Variation in the Wild"**, a topic that aims at showcasing research taking place at Radboud and highlighting the need for methodologies, techniques, and approaches that allow speech research to move out of the laboratory and handle the challenges of speech production, perception, processing, and acquisition in ecologically valid settings. We were very pleased to see a strong response to our theme. As a result, the conference includes talks and presentations on research and new methodologies that deal with speech in the wild, while not neglecting more traditional, and equally valid, studies. Indeed, the programme of PaPE 2023 – which includes 33 oral presentations and 70 posters – is extremely rich, covering acoustics, perception, articulation, sociophonetics, acquisition, individual variation, multimodality in speech, as well as formal phonology. We are very pleased to witness this richness of perspectives that testifies to the vitality of the field.

This reflected in the **keynotes** as well: we are very happy and honoured to have presentations by **Professor Andries Coetzee**, **Professor Tamara Rathcke**, and **Dr. Esther de Leeuw**, who will cover different aspects of "wild speech", from language contact to individual variation to multilingualism.

Finally, the richness of PaPE 2023 is also illustrated in the theme of the **satellite workshop on metaphony** which highlights the importance of both formal and empirical approaches to speech. Thus, we hope that PaPE 2023 will be a lively forum for dialogue among researchers approaching speech from different perspectives and will contribute to the study of speech in its natural habitat. To further facilitate dialogue, we have kept the poster sessions non-thematic, so that delegates have the opportunity to learn about the work of other authors with similar interests.

In the interest of **sustainability** and in line with the Radboud Sustainable initiative, we will not provide bags, paper, or electronic devices as gifts for delegates. In addition, we ask you to hand in your name badge at the end of the conference so that it can be reused in the future. We hope that you will support us in our sustainability goals.

PaPE 2023 would not be possible without our **attendees** and numerous **reviewers** whom we wish to **warmly thank** for their contributions. Finally, PaPE 2023 owes a great deal to our many **sponsors**, Radboud's CLS and Internationalization Office, ISCA, LabPhon, and NVFW (Nederlandse Vereniging voor Fonetische Wetenschappen, Dutch Association for Phonetic Sciences). We are most grateful for their generous support.

We wish you all a fruitful and exciting conference!

The organising committee

Amalia Arvaniti, CLS, Radboud Universiteit Nijmegen
Mirjam Broersma, CLS, Radboud Universiteit Nijmegen
Lieke van Maastricht, CLS, Radboud Universiteit Nijmegen
Marc van Oostendorp, CLS, Radboud Universiteit Nijmegen

Committees

Local organizing committee

Amalia Arvaniti, CLS, Radboud Universiteit Nijmegen
Mirjam Broersma, CLS, Radboud Universiteit Nijmegen
Lieke van Maastricht, CLS, Radboud Universiteit Nijmegen
Marc van Oostendorp, CLS, Radboud Universiteit Nijmegen

Local advisory committee

Louis ten Bosch, CLS, Radboud Universiteit Nijmegen
Mirjam Ernestus, CLS, Radboud Universiteit Nijmegen
Stella Gryllia, CLS, Radboud Universiteit Nijmegen
Esther Janse, CLS, Radboud Universiteit Nijmegen
Jiseung Kim, CLS, Radboud Universiteit Nijmegen
Riccardo Orrico, CLS, Radboud Universiteit Nijmegen

Standing committee

Gorka Elordieta, Universidad del País Vasco
Sónia Frota, University of Lisbon
Barbara Gili-Fivela, Università del Salento
Martine Grice, Cologne University
Brechtje Post, University of Cambridge
Pilar Prieto, ICREA Universitat Pompeu Fabra
Joaquín Romero, Universitat Rovira i Virgili
Maria Josep Solé, Universitat Autònoma de Barcelona
Marina Vigário, University of Lisbon

Program at a glance

Thursday 1 June 2023, Erasmus Building, E2.54	
09:00 - 17:30	**Satellite Workshop** **Metaphony: Theoretical, descriptive and typological issues** **Organisers: Michela Russo & Rachel Walker**
17:30 - 18:30	**Welcome Reception & Registration** The Refter, Erasmus Building

Friday 2 June 2023, Berchmanianum	
08:50 - 09:20	**Registration**
09:20 - 09:30	**Opening Ceremony**
09:30 - 10:30	**Keynote: Andries Coetzee**
10:30 - 11:30	**Oral session: Variation in Production & Perception**
11:30-11:50	**Coffee Break (20 minutes)**
11:50-13:10	**Oral session: Articulation**
13:10-14:10	**Lunch Break (60 minutes)**
14:10-15:30	**Oral session: Phonetic Variation**
15:30-15:50	**Coffee Break (20 minutes)**
15:50-17:10	**Oral session: Word-level Prosody**

Saturday 3 June 2023, Berchmanianum	
09:00 - 10:00	**Keynote: Tamara Rathcke**
10:00 - 11:20	**Coffee Break and Poster Session 1 (80 minutes)**
11:20 - 12:40	**Oral session: Prosody & Individual Differences**
12:40 - 13:40	**Lunch Break (60 minutes)**
13:40 - 15:00	**Oral session: L2 Production & Perception**
15:00 - 16:20	**Coffee Break and Poster Session 2 (80 minutes)**
16:20 - 17:20	**Oral session: Prosody Perception**
19:30 - 21:30	**Conference dinner** De Waagh restaurant, Nijmegen

Sunday 4 June 2023, Berchmanianum	
09:00 - 10:00	**Keynote: Esther de Leeuw**
10:00 - 11:20	**Coffee Break and Poster Session 3 (80 minutes)**
11:20 - 12:40	**Oral session: Language Acquisition & Development**
12:40 - 13:40	**Lunch Break (60 minutes)**
13:40 - 14:40	**Oral session: Multimodal Prosody**
14:40 - 15:00	**Coffee Break (20 minutes)**
15:00 - 15:15	**Closing Ceremony + Student Awards**

INVITED SPEAKERS

(in order of appearance in the programme)

Andries Coetzee

Andries Coetzee is a professor of linguistics at the University of Michigan. His research focuses on variation in speech production and perception, the structure and limits of such variation, and how such variation contributes to sound change. He often relies on data from Afrikaans to inform his research.

The Perception-Production Link: Three Case Studies

Although phonetic theory assumes a link between perception and production, the exact nature of this link is not well-understood. In this talk, I will present three case studies that investigate the nature of this link, probing questions such as the following: (i) At what level does this link exist? At the level of the speech community, the individual speaker-listener, or both? (ii) What factors mediate the strength of the link? Can the link be relaxed (or even severed) in the context of an ongoing sound change or in socially structured variation? Case Study 1: The first case study focuses on socially neutral variation in the extent of coarticulatory nasalization in American English, and establishes evidence for the perception-production link at the level of both the speech community and the individual speaker-listener. Case Study 2: The second case study also investigates coarticulatory nasalization, but this time in an Afrikaans speech community where the extent of nasalization is socially structured. This study finds results similar to that for American English, showing that social structuredness of variation doesn't necessarily impact the nature of the perception-production link. Case Study 3: The third case study investigates an ongoing process of tonogenesis in Afrikaans. The results show that there are members of the speech community for whom perception and production are not well-aligned, indicating that the perception-production link can be weakened in the context of an ongoing sound change.

Tamara Rathcke

Tamara Rathcke is a professor of English Linguistics at the University of Konstanz. Her research focuses on speech prosody, its phonological representations, and variability across languages, dialects, and populations, and asks what speech prosody can reveal about human cognition. She often talks to other disciplines before making methodological choices.

Individual Variation *is* the Wild

Even though individual variation is ubiquitous and has been documented in a growing number of studies, research into the nature of phonological contrasts and their phonetic correlates tends to treat participants as a homogenous group. In this talk, I will argue that many conflicting findings in speech production and perception that have been considered controversial may be resolved once individual variation in participants' performance has been adequately captured. I will focus primarily on controversial findings in rhythm research (such as p-centre, perceptual regularisation, and rhythmic typology) and discuss recent evidence that is starting to shed light on the individual origins of the controversies surrounding speech rhythm. While I recognise that a hypothesis-driven sampling of individual characteristics and traits remains a challenge across many subfields of the study of speech, I will highlight a number of instruments available for capturing individual abilities relevant to prosody research and outline some best practices in considering individual variability by design of production and perception experiments.

Esther de Leeuw

Esther de Leeuw is Reader in Experimental Linguistics and Phonetics at Queen Mary University of London. Her research concerns the production, processing, representation, and development of speech sounds in the context of multilingualism and language contact, the cognitive organization of dual phonological systems, and the phonetic and phonological interactions which occur during bilingualism, L1 attrition and L2 acquisition.

Variation in phonetic and phonological L1 attrition

Within the context of bilingualism, L1 attrition is generally defined as change in an individual's native language, i.e. L1, upon acquisition of an L2 post adolescence, and is most often investigated within the scope of long-term immigration. Research into phonetic and phonological L1 attrition (henceforth attrition) is of great importance to our understanding of the plasticity of native speech as well as how languages and dialects interact and are represented in the mind across the lifespan. My presentation will discuss a number of studies which suggest that different phonetic and phonological variables are not affected equally by the process of attrition, i.e. some phonetic and phonological variables undergo more attrition than others. The results from these studies support the understanding that variation in attrition is modified by sociolinguistic factors.

ORAL SESSIONS

(in order of appearance in the programme)

Variation in Production & Perception

Friday, oral session 1

A corpus study of naturalistic misperceptions in German sung speech

Jessica Nieder[a] & Kevin Tang[b]
nieder@phil.hhu.de & Kevin.Tang@hhu.de
[a]Department of Linguistics & [b]Department of English Language and Linguistics,
Heinrich-Heine-Universität Düsseldorf, Germany

When listening to a conversation, the listener's task is to identify speech segments and subsequently segmenting the continuous speech stream they receive into meaningful word forms. In some cases, however, the segmentation process fails and results in naturally-occurring instances of misperceptions, also called *slips of the ear* (Tang, 2015; Tang & Nevins, 2014). These slips of the ear give us insights into speech perceptual processes and were investigated in corpus-based and/or experimental studies over the past years (e.g. Bond, 1999; Cutler & Butterfield, 1992; Marxer et al., 2016; Tang, 2015; Tang & Nevins, 2014; Tóth, 2017; Vroomen et al., 1996). Another source of naturalistic *slips of the ear* is given when focusing on misperceiving sung speech, resulting in misperceptions such as "It doesn't make a difference if we're naked or not" for the line from Bon Jovi's Livin On a Prayer "It doesn't make a difference if we make it or not" (Asaridou & McQueen, 2013; Hirjee & Brown, 2010).

In this study, we report on a corpus of 176 instances of misheard German sung speech. To confirm the ecological validity of our corpus, we examined segmental confusions and word mis-segmentations. The approximately 1,000 naturalistic segment confusions found in this study were significantly correlated with vowel acoustic distances from Sendlmeier and Seebode (2006) (r = 0.559) as well as with speech-in-noise-induced confusions from an experimental study presented in Jürgens and Brand (2009) (vowel: r = 0.364; consonant: r = 0.210), suggesting that confusions from misheard sung speech are strongly influenced by phonetics. Given that the experimental conditions differ greatly from the naturalistic one, such as background noise in the form of instrumentals or beat, this is a surprising finding indicating that despite influences of music on sung speech, listeners still heavily rely on acoustic similarity during processing.

Our mis-segmentation patterns, however, only partially confirmed the rhythmic segmentation hypothesis presented by Cutler and Butterfield (1992) and findings from previous studies on misperceptions in German sung speech such as Kentner (2015). While boundaries inserted before strong syllables created content words following the preferred rhythmic properties of German – confirming the rhythmic segmentation hypothesis – we find a surprising amount of boundary deletion before strong syllables as compared to weak syllables – against the rhythmic segmentation hypothesis. A great amount of these deletions results in humorous nonce percepts which highlight the importance of affective signals in perception (Beck et al., 2014) and might be the consequence of the experience of listeners with neologisms in song lyrics (Werner, 2012).

To sum up, in this study we are able to demonstrate the role of bottom-up (phonetics) and top-down (lexical expectation) factors in the processing of sung speech, despite the relatively small size of our corpus and the inherent differences of listening to spoken speech vs. sung speech.

References

Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Frontiers in Psychology*, *4*, 321. https://doi.org/10.3389/fpsyg.2013.00321

Beck, C., Kardatzki, B., & Ethofer, T. (2014). Mondegreens and Soramimi as a Method to Induce Misperceptions of Speech Content – Influence of Familiarity, Wittiness, and Language Competence. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0084667

Bond, Z. (1999). *Slips of the ear: Errors in the perception of casual conversation*. Brill.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236. https://doi.org/10.1016/0749-596X(92)90012-M

Hirjee, H., & Brown, D. G. (2010). Solving misheard lyric search queries using a probabilistic model of speech sounds. In J. S. Downie & R. C. Veltkamp (Eds.), *Proceedings of the 11th international society for music information retrieval conference* (pp. 147–152). International Society for Music Information Retrieval.

Jürgens, T., & Brand, T. (2009). Microscopic prediction of speech recognition for listeners with normal hearing in noise using an auditory model. *The Journal of the Acoustical Society of America*, *126*(5), 2635–2648.

Kentner, G. (2015). Rhythmic segmentation in auditory illusions - evidence from cross-linguistic mondegreens. *Proceedings of 18th ICPhS*.

Marxer, R., Barker, J., Cooke, M., & Garcia Lecumberri, M. L. (2016). A corpus of noise-induced word misperceptions for english. *The Journal of the Acoustical Society of America*, *140*(5), EL458–EL463.

Sendlmeier, W. F., & Seebode, J. (2006). *Formantkarten des deutschen Vokalsystems* (tech. rep.). TU Berlin, Institut für Sprache und Kommunikation.

Tang, K. (2015). *Naturalistic speech misperception* (Doctoral dissertation). University College London.

Tang, K., & Nevins, A. (2014). Measuring segmental and lexical trends in a corpus of naturalistic speech. In H.-L. Huang, E. Poole, & A. Rysling (Eds.), *Proceedings of the 43rd meeting of the north east linguistic society* (pp. 153–166). GLSA (Graduate Linguistics Student Association).

Tóth, M. A. (2017). *A microscopic analysis of consistent word misperceptions* (Doctoral dissertation). Universidad del Paıś Vasco.

Vroomen, J., Van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, *24*(6), 744–755. https://doi.org/10.3758/BF03201099

Werner, V. (2012). Love is all around: A corpus-based study of pop lyrics. *Corpora*, *7*(1), 19–50. https://doi.org/10.3366/cor.2012.0016

Explaining variation in three dialects of Meru, Kenya

Conceição Cunha[1], Fridah Kanana[1,2], Jonathan Harrington[1]
[1]*Institute of Phonetics, LMU Munich, Germany* [2]*Kenyatta University, Kenya*

The main aim of this study is to document the phonetic variation for Kenyan dialects of the Meru-Tharaka group, with a particular focus on the inclusion of unstudied varieties. The focus is on the morphophonological variation in the palatalization of the plural prefixes in class 8 nouns in three dialects of the Meru language, of Bantu origin, and spoken on the north eastern slopes of Mount Kenya: Imenti, which is considered the standard variety, Tigania in the northern and Chuka in the southern. Imenti has the most developed literature and is also used in formative years of schooling. The literature development in Chuka is more recent, while Tigania exists almost exclusively in oral forms and social media. In earlier studies of the dialects of the Meru-Tharaka group, Kanana ([1-3]) showed that the consonant of the plural prefix derived from a proto-Bantu bilabial stop /*bi/ ([4]) is produced with different kinds of palatals in Imenti and Chuka. The so far unstudied region of Tigania which is part of the present study was predicted to be predominantly influenced by the geographical proximal Imenti dialect region (with which it shared a border to the north and because Imenti has some characteristics of a standard accent).

For the present study, the focus of the analysis was on five pairs of singular-plural class 7/8 nouns that were recorded in 2022 from 75 multilingual adult speakers from Chuka (n = 26, 14F), Imenti (n = 23, 6F), and Tigania (n = 26, 9F). The participants provided informed written consent and were compensated for their participation. The experiment was reading task consisting of a randomized order of 2-3 repetitions of 96 words, with one word at a time presented on a computer monitor using SpeechRecorder [5]. Since the participants were educated in English and Swahili and we wanted to avoid orthographic forms from these dialects, the words were presented in both English and in Swahili and the task was to produce the equivalents in the local dialect. Words were repeated if the participants gave an incorrect equivalent (e.g., 'woman' or 'small girl' for the targeted 'girl'). The singular and plurals forms were presented together. Following an orthographic transcription by a native speaker of each variety, the speech signals were forced-aligned with WebMAUS [6] and manually corrected. The output was structured into a speech database using EMU-SDMS [7] for further processing.

Consistently with earlier findings [1-3], the results showed the plural prefix was produced with palatalised labials in Imenti, but with palatalised lingual consonants in Chuka. The prefix in Tigania (Fig.1) ranged over all these places of articulation. A further analysis of Tigania showed two main findings: (1) participants were more likely to produce a palatalised labial when the investigator conducting the experiment was an Imenti speaker (Fig. 2) whereas lingual consonants were more likely when the investigator was from the Tigania region, (2) within lingual prefixes, Tigania participants preferred dorsal /ɕ, tɕ/ whereas apical prefixes /sj, ts, tʃ, ʃ/ were more likely in Chuka.

We interpret the first finding as a form of style-shifting [8] in which the Tigania participants adapted their speaking style towards the dialect of the investigator resulting in more productions of an Imenti-style palatalised labial prefix with the Imenti investigator. To explain the second finding, we assume that both Chuka and Tigania have undergone a sound change of labial palatalisation [references] in which palatalised labials became palatal consonants with a dorsal constriction /ɕ, tɕ/. Chuka may then have introduced a further innovation by which these dorsals have undergone velar palatalisation [9, 10] resulting in prefixes with an apical constriction. Chuka could be at the forefront of this change given that it is geographically and administratively more removed from Imenti than Tigania: this could explain why apical prefixes are more common in Chuka than in Tigania. Further analyses of the many other dialects in this region are necessary to further substantiate this hypothesis.

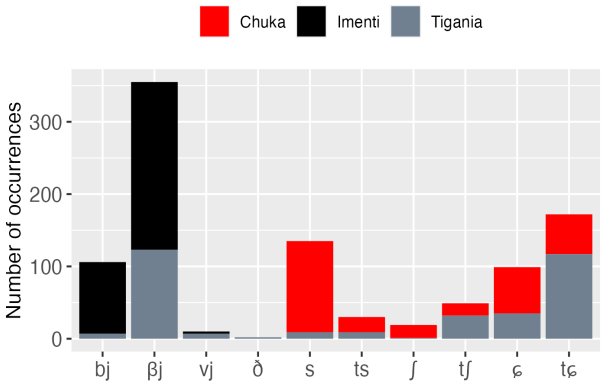


Figure 1. *A count of the different forms of the initial plural prefix consonant in three Meru dialects.*

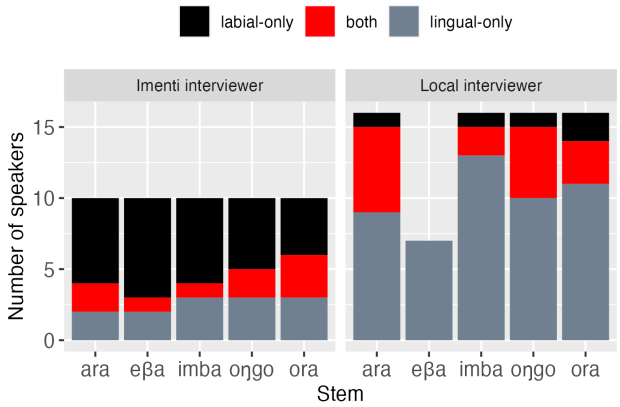


Figure 2. *Number of Tigania speakers by interviewer and stem who produced labial, lingual or both types of plural prefixes*

Selected references

[1] Kanana, F.E. 2011a. Dialect convergence and divergence: A case of Chuka and Imenti. *Selected Proceedings of the 40th Annual Conference on African Linguistics*, ed. Eyamba G. Bokamba et al., 190-205. Sommerville, MA: Cascadilla Proceedings Project.

[2] Kanana, F.E. 2011b. Meru dialects: The linguistic evidence. *Nordic Journal of African Studies* 20(4): 300–327.

[3] Kanana, F.E. 2014. *Lexical-phonological comparative analysis of selected dialects of the Meru-Tharaka Group*. Frankfurt: Peter Lang, edited by Rainer Vossen, doi: 10.3726/978-3-653-05161-2

[4] Guthrie, M. 1967-71. Comparative Bantu: An introduction to the comparative linguistics and pre-history of Bantu languages. 4 vols. Farnborough: Gregg.

[5] Draxler, C. & Jänsch, K. 2004. SpeechRecorder – A universal platform independent multi-channel audio recording software, *Proc. of the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal, 559–562.

[6] Kisler, T., Reichel, U. D. & Schiel, F. 2017. Multilingual processing of speech via web services, *Computer Speech & Language*, vol. 45, pp. 326–347, doi: 10.1016/j.csl.2017.01.005.

[7] Winkelmann, R., Harrington & J., Jänsch, K. 2017 EMU-SDMS: Advanced speech database management and analysis in R, *Computer Speech & Language*, vol. 45, 392–410, doi: 10.1016/j.csl.2017.01.002.

[8] Hay J, Jannedy S & Mendoza-Denton N. 1999. Oprah and /ay/: lexical frequency, referee design and style. *Proc. of the 14th international congress of phonetic sciences*, San Francisco.

[9] Bateman, N. On the typology of palatalization. *Language & Linguistic Compass*, 5, 588-602.

[10] Krämer, M. & Urek, O. 2016. Perspectives on palatalization. *Glossa*, 1(1): 31. 1–17.

Selective adaptation between allophones of /r/ in German

Holger Mitterer[1,2] and Eva Reinisch[2]
[1]*University of Malta,* [2]*Hanyang University,* [3]*Austrian Academy of Sciences*

Selective adaption (SA) is an experimental paradigm that is often used to investigate pre-lexical representations of the speech signal [1]. In this paradigm, participants are exposed to a series of adaptor stimuli before categorizing a test stimulus from a continuum between two speech sounds. SA seems to reflect general perceptual principles, with counterparts in visual perception, such as the waterfall illusion [2], in which observers watch a waterfall for around 30s and then perceive stationary objects moving upwards. Such a contrastive effect is also observed in speech perception, with listeners perceiving the test stimuli as contrasting with the adaptors. That is, after hearing a serious of /b/-initial words, a stimulus from a [ba]-[da] continuum is more likely to be perceived as /da/, contrasting with the /b/-initial adaptors [3].

The paradigm has had a waxing and waning popularity [3] and critics pointed out potential post-perceptual influences on SA [1]. More recently, SA has been used to investigate pre-lexical representations in spoken-word recognition with the rationale that SA between adaptors in coda position and test stimuli in onset position would reflect position-invariant phonemic representations at a pre-lexical level. While one study found such an effect [4], others ([5], [6]) did not and pointed out phonetic confounds in [4]. One critical finding was that there was no adaptation between word-initial trilled /r/ and a sonorant /r/ in the coda position in Dutch [6]. Here we present three SA experiments, conducted online with about 30 participants each (advertised via *prolific.co*), that aim to replicate and extend this finding in German. Similar to Dutch, German has a large variety of allophones of /r/, including uvular and alveolar trills and a vocalized /r/ in the coda position (e.g., *Fischer*, Engl., 'fisherman', [fiʃɐ]).

The first experiment used an alveolar trill-lateral continuum ([rozə]-[lozə], Engl. 'rose'-'lottery tickets') as target stimuli, and different adaptor series containing either alveolar trills [r] (the maximal overlap with the test stimuli), uvular fricatives [ʁ], or vocalized versions of /r/ ([ɐ]). A control-adaptor condition was generated from words that did not contain any variant of /r/ or /l/. Results replicated [5] for Dutch, such that the most sonorant adaptor, the vocalized [ɐ], did not trigger any SA for the test stimuli containing an alveolar trill (with a Bayes Factor supporting the null over an alternative hypothesis, see Table 1 for a summary of the results). However, the uvular fricative, though phonetically different from the trill, led to SA. Experiments 2 and 3 focused on the uvular fricative and used an [ʁ]-[h] continuum ([ʁozə]-[hozə], Engl. 'rose'-'trousers', note that in German /h/ is the phoneme closest to [ʁ]) as test stimuli. The adaptor series contained either alveolar trills, uvular trills, or uvular fricatives. In Experiment 2, /r/ in the adaptors was word-initial (e.g., [ra:t], Engl. 'advice'), while in Experiment [3], it was word-medial but still in the syllable onset (e.g., *Barock,* Engl., 'baroque', [barɔk]). In both experiments, the surprising result was that the [r] adaptors caused stronger adaptation effects on the uvular-fricative stimuli from the test continuum than [ʁ] adaptors. In fact, in Experiment 3, the [ʁ] adaptors, with the same allophone as the test stimuli, even failed to produce any selective adaptation at all, while the trill adaptors did.

Overall, the results show that phonemic overlap is not sufficient to generate SA, which questions the assumption of phonemic representations at a pre-lexical level. However, in some cases SA is observed between different allophones, with the surprising result that alveolar trilled /r/ leads to stronger adaptation for both alveolar-trill and uvular-fricative test stimuli. This indicates that SA, as early critics already suggested [1], may also arise at later, post-perceptual (rather than prelexical) levels of processing. For [r], this may be due to saliency of the amplitude modulation in trills [7]. With such post-perceptual influences, selective adaptation may not be the ideal paradigm to reveal prelexical representations in spoken-word recognition.

Table 1: Overview of selective-adaptation effects in the current study

	/r/ target stimulus	Adaption effects		
		strongest	→	weakest
Exp1	[rozə] (alveolar trill)	[#rV…] >	[#ʁV…] >	[…Vʁ(C)#] = ∅
Exp2	[ʁozə] (uvular fricative)	[#rV…] =	[#ʀV…] >	[#ʁV…] > ∅
Exp3	[ʁozə] (uvular fricative)	[#...rV…] =	[#...ʀV…] >	[#...ʁV…] = ∅

Note: "= ∅" means that an adaptor condition is similar to the control condition according to a Bayes Factor. "#" indicates a word boundary.

References

[1] S. Harnad, *Categorical perception: The groundwork of cognition*. Cambridge.: Cambrigde University Press, 1987.

[2] A. G. Goldstein, "On the after-effects of the 'waterfall' and 'spiral' illusions," *The American Journal of Psychology*, vol. 71, pp. 608–609, 1958, doi: 10.2307/1420264.

[3] D. F. Kleinschmidt and T. F. Jaeger, "Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning?," *Psychon Bull Rev*, vol. 23, no. 3, pp. 678–691, Oct. 2015, doi: 10.3758/s13423-015-0943-z.

[4] J. S. Bowers, N. Kazanina, and N. Andermane, "Spoken word identification involves accessing position invariant phoneme representations," *Journal of Memory and Language*, vol. 87, pp. 71–83, Apr. 2016, doi: 10.1016/j.jml.2015.11.002.

[5] H. Mitterer, E. Reinisch, and J. M. McQueen, "Allophones, not phonemes in spoken-word recognition," *Journal of Memory and Language*, vol. 98, no. Supplement C, pp. 77–92, Feb. 2018, doi: 10.1016/j.jml.2017.09.005.

[6] A. G. Samuel, "Psycholinguists should resist the allure of linguistic units as perceptual units," *Journal of Memory and Language*, vol. 111, p. 104070, Apr. 2020, doi: 10.1016/j.jml.2019.104070.

[7] B. Delgutte and N. Y. S. Kiang, "Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics," *The Journal of the Acoustical Society of America*, vol. 75, no. 3, pp. 897–907, Mar. 1984, doi: 10.1121/1.390599.

Articulation

Friday, oral session 2

Articulatory variation in ejective production: A real-time MRI study of Amharic ejective plosives

Lavinia Price[1], Marianne Pouplier[1], Phil Hoole[1]
[1]*Institute of Phonetics and Speech Processing, Ludwig-Maximilians Universität Munich*

An ongoing debate surrounds the hypothesis that ejectives may result from a range of different articulatory mechanisms which may or may not involve larynx raising [1, 3, 4, 5, 6, 7, 8]. While the IPA classifies ejectives as non-pulmonic, it is increasingly recognized that ejectives may be much more variable than assumed so far and may in fact be initiated by either a glottalic or pulmonic airstream. How and under what circumstances non-glottalic initiation exactly occurs remains unclear. It is, however, increasingly evident that our knowledge of ejective production is fundamentally incomplete. This study contributes to this current debate by investigating articulatory variability in Amharic ejectives using rt-MRI.

Traditionally, ejectives have been defined as glottalic sounds produced by raising of the larynx during simultaneous constrictions at the glottis and in the oral cavity [2]. Larynx-raising reduces the supraglottal cavity, thereby increasing the intraoral air pressure (IOP) which results in the auditorily distinct quality of these sounds. The initiatory contribution of the larynx, however, is now being questioned. For one, larynx raising alone is deemed insufficiently effective at increasing IOP to justify the intense bursts found in some ejectives [3, 8]. Secondly, other ejective types appear not to rely on larynx raising at all [1, 6, 7]. Both of these observations suggest that additional strategies of supraglottal volume reduction may contribute to IOP build-up instead of larynx raising. Yet, beyond Kingston's suggestion of tongue root backing [3], it is to date completely unclear what these additional strategies may be. Overall, these findings underscore how little we know about the range of possible mechanisms by which ejectives may be produced. The current study addresses this gap by investigating the articulatory details of ejectives in Amharic, an Ethio-Semitic language of Ethiopia, using rt-MRI data from five native speakers (2 male, 3 female) acquired at 50 frames per second. Our goals are to observe the movements of the larynx and of the supraglottal articulators likely to assist in volume reduction during the ejective's closure interval. We recorded 271 lexical items containing the pulmonic and ejective plosives of Amharic at three places of articulation (/p, t, k, p', t', k'/), as singletons and geminates (e.g., /t' vs. tt'/) in three word-positions (initial, medial, final). First qualitative analyses suggest a high degree of inter- and intra-speaker variation in ejective realization, with larynx raising occurring only some of the time. We observe cases without larynx raising but with otherwise clearly reduced supraglottal cavities compared to pulmonic stops (Fig.1). This would be consistent with a higher IOP typical for ejectives. We will also present first results from a recording exploring the use of dual-slice sagittal/axial recordings to capture lateral activity of the upper pharyngeal wall, which may assist in cavity reduction.

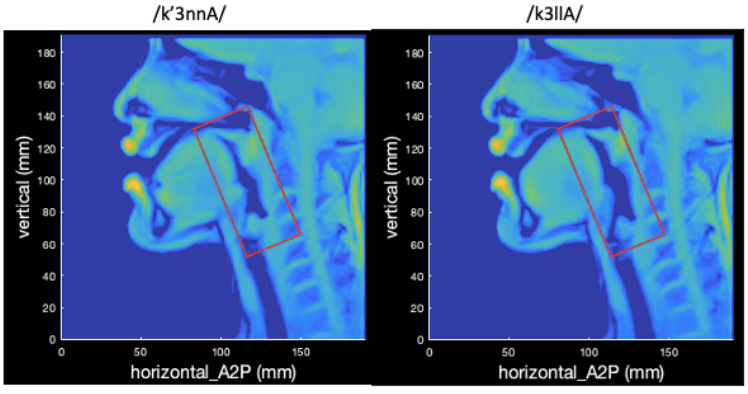


Figure 1. *Speaker 1 at the moment of velar constriction prior to burst release of the initial ejective in item /k'3nnA/, left panel, and initial pulmonic plosive /k3llA/, right panel. The supraglottal cavity volume visibly reduced on the left, compared to that on the right, suggests a higher IOP, which is a typical feature of ejectives.*

References

[1] Brandt, E., Simpson, A.P. 2021. The production of ejectives in German and Georgian. *Journal of Phonetics* 89, 101111.

[2] Catford, J. 1977. *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.

[3] Kingston, J.C. 1985. *The phonetics and phonology of the timing of oral and glottal events*. Ph.D. thesis, University of California, Berkeley.

[4] Ladefoged, P., Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

[5] Lindau, M. 1984. Phonetic differences in glottalic consonants. *Journal of Phonetics* 12, 147–155.

[6] Price, L., Pouplier, M., Hoole, P. 2022. Kehlkopfanhebung in englischen wortfinalen Ejektiven: eine Echtzeit-MRT-Studie. P&P 18, Bielefeld.

[7] Simpson, A.P. 2014. Ejectives in English and German – linguistic, sociophonetic, interactional, epiphenomenal? In: Celata, C., Calamai, S. (eds), *Advances in Sociophonetics*. Amsterdam: John Benjamins, 187–202.

[8] Simpson, A.P., Sulaberidze, N. 2022. Ejectives in English: elicitation und analysis. P&P 18, Bielefeld.

[9] Wright, R., Hargus, S., Davis, K. 2002. On the categorization of ejectives: Data from Witsuwit'en. *Journal of the International Phonetic Association* 32, 43–77.

A new articulatory tool: Comparison of EMA and MARRYS

Malin Svensson Lundmark[1,2], Donna Erickson[3], Oliver Niebuhr[2], Mark Tiede[3] and Wei-Rong Chen[3]

[1]*Lund University,* [2]*University of Southern Denmark,* [3]*Yale University*

Syllable articulation involves opening and closing the mouth. For stressed/prominent syllables, we open our mouths more, and the tongue moves accordingly to accommodate for the increased jaw opening (e.g. [1]). Babies acquire adult-like jaw articulation earlier than lip or tongue articulation ([2]). Languages vary in terms of organization of syllable prominence patterns, which is reflected in the jaw displacement patterns of speakers of a specific language (e.g. [3]). Second language learners tend to transfer their first language jaw patterns when learning a new language ([4]). Given this, jaw movement patterns, we suggest, are important underpinnings of spoken language.

The method currently used to measure jaw displacement is electromagnetic articulography (EMA) ([5]). The advantage of this method is that we can also measure the articulators used for consonant production (e.g. [6]); the disadvantage is that the expense of EMA precludes easy access of data acquisition for a large number of speakers. To remedy this, a new method for collecting jaw data is being developed at the University of Southern Denmark ([7], [8]). The new method, the MARRYS helmet, records jaw displacement via two bending sensors in the cheek straps on either side, time-aligned with the acoustic speech signal; it is economically much cheaper, and also requires drastically less preparation and processing time, thus allowing jaw data collection from a large number of speakers – and over a long time, even in the field given its mobility. By means of the two bending sensors, the MARRYS system additionally provides information about the symmetry of jaw movements. The helmet also has an attached microphone, allowing for good intensity recordings as well as a constant mouth-to-microphone distance.

This paper reports on a comparison of jaw displacement data recorded by EMA articulography and the MARRYS helmet. The co-collection of EMA and MARRYS data was done simultaneously, Fig. 1(a). The results show a very good correlation of the two sets of data. Details are explained in the full paper. Fig. 1(b) provides an example of the MARRYS signals (2 channels, showing left and right jaw movements). Each syllable provides a spike in the trajectory representing the jaw movement, and the size of the spike corresponds to jaw displacement. In Fig 1(b) larger displacement can be seen for the open vowels, for phrase-final words, but more noticeably, for the contrastive focused words “mat” and “cat” (capitalized words in fig. 1(b)).

The MARRYS helmet provides a tool for observing jaw displacement patterns in a number of situations, including basic understanding of speech articulation, clinical applications, teaching of second language prosody or public speaking, etc.

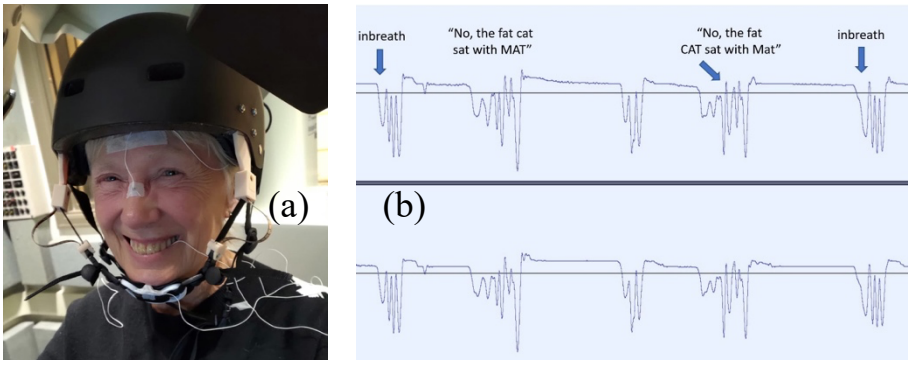


Figure 1. *(a) Jaw data co-collection with MARRYS and EMA, (b) 2-ch MARRYS signal*

References

[1] Erickson, D. 2002. Articulation of extreme formant patterns for emphasized vowels. *Phonetica* 59, 134-149.

[2] Green, J. R., Moore, C. A., Reilly, K. J. 2002. The sequential development of jaw and lip control for speech, *Journal of Speech Language and Hearing Research* 45.1., 66–79. https://doi.org/10.1044/1092- 4388(2002/005)

[3] Erickson, D. & Kawahara, S. 2016. Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistic Vanguard* 2, 102-110. De Gruyter Mouton. https://doi.org/10.1515/lingvan-2015-0025

[4] Wilson, I., Erickson, D., T. Vance, T., and Moore, J. 2020. Jaw dancing American style: A way to teach English rhythm, *Speech Prosody 2020*.

[5] Hixon, T.X. 1971. An electromagnetic method for transducing jaw movements during speech. *The Journal of the Acoustical Society of America* 49, 603-606.

[6] Svensson Lundmark, M. 2023. Rapid movements at segment boundaries. *The Journal of the Acoustical Society of America* 153, 1452-1467. https://doi.org/10.1121/10.0017362

[7] Erickson, D., Niebuhr, O., Gu, W., Huang, T., Geng, P. 2020. The MARRYS cap: A new method for analyzing and teaching the importance of jaw movements in speech production *International Seminar of Speech Production*.

[8] Niebuhr, O., & Gutnyk, A. 2021. Pronunciation engineering: Investigating the link between jaw-movement patterns and perceived speaker charisma using the MARRYS cap. *Proc. 3rd IEEE International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, Kuala Lumpur, Malaysia, 1-6.

Ultrasound measurements of turn-taking in interactive question-answer sequences: Articulatory preparation is delayed but not tied to the response

Sara Bögels[1,2] and Stephen C. Levinson[2]
[1]*Department of Communication and Cognition, Tilburg University,* [2]*MPI Nijmegen*

Previous research has shown that speech planning in conversational turn-taking can happen in overlap with the previous turn. Most of this research, using methods such as EEG [e.g., 1-3] and eye-tracking [e.g., 4], suggests planning of one's next turn starts as early as possible, that is, as soon as the gist of the current turn becomes clear. For example, [1] manipulated this position by creating quiz questions that allowed either early planning (example 1) or late planning of the answer (example 2), because the critical information, needed to answer the question (i.e., *007*), appears in the middle or at the end of the question, respectively. The EEG results showed a neural signature (positivity) starting around 500 milliseconds after the start of the critical word, which was interpreted as indicating the start of production planning. Although the studies cited above suggest an early start of planning, it is not clear how far (i.e., until what phase of production planning) and how fast language production proceeds after it has started. One study [5] showed that the 'phonological planning' phase of production planning is ongoing more than one second after critical word onset, suggesting that planning proceeds at least until this stage, but also appears to be delayed with respect to planning 'in the clear'.

The present study aimed to investigate whether early planning immediately proceeds all the way up to the last stage, which we refer to as 'articulatory preparation' (i.e., putting the articulators in place for the first phoneme of the response) and to gain more insight into the timing of this process. Participants answered pre-recorded quiz questions taken from [1], while their tongue movements were measured using ultrasound. Even though the questions were prerecorded, ensuring experimental control over when planning could start, participants were under the impression that the questions were asked live, and they indeed received live feedback from the same speaker. As illustrated by examples (1) and (2), the moment at which planning could start was manipulated to be early (i.e., midway during the question, example 1) or late (i.e., only at the end of the question, example 2). Ultrasound measurements were analyzed by calculating frame-to-frame changes in pixel hues between subsequent ultrasound images.

First, the results showed faster response times for the early-planning (example 1) than the late-planning condition (example 2) and, correspondingly, earlier changes in tongue movements with respect to question offset in the early than in the late condition. More importantly, tongue movements did not differ between the early and late conditions for at least two seconds after planning could start in early-planning questions (see Figure 1), suggesting that speech planning in overlap with the current turn proceeds (much) more slowly than in the clear. On the other hand, when time-locking to speech onset and looking backwards (Figure 2), tongue movements differed between the two conditions from up to two seconds before this point. This latter result suggests that articulatory preparation can occur a few seconds in advance of articulation and is not fully tied to the overt articulation of the response itself.

In conclusion, the present study shows, first, that (preparatory) articulatory movements can be measured using ultrasound in an interactive turn-taking setting. In future research, its application may be expanded to even more natural, conversational turn-taking situations. Furthermore, the results show that early planning during interactive turn-taking can proceed all the way up to articulatory preparation in the form of tongue movements, and arrive there some time (here: up to two seconds) before response onset. This is in accordance with earlier observational work [6] suggesting that articulatory preparation (in that case in the form of lip apertures) is not strictly tied to the response itself. However, it also adds to the body of evidence showing that the planning process is clearly drawn out and less efficient when performed in a turn-taking situation, where it overlaps with understanding of the ongoing turn.

Examples:

(1) Which character, also called [1]**007**, appears in the famous movies? [2]*James Bond*

(2) Which character from the famous [1]movies, is also called **007**? [2]*James Bond*

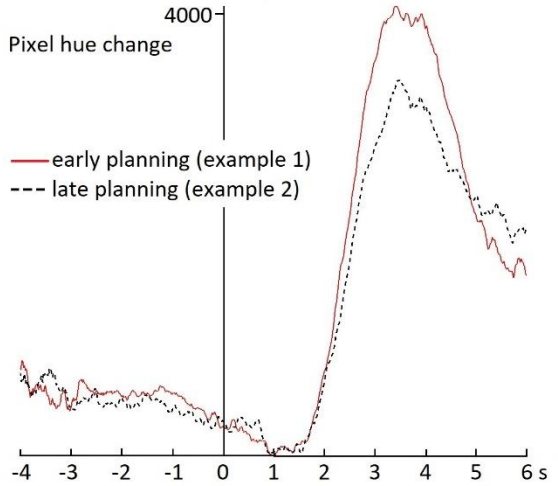


Figure 1. *Ultrasound results (frame-to-frame change in pixel hues) relative to onset of a word in the middle of the question (0-point; [1] in the examples). For the early-planning condition (red solid line), the 0-point corresponds to the onset of the critical word (i.e., onset of* 007 *in example 1). For the late-planning condition (black dashed line) the 0-point corresponds to an equivalent position (onset of* movies *in example 2), where planning cannot start yet.*

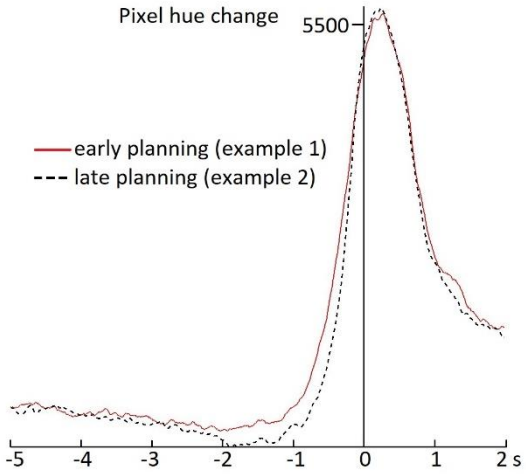


Figure 2. *Ultrasound results (frame-to-frame change in pixel hues) relative to response onset (0-point; [2] in the examples) for the early-planning (red solid line) and late-planning (black dashed line) conditions.*

.

References

[1] Bögels, S., Magyari, L., & Levinson, S.C. 2015. Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports* 5, 12881.

[2] Bögels, S., Casillas, M., & Levinson, S.C. 2018. Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia* 109, 295–310.

[3] Bögels, S. 2020. Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition* 203, 104347.

[4] Barthel, M., Sauppe, S., Levinson, & S.C., Meyer, A.S. 2016. The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology* 7.

[5] Barthel, M., & Levinson, S.C. 2020. Next speakers plan word forms in overlap with the incoming turn: evidence from gaze-contingent switch task performance. *Language and Cognitive Neuroscience* 35(9), 1183–202.

[6] Krause, P.A., & Kawamoto, A.H. 2021. Predicting one's turn with both body and mind: Anticipatory speech postures during dyadic conversation. *Frontiers in Psychology* 12.

Changes in manner of articulation explained by aerodynamic factors

Maria-Josep Solé
Universitat Autònoma de Barcelona, Spain

Changes in manner of articulation of consonants –found in allophonic variation, phonological alternations, and sound change– are traditionally attributed to differences in degree of oral constriction or stricture. Thus changes in manner are described in terms of 'lenition' (a more open and/or shorter articulatory constriction) or 'strengthening' (a closer and/or longer constriction). These terms, however, are used for a multiplicity of phenomena (e.g. 'strengthening' is used to refer to devoicing, ejectivization, aspiration, stop insertion, occlusivization, frication and affrication). Despite the various definitions and outcomes of 'strengthening' and 'lenition' [1], what these phenomena have in common is that they involve higher- or lower-than-normal oral pressure build up (strengthening and lenition, respectively). In this paper it is argued that a great number of changes in manner of articulation affecting consonants (specifically, obstruents and sonorants) can be accounted for by aerodynamic factors.

Many cases of changes in manner of articulation are indeed due to variation in the magnitude (and duration) of the oral constriction due to segmental context or coarticulation with adjacent segments (e.g., [2], [3]), syllable position [4], position in the sentence [5], and prosodic factors [6]. However, changes in manner may also result from changes in glottal aperture or velopharyngeal closure without any variation in the supraglottal constriction. This is because the vocal tract is a system of interconnected mechanisms (glottal, nasal, oral) that influence each other, that is, local articulatory gestures at the glottis or the velum have consequences on the pressure and flow at the supraglottal constriction.

We examine patterns involving loss/generation of audible frication, stops bursts and tongue tip trilling with variation in the aerodynamic conditions due to the action of distant gestures (glottal action, nasal aperture). Thus, without any variation in the configuration of the oral constriction, fricatives may become approximants and approximants may become fricatives, just by changes in glottal aperture. Reduced transglottal flow through the vibrating glottis for voiced (compared to voiceless) fricatives results in a lower intensity of noise generated at the oral constriction which may be perceptually missed. Hence voiced fricatives tend to weaken into glides and approximants (example 1) or rhotics (example 2), and are lost earlier than voiceless fricatives. By the same principle, approximants, glides and high vowels which are usually voiced may become fricatives if devoiced. This is due to the increased airflow (caused by the larger glottal opening) passing through the oral constriction for these segments (example 3). Other obstruents sounds, such as stops, affricates and tongue-tip trills present comparable effects of voicing (i.e., degree of glottal aperture) on manner.

Changes in velopharyngeal opening also have an impact on manner. Concurrent and coarticulatory nasalization, which allows the airflow to escape through the nasal cavity and reduces the pressure difference across the oral constriction, impairs strong supralaryngeal frication, high intensity noise bursts and tongue-tip trilling. Thus fricatives preceding nasals may become glides or be lost (example 4) due to anticipatory velum opening. By contrast, fricatives following nasals tend to develop a transitional epenthetic stop (due to an early raising of the velum during the acoustic duration of the nasal) which may result in affrication (example 5). Though these changes may be accounted for by differences in interarticulatory coordination, the generation/loss of frication or stop bursts is an aerodynamic event. Similarly, the strong release burst characteristic of voiceless stops is impaired by a following nasal, resulting in stop voicing, nasal assimilation (or replacement by a glottal stop to preserve the high intensity burst) (ex. 6).

Understanding the specific aerodynamic conditions for a given manner of articulation and the trade-offs between articulator movements and aerodynamic forces, and their acoustic result, (i) allows us to provide a unified account of changes in manner of articulation (i.e., including not only changes in stricture), and (ii) questions notions such as the scale of 'consonantal strength' defined exclusively in terms of 'articulatory constriction' or 'gestural strength' [7].

Examples cited

(1) Voiced fricative weakening or gliding
Old English *næ**g**l* [ʝ] > *na**i**l*, *to**g**ian* [ɣ] > *to**w***

Welsh 'soft mutation', voiceless fricative [ɬ] and [ɽ] > voiced sonorant [l], [r]
***rh**ywbeth* [ɽ] 'something' *ei **r**ywbeth* [r] 'his something'
***ll**yfr* [ɬ] 'book' *ei **l**yfr* [l] 'his book'

(2) Rhotacism of voiced fricatives
Old English *we**s**an* > *wa**s** – we**r**e*, *forleo**s**an* > *lo**s**t – forlo**r**n*

(3) Before the devoiced high vowels /i/ and /u/, Japanese /h/ is realized as a palatal fricative [ç], and labial fricative [ɸ], respectively, while it remains [h] before non-high vowels.
hito /hito/ [**ç**i̥to] 'person' *fune* /hune/ [**ɸ**u̥ne]'ship, boat' vs. *hata* /hata/ [**h**ḁta] 'flag'

(4) Prenasal fricative weakening and loss
Latin *spa**s**mu* > Gascon *espa**u**ma* 'spasm', Latin *elemo**s**(i)na* > Catalan *almo**i**na* 'alms'
English *isn't* [ɪ**n**n̩t], *give me* [ˈgɪ**m**mɪ]

(5) Post nasal affrication
Shekgalagari /n+zípa/ [n.**t**sípa] 'zip up for me!' vs. /χʊ+zípa/ [χʊzípa] 'to zip up'
English *Ham**p**stead* < Old English *hām + stede*, *Prince* [n**t**s], *length* [ŋ**k**θ]

(6) Voiceless stop weakening
Post nasal voicing, Japanese *yom+**t**a* > *yon**d**a* 'read PAST'
Assimilation, Indonesian /məN/+**k**asih [məŋasih] 'give' vs. /məN/+gambar [məŋgambar] 'draw'
Glottalization, American English, *moun**t**ain* [ˈmãðn**ʔ**n̩] vs. abandon [əˈbæn**d**ən]

References

[1] Honeybone, P. 2008. Lenition, weaking and consonantal strenght: tracing concepts through the history of phonology. In Joaquim Brandão de Carvalho, Tobias Scheer and Philippe Ségéral (eds.), *Lenition and Fortition*. Berlin, New York: De Gruyter Mouton, pp. 9-92.
[2] Romero, J. 1996. *Gestural Organization in Spanish: An Experimental Study of Spirantization and Aspiration* (Doctoral Dissertation). University of Connecticut.
[3] Zhao, S. (2010). Stop-like modification of the dental fricative /ð/: An acoustic analysis. *The Journal of the Acoustical Society of America*, 128(4), 2009-2020.
[4] Solé, M.J. 2010. Effects of syllable position on sound change: An aerodynamic study of final fricative weakening. *Journal of Phonetics* 38(2), 289-305.
[5] Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29(2), 155–190.
[6] De Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, 26(3), 283–310.
[7] Trask, R.L. 2000. *The Dictionary of Historical and Comparative Linguistics*. Edinburgh: Edinburgh University Press.

Phonetic Variation

Friday, oral session 3

Age-dependent variation in the realization of schwa in Canavesano Piedmontese

Alec Gallo and Gorka Elordieta
University of the Basque Country (UPV/EHU)

This paper presents an acoustic analysis of the vowel system of Canavesano, a variety of Piedmontese spoken to the north of Turin. [1] defined for this variety 9 phonemes in tonic position: /i,y,e,ɛ,ø,a,o,u,ä/. The inclusion of /ä/, characterized by [1] as an open-mid front round vowel, is certainly mysterious: the ¨ diacritic in IPA indicates centralization but /a/ is a central vowel already, and /a/ is open and unrounded. Moreover, the author does not provide minimum pairs to justify its existence as a phoneme. On the other hand, [1] does not include the mid-central vowel /ə/, although she does transcribe this vowel in examples that constitute minimum pairs (e.g., *pəs/pes* 'fish/worse'). Within the general goal of providing a detailed acoustic analysis of Canavesano vowels, we have two specific objectives: to clarify the possible existence of /ä/, and to test the intuition of the first author that young speakers produce /ə/ as more front and more open concerning schwas in other languages or other Italic varieties such as Neapolitan.

Twelve speakers of Canavesano were interviewed, six men and six women. Given the importance of the age factor in producing /ə/, we distinguish three age groups (18-40, 41-60, and 61-80), with four speakers in each group. Each participant had to perform a task naming the object reflected in an image displayed on a computer monitor. The vowels analysed created minimum pairs, with which their phonemic value could be verified. As a control population, 6 speakers of a variety of Neapolitan (from the town of Massa Lubrense, specifically) were recorded so that their schwas could be analyzed in comparison to those of Canavesano. We analysed 1980 stressed and unstressed vowels, whose frequency values for the first two formants (F1 and F2) were analysed in Praat.

The acoustic analysis confirms that Canavesano has nine vowel phonemes in stressed positions: /a,ɛ,e,ə,ø,o,i,y,u/. No open-mid front round vowel that was supposed to be /ä/ was found. The vowel /ə/ is not exactly a middle central vowel but presents great variability in the acoustic space. With respect to the average F1 and F2 values reported for /ə/ in different languages, which are 511 Hz and 1428 Hz ([2]), /ə/ in Canavesano presents higher values for both formants (618 Hz and 1577 Hz, respectively). That is, it is more open and slightly more fronted than a schwa with the average F1-F2 values in [2]. Conversely, /ə/ in the Neapolitan variety of Massa Lubrense presents average values of 573 Hz and 1457 Hz for F1 and F2, more in line with the mean values of schwa in other languages as summarized in [2]. However, schwa in Canavesano varies according to the age of participants. Older speakers have schwas with F2 values that are very close to those of Massa Lubrense: 1474 Hz, very similar to the 1457 Hz in Massa Lubrense, and the 1428 Hz for /ə/ in [2]). On the other hand, young and middle-aged speakers have significantly higher F2 values (1647 and 1611 Hz, respectively). This confirms the first author's intuition that younger generations have a more front realization of schwa, with partial overlap with /ø,ɛ/, which have average F2 values of 1643 and 1763 Hz, respectively. The term *partial overlap* needs to be stressed, as F1 serves to distinguish /ø,ɛ,ə/ among young and middle-aged speakers: an average of 515 and 685 Hz for /ø/ and /ɛ/, respectively. The F1-F2 plot charts in Figures 1-3 show the relative differences among age groups in the distribution of Canavesano vowels.

Among other results, the present investigation also confirms [1]'s description that /e/ and /ɛ/ neutralize to /e/ in pretonic position, and it brings forth the novel description that in posttonic positions only /i,u,ə/ are possible, and word-finally only /e,o,a/.

In all, this work constitutes a precise phonetic and phonological documentation of an understudied Romance variety, Canavesano Piedmontese, refining that of [1]. It also provides evidence for a possible change in progress in this variety regarding /ə/.

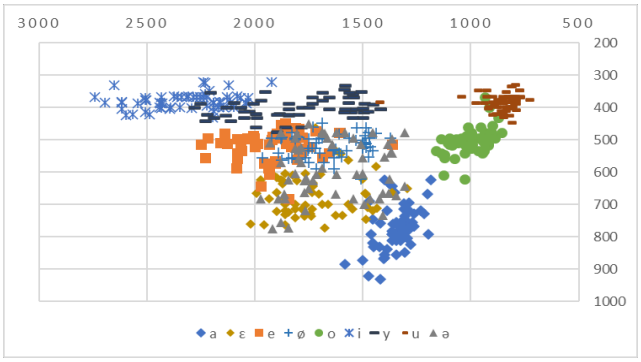


Figure 1. F1-F2 (Hz) vowel chart of Canavesano younger speakers

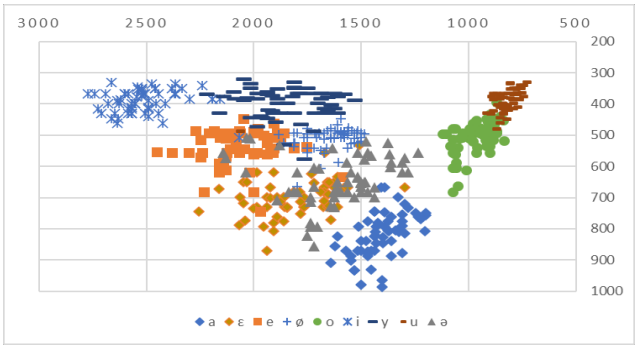


Figure 2. F1-F2 (Hz) vowel chart of Canavesano middle-aged speakers

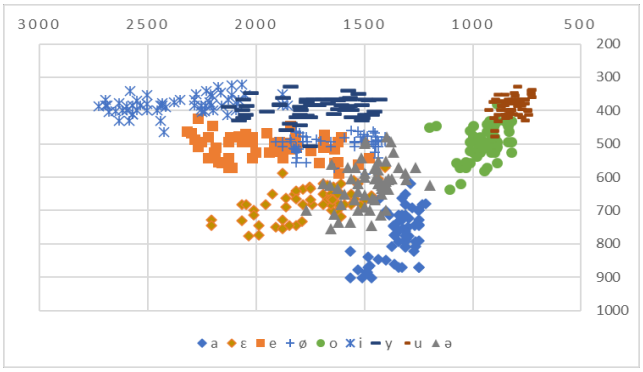


Figure 3. F1-F2 (Hz) vowel chart of Canavesano older speakers

References

[1] Zörner L. 1998. *I dialetti canavesani di Cuorgnè, Forno e dintorni. Descrizione fonologica, storico-fonetica e morfologica*. Cuornè: Edizioni CORSAC.

[2] Recasens, D. 2022. Acoustic characteristics and placement within vowel space of full schwa in the world's languages: a survey. *Journal of the International Phonetic Association* 52(1), 59–94.

Glottal and velar fricatives in a northern Iberian Spanish variety

Alejandra Mella[1], Gorka Elordieta[1] and Magdalena Romera[2]
[1]University of the Basque Country (UPV/EHU), [2]Public University of Navarre (UPNA)

In this paper we provide empirical evidence for the presence of a glottal fricative in alternation with the voiceless velar fricative /x/ of Castilian Spanish in the Cabuerniga Valley, (henceforth CV) in the northern Spanish region of Cantabria. Examples of this alternation are provided in (1). [1] assumes that the glottal fricative competed with /x/ after this phoneme had been introduced in the XVII century evolving from /ʃ/. Descriptions on the glottal fricative in CV are scarce and vague ([2], cf. also [3]-[6]). In a pilot study, [7] shows that the glottal fricative is mainly voiced, /ɦ/, rather than the voiceless /h/ which has been assumed previously in the literature. We also test the hypothesis that age, gender and geographical location within CV might be crucial factors affecting the alternation of the glottal and velar fricatives.

12 speakers were recorded (6 men and 6 women), in three age groups: 18-35, 36-59 and 60+ years of age, with 4 subjects in each group. 6 speakers lived in the higher part of CV, more isolated, and 6 speakers lived in the lower part, closer to other areas. The data was produced through sociolinguistic interviews and pronunciations of isolated words. In the interviews, the first author asked questions to the subjects on life in CV (e.g., marital status, number of children, whether they had ever spent time outside CV, their habits in everyday life, whether they considered life in CV optimal or satisfactory, their perception of their own vernacular linguistic variety and of the inhabitants of CV). The goal was to gather naturalistic data. To obtain more fricatives, the subjects were asked to pronounce 21 words containing a velar fricative (thus potentially in alternation with a glottal fricative). The words were all familiar objects, to facilitate a natural reading. In addition, 40 filler words were included. As a comparison, we also analyzed the velar fricatives from a control group of 4 speakers from Madrid and other areas outside CV. In total, 837 target fricatives were obtained and analyzed (676 from CV and 118 from the control group). The following spectral properties were analyzed ([8]-[12], among others): Center of Gravity (CG), FFT and LPC peaks, standard deviation, skewness and kurtosis. Additionally, degree of voicing, duration, intensity, and difference in intensity with respect to that of the following vowel were also measured. We classified the fricatives as velar or glottal based on auditory perception and analyzed acoustically with respect to the parameters listed above.

In CV, 59.3% of the target fricatives were velar and 38.5% were glottal (2.2% could not be classified as either glottal or velar). Thus, although not prevalent, a glottal realization alternating with the widespread velar fricative /x/ is still significantly present in CV. On the contrary, the control group only produced velar fricatives, as expected. Compared to the velar fricatives, the glottal fricatives had significantly lower frequency CG and FFT peaks and a lower standard deviation, which is expected if CG and spectral peaks are in inverse proportion with the anteriority of the constriction (i.e., the further back the constriction is, the lower the CG and the frequencies of the spectral peaks). The glottal fricative also had a significantly higher value for skewness and kurtosis than the velar fricative. The velar fricative was voiceless, as expected, and the glottal fricative was voiced, as already shown by [7], i.e., /ɦ/. The glottal fricative also has significantly shorter duration and higher intensity than the velar fricative. The acoustic values of the velar fricatives in CV and the control group were similar.

Regarding the future outlook of /ɦ/, the results show that males, speakers over 60 years of age and speakers from the higher part of CV produced significantly more glottal fricatives than females, younger speakers and speakers from the lower part of CV, respectively. The fact that the presence of /ɦ/ is virtually 0% among women below 60 years of age and only 33% among males in the 18-35 age group suggest that the glottal variant is very likely to disappear in CV this century.

(1) [xaˈmon] ~ [h/ɦaˈmon] ‘ham’
[oˈβ̞exa] ~ [oˈβ̞eh/ɦa] ‘sheep’

	Glottal fricat. N = 260	**Velar fricat. N = 401**	
Center of gravity	1615 Hz	2893 Hz	$p < 0.001$
1st FFT peak	819 Hz	1272 Hz	$p < 0.001$
2nd FFT peak	1749 Hz	2352 Hz	$p < 0.001$
3rd FFT peak	2703 Hz	3442 Hz	$p < 0.001$
Standard Deviation	2742 Hz	3980 Hz	$p < 0.001$
Skewness	3	2	$p < 0.001$
Kurtosis	16	7	$p < 0.001$
Voicing	84%	14%	$p < 0.001$
Duration	30 ms	80 ms	$p < 0.001$
Intensity	56 dB	48 dB	$p < 0.001$
Diff. intensity w/ following vowel	-3.25 dB	-6.82 dB	$p < 0.001$

Table 1. *Quantitative values of the acoustic parameters for the glottal and velar fricatives*

References

[1] Lapesa, R. 1980. *Historia de la lengua española.* Madrid: Gredos.
[2] García González, F. 1972. Sobre la aspiración en la Provincia de Santander. *Publicaciones del Instituto de Etnografía y Folklore*. Santander, 221-241.
[3] Alvar, M. 2016. *Manual de dialectología hispánica: el español de España.* Barcelona: Ariel Lingüística.
[4] Zamora Vicente, A. 1974. *Dialectología española*. Madrid: Gredos.
[5] Arias, X. L. G. 2003. *Gramática histórica de la lengua asturiana: fonética, fonología e introducción a la morfosintaxis histórica* (Vol. 15). Academia Llingua Asturiana.
[6] Martínez Álvarez, J. 2016. Las hablas asturianas. In Alvar, M. (Ed.), *Manual de dialectología hispánica. El español de España*. Barcelona: Ariel lingüística, 119-133.
[7] Mella Casar, A. 2020. *La fricativa glotal sonora del Valle de Cabuérniga: un análisis sociofonético*. Master’s thesis, University of the Basque Country.
[8] Stevens, P. 1960. Spectra of fricative noise in human speech. *Language and Speech* 3, 32-49.
[9] Baum, S. R., & Blumstein, S. E. 1987. Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. *Journal of the Acoustical Society of America* 82, 1073-1077.
[10] Jongman, A., Wayland, R., & Wong, S. 2000. Acoustic characterization of English fricatives. *Journal of the Acoustical Society of America* 198, 1252-1263.
[11] 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.
[12] McMurray, B., & Jongman, A. 2011. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review* 118(2), 219-246.

A socio-phonetic study of reduced pronunciation variants in Romanian monologue speech: the case of *(nu) ştiu*

Oana Niculescu
Romanian Academy Institute of Linguistics "Iorgu Iordan – Al. Rosetti"

Acoustic analyses regarding connected speech phenomena on Romanian data are still rare, partly due to lack of available speech corpora designed for research at the interface between phonetics and laboratory phonology. Consequently, based on a recently developed Romanian speech corpus, in this paper we investigate an understudied topic related to the phonetic realizations of the first person singular verb *I know* 'ştiu' /ʃtiw/, both in affirmative and negative clauses. The few studies addressing this topic ([1], [2]) focus only on the pragmatic uses the cognitive verb in written texts. Complementary to this line of research, we propose a socio-phonetic account of reduced pronunciation variants of /ʃtiw/ in Romanian monologue speech. First, we want to examine the extent to which this form is reduced in connected speech and whether or not this process is conditioned by the phonological context, duration and spectral changes, as well as the gender of the speakers. Second, as discussed in [3], we want to account for reduction patterns depending on whether the string was used with a referential function (non-phraseological use) or pragmatic function (discourse marker).

Our analysis is carried out on 6hs of spontaneous speech pertaining to 6 native speakers (3 female, 3 male), adults (30-40 years of age), taken from a larger Romanian corpus designed and recorded by the author [4]. A total of 284 tokens were manually extracted from the corpus. All outputs underwent *temporal* (overall duration of the verb, as well as the diphthong or GV sequence in comparison to the reduced monophthong outputs) and *frequency measurements* (first two formants excerpted at the temporal midpoint, pertaining to the steady-state part of the vowel or glide). When compared to other cognitive verbs from the corpus, such as *think* (151 occurrences, with 92% of the cases present in affirmative clauses), we observe that *know* is the highest frequent verb, with over 87% of the tokens preceded by the negation 'nu'. Examples (1) to (8) showcase the gradient nature of the reduction processes involving /ʃtiw/, extending from the canonical output to the highly contracted forms. Our results show that [ʃtu] is the preferred outcome, while the canonical form only surfaces in 6% of the data.

Similar to previous research on Romance data [5], we organized the tokens into 5 categories depending on the following context (*voiceless obstruent* (56%, N = 158), *voiced obstruent* (11%, N = 31), *sonorant* (12%, N = 34), *vowel* (9%, N = 26), and *pause* (12%, N = 35); see also Table 1, Figure 1). The ANOVA t-test did not reveal a significant correlation between the phonological context and the degree of phonetic erosion. Instead, when comparing canonical vs reduced variants in terms of context function (referential vs pragmatic), we observed that pragmatic uses favored reduced variants (Table 2). Also, when the string is used with a pragmatic function, in 81% of the data it surfaces when preceded by the negative particle 'nu'. These results are also visible in the temporal domain, where referential uses have a higher average duration (205ms), and a higher degree of phonetic reduction is correlated to a decrease in the output duration (Table 3). As expected, our results also show a decrease in duration from diphthong (160ms, SD = 164ms) to GV sequence (104ms, SD = 90ms) and monophthong (55ms, SD = 48ms). These observations translate also in the formant frequencies, where we noticed a wider Euclidean distance related to the diphthong as opposed to the GV sequence (Δ_{iw} = 817Hz and Δ_{ju} = 517Hz for female speakers vs Δ_{iw} = 953Hz, Δ_{ju} = 558Hz in the case of male speakers). Furthermore, due to coarticulatory processes, the monophthong surfaces as a central, mid-open vowel (female speakers: F1 = 393Hz/F2 = 1591, male speakers: F1 = 290Hz, F2 = 1590Hz). As a final observation, when looking at the reduced pronunciation variants as a function of gender, we noticed that the canonical form exclusively surfaces in female speech (Table 4). Out of the seven possible reduction patterns, only 3 are employed by female speakers

([ʃtju], [ʃtu] and [nuʃ]/[nuʒ]), as compared to all reduction patterns appearing, in different degrees, in male speech, with [ʃtu] being the preferred outcome. To allow for a more comprehensive analysis, we consider including mixed-effects models in future stages of the study.

(1) obstruent cluster + diphthong	(‘CCVG’)	(6%, N = 17)	[nu.ʃtiw.ʧe]; [ʃtiw.kə]
(2) obstruent cluster + GV sequence	(‘CCGV’)	(16%, N = 45)	[nu.ʃtju.sə][ʃtju.tot]
(3) obstruent cluster + vowel	(‘CCV’)	(35%, N = 100)	[nu.ʃtu.ka.re]; [ʧe.ʒtu.sə]
(4) obstruent cluster + devoiced vowel	(‘CC̥V’)	(10%, N = 29)	[nu.ʃtu̥.kum]
(5) obstruent cluster	(‘CC’)	(1%, N = 5)	[nuʃt.sə]
(6) fricative + vowel	(‘CV’)	(12%, N = 33)	[nu.ʃu.kum]; [nu.ʒu.ʧe]
(7) fricative + devoiced vowel	(‘CV̥’)	(8%, N = 22)	[nu.ʃu̥#]
(8) fricative	(‘C’)	(12%, N = 33)	[nuʃ.ʧe] [nuʒ.da.kə]

	voiceless obstruent	voiced obstruent	sonorant	vowel	pause
CCVG	59%	6%	6%	6%	23%
CCGV	64%	2%	5%	7%	22%
CCV	56%	1%	16%	13%	14%
CCV̥	59%	3%	14%	17%	7%
CC	80%	0%	20%	0%	0%
CV	58%	21%	12%	3%	6%
CV̥	23%	32%	23%	13%	9%
C	55%	39%	3%	0%	3%

Table 1. *Percentage of outputs depending on the phonological context*

	Referential function	Pragmatic function
CCVG	20%	0%
CCGV	31%	37%
CCV	22%	14%
CCV̥	4%	9%
CC	2%	13%
CV	17%	9%
CV̥	4%	15%
C	0%	3%

Table 2. *Percentage of outputs in relation to context function*

	mean	median	SD
CCVG	336	278	220
CCGV	241	192	109
CCV	194	173	71
CCV̥	156	154	29
CC	130	133	26
CV	135	125	38
CV̥	115	112	23
C	68	69	16

Table 3. *Outputs duration (ms)*

	female	male
CCVG	17%	0%
CCGV	25%	11%
CCV	52%	26%
CCV̥	0%	16%
CC	0%	3%
CV	0%	18%
CV̥	0%	12%
C	6%	14%

Table 4. *Percentage of outputs in relation to gender*

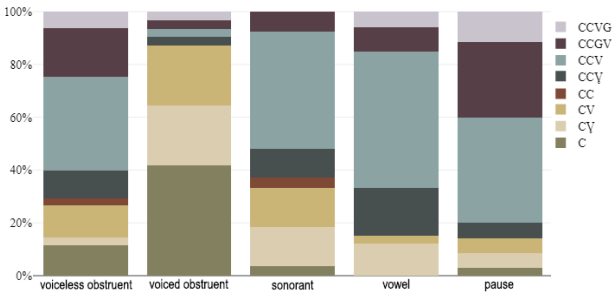


Figure 1. *Outputs frequency (stacked in group) as a function of the following context*

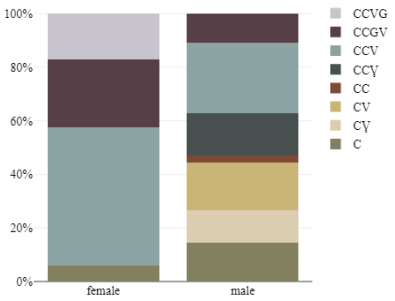


Figure 2. *Outputs frequency (stacked in group) as a function of gender*

References [1] Dascălu-Junga, L. 2012. Verbul a şti: funcţii şi valori discursive [The verb to know: functions and discursive values]. In Colceriu, St. (Ed.) Bătrânul înţelept de la Pylos. Volum omagial dedicat lui Andrei Avram la optzeci de ani. Bucharest : Romanian Academy Publisher, 27-38. [2] Zafiu, R. 2004. Nu ştiu ce. *România literară* 4. [3] Bybee, J., & Scheibman, J. 1999. The effect of usage on degrees of constituency: The reduction of don’t in English, *Linguistics* 37(4), 575-596. [4] Niculescu, O. 2021. Developing linguistic resources for Romanian written and spoken language, In Rebreja, P., Onofrei, M., Cristea, D., Tufiş, D. (eds) *Proceedings of the 16th International Conference „Linguistic Resources and Tools for Natural Language Processing”*. Iaşi: Editura Universităţii „Alexandru Ioan Cuza”, 21-36. [5] Hutin, M., Wu, Y., Jatteau, A., Vasilescu, I., Lamel, L., & Adda-Decker, M. 2021. Synchronic Fortition in Five Romance Languages? A Large Corpus-Based Study of Word-Initial Devoicing. In *Proceedings of Interspeech 2021*, Brno, Czech Republic, 996-1000.

The effect of word frequency on geminate duration in Hungarian spontaneous speech

Tilda Neuberger
Hungarian Research Centre for Linguistics

Introduction: One purpose of focusing on **casual speech** is to test whether hypotheses that have been confirmed in careful speech also hold for everyday communication [21]. Spontaneous or casual speech is generally characterized by faster articulation rate, more reduction and shorter segment duration compared to read or careful speech [2, 6, 20]. Besides speech style, **word frequency** can affect the phonetic properties of speech sounds [12, 13, 17]. According to usage-based models of language, words that are frequently used often undergo reduction as part of the move to automate speech [4]. Since both speech style and word frequency can induce durational reduction, the question arises as to how these factors affect the realisation of phonological contrasts that are phonetically implemented mainly in the temporal dimension of speech, such as the consonant length contrast. Acoustic correlates of **consonant length** have been established in various languages based on words, or (near) minimal pairs in isolation, or embedded in carrier sentences [1, 8, 9, 11, 16, 19]. However, there is only limited research on geminate production in spontaneous speech [10, 15]. Besides, the influence of word frequency on geminate production, to the best of our knowledge, has not been investigated so far.

Aim: In the present study, we observe how the consonant length contrast is realised in Hungarian spontaneous speech. Since the acoustic structure of a word may be reduced in casual speech, it is plausible that geminate durations show considerable overlap with their singleton counterparts, which results in an increased role of supplementary enhancing features (e.g., adjacent vowel duration). Our specific research question is whether and how word frequency influences the phonetic realisation of singletons and geminates. It is hypothesized that the extent of geminate reduction is correlated with token frequency. In other words, phonetic realisations of geminates in frequent words may show shorter duration, thus, more similarity to singleton forms, than in low-frequency items.

Method: Research material consists of spontaneous speech samples from 20 monolingual, Hungarian-speaking adults (aged between 20 and 31 years) from the BEA database [14]. V1CV2 segments (where C is a short/long voiceless stop and V1 and V2 are any Hungarian vowels) are chosen for acoustic analysis. Several absolute and relative durational parameters, including duration of the target consonants and neighbouring vowels, are measured in case of more than 1800 consonants and 2700 vowels using Praat [3]. Token frequency is measured based on the Hungarian Webcorpus [7]. Statistical analysis (linear mixed effect models on duration(s) as dependent variable(s), with length and word frequency as fixed factors, and speaker as random factor, as well as correlation analysis between segment/word duration and token frequency) is carried out in R [18].

Results: In spite of the considerable overlap in the absolute (C and V) durations between the two length categories, significant differences are found between singletons and geminates in relative and absolute terms. Adjacent vowel duration contributes to the C-length contrast in Hungarian spontaneous speech. Our preliminary results show that high token frequency has a reductive effect on both consonant and vowel duration. Moreover, high-frequency lexemes are more likely to be articulated with a faster tempo than rare words.

Conclusion: These findings on segment reduction are in line with those found for other phonological phenomena than length contrast, e.g., word-final t/d-deletion in American English [5]. The findings highlight how low-level phonetic phenomena interact with the meaningful end of lexicon. Results are discussed in light of related issues, such as neighbours or word predictability. The durational patterns of length found in this study may shed more light on the representation and realisation of the consonant length in the world's languages.

References

[1] Arvaniti, A., & Tserdanelis, G. 2000. On the phonetics of geminates. Evidence from Cypriot Greek. *Proceedings of 6th International Conference on Spoken Language Processing*, vol. 2. Beijing, China. 559-562.

[2] Audibert, N., Fougeron, C., Gendrot, C., & Adda-Decker, M. 2015. Duration- vs. style-dependent vowel variation: A multiparametric investigation. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS'15),* hal-01251372.

[3] Boersma, P., & Weenink, D. 2020. *Praat: doing phonetics by computer*. Computer program.

[4] Bybee, J. 1999. Usage-based phonology. *Functionalism and formalism in linguistics*, *1*, 211-242.

[5] Bybee, J. L. 2000. The phonology of the lexicon: evidence from lexical diffusion. In Barlow, M. & Kemmer, S. (eds.) *Usage based models of language*, Stanford: CSLI, 65–85.

[6] Duez, D. 1995. On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics*, *23*(4), 407-427.

[7] Halácsy, P., Kornai, A., Németh, L., Rung, A., Szakadát, I., Trón, V. 2004. Creating open language resources for Hungarian. *Proceedings of the 4th international conference on Language Resources and Evaluation (LREC2004)*, 1201-1204.

[8] Hermes, A., Tilsen, S., & Ridouane, R. 2020. Cross-linguistic timing contrast in geminates: A rate-independent perspective. *Proceedings of the 12th International Seminar on Speech Production (ISSP2020),* 52-55.

[9] Hirata, Y., & Whiton, J. 2005. Effects of speaking rate on the single/geminate stop distinction in Japanese. *The Journal of the Acoustical Society of America*, *118*(3), 1647-1660.

[10] Khattab, G. 2007. A phonetic study of gemination in Lebanese Arabic. *Proceedings of the XVI. International Congress of Phonetic Sciences,* 153-158.

[11] Lahiri, A., & Hankamer, J. 1988. The timing of geminate consonants. *Journal of Phonetics*, *16*(3), 327-338.

[12] Meunier, C., & Espesser, R. 2011. Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, *39*(3), 271-278.

[13] Munson, B., & Solomon, N. P. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47, 1048-1058.

[14] Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy, M., & Beke, A. 2014. Development of a large spontaneous speech database of agglutinative Hungarian language. *Proceedings of the International Conference on Text, Speech, and Dialogue (TSD2014),* Springer, Cham. 424-431.

[15] Neuberger, T. 2015. Durational correlates of singleton-geminate contrast in Hungarian voiceless stops. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS'15),* Paper0422. 1-5.

[16] Payne, E. M. 2006. Non-durational indices in Italian geminate consonants. *Journal of the International Phonetic Association*, *36*(1), 83-95.

[17] Pluymaekers, M., Ernestus, M., & Baayen, R. H. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561-2569.

[18] R Core Team 2020. R: A language and environment for statistical computing.

[19] Ridouane, R. 2010. Geminates at the junction of phonetics and phonology. *Papers in laboratory phonology*, *10*, 61-90.

[20] Skrelin, P. A. 2004. Segment features in different speech styles. *9th Conference Speech and Computer*. St. Petersburg, Russia.

[21] Tucker, B. V., & Ernestus, M. 2016. Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The mental lexicon*, *11*(3), 375-400.

Word-level Prosody

Friday, oral session 4

A corpus study of forestressing in African American English

Barış Kabak[1] & Janne Lorenzen[2]
[1]*University of Würzburg,* [2]*University of Cologne, Germany*

Prosodic features, which are vivid markers of ethnic identity [1], constitute an understudied area in sociophonetic research on African American English (AAE). One such prosodic pattern is forestressing, which is characterized by primary stress placement on word-initial syllables that carry non-initial stress in mainstream Englishes, e.g., *'November* (AAE) vs. *No'vember* **(**General American, GA). Due to the relative scarcity of non-initially stressed words in English, an inquiry into possible deviations from non-initial prominence is akin to a search for "variation in the wild". Therefore, we conducted a corpus study using the *Corpus of Regional African American Language* (CORAAL; [2]), which consists of audio recordings with time-aligned orthographic transcription from over 150 interviews with speakers born between 1891-2005 from different regions in the United States.

Extralinguistic factors such as age, region and socioeconomic status have been postulated to modulate the frequency of forestressing in AAE [1], although these observations have not been corroborated by systematic data. Furthermore, words of the structure CV.CVC were suggested to be amenable to forestressing [3], however, little is known as to whether other structural factors (e.g., part of speech, segmental and prosodic structure of words) modulate this phenomenon. As a feature that significantly lost ground at least in urban Southern American English [4], to which AAE shares similarities, an investigation of stress deviations across time, location and other social and structural variables in AAE requires systematic inspection of large amounts of spoken data for the AAE realizations of words that bear non-initial prominence in other English varieties.

In this paper, we ask (1) how **prevalent** is forestressing among AAE speakers across the USA and (2) which **linguistic** (e.g., syllable structure, word class) and **extralinguistic** (e.g., age, gender, region) factors modulate the rate of forestressing in AAE? To that end, we selected 84 words that are non-initially stressed in GA and analyzed their realization across all sub-components of CORAAL, which yielded 3200 tokens altogether. Two phonetically trained annotators auditorily determined the stress position in these tokens (agreement rate: 89%). In cases of disagreement, a consensus annotation was reached. Excluding tokens know to be prone to stress reversals (e.g., *Chinese restaurant*) and cases where no consensus was reached, the annotated corpus included 3070 tokens. The overall rate of forestressing in the data was 5.3%.

Logistic regression analyses revealed that the degree of forestressing is modulated by speakers' region and year of birth. Forestressing is more conspicuous in southern regions and among older speakers (Figure 1). Furthermore, nouns and words with heavy initial syllables bearing secondary stress in GA are more conducive to forestressing. Building on these results, we will first provide an account to explain why and how forestressing emerged in AAE by resorting to cognitive mechanisms such as majority pattern analogy as well as to other processes largely observed in second language varieties such as the overgeneralization of the Stress-to-Weight Principle. Second, we will weigh these factors against a top-down prosodic account, which dwells on the fact that AAE is characterized by a higher density of pitch accents, in comparison to GA [1],[5],[6]. Higher tonal density arguably forms a breeding ground for pitch accent clashes, which are known to yield reversals of primary and secondary prominences [7]. We thus hypothesize that the resolution of these pitch accent clashes may have led to the reanalysis of intonational prominence as lexical stress, especially in cases where words had heavy initial syllables. As such, we suggest that forestressing, although on the decline and by now a regional feature prevalently observed in the South, is not random but systematically determined by well-known principles of change and widely attested prosodic patterns, which are further modulated by a set of sociolinguistic variables.

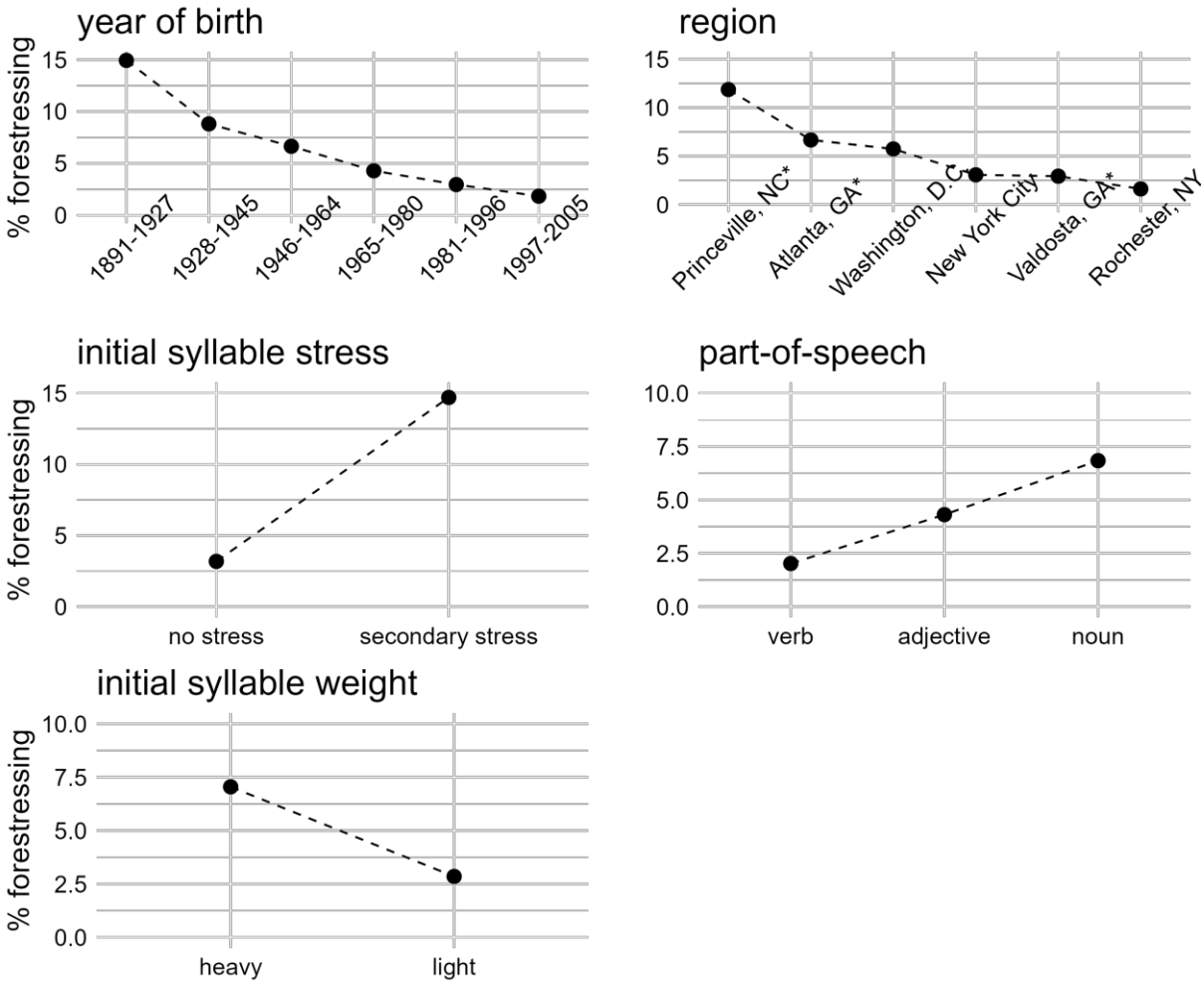


Figure 1. *Proportion of forestressing (in %) as modulated by social and structural factors*

References

[1] Thomas, E. R. 2015. Prosodic Features of African American English. In J. Bloomquist, L. Green, & S. Lanehart (Eds.), *The Oxford Handbook of African American Language* (pp. 420-436). Oxford Univ. Press.

[2] Kendall, T. & Farrington, C. 2021. *The Corpus of Regional African American Language.* Version 2021.07. Eugene, OR: The Online Resources for African American Language Project. http://oraal.uoregon.edu/coraal.

[3] Baugh, J. 1983. *Black Street Speech: Its History, Structure, and Survival*. Austin: Univ. of Texas Press.

[4] Tillery, J., & Bailey, G. 2004. The urban South: phonology. In E. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English: Vol. 1: Phonology* (pp. 325–337). Berlin: De Gruyter.

[5] Holliday, N. 2016. "Intonational Variation, Linguistic Style and the Black/Biracial Experience." Ph.D. diss., New York University.

[6] McLarty, J. 2018. African American Language and European American English intonation variation over time in the American South. *American Speech* 93(1), 32-78.

[7] Shattuck-Hufnagel, S., Ostendorf, M. & Ross, K. 1994. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22, 357-388.

Learnability of prosodic end-weight effect in Malay echo reduplication: A substantive bias account

Jian-Leat Siah (*University of California, Los Angeles*)

Background: Echo reduplication involves copying of a word with some minor alternation, such as a change in onset consonant (e.g., *helter-skelter*) or a change in vowel (e.g., *pitter-patter*). It often respects prosodic end-weight, whereby the prosodically heavier constituent tends to come second. Several prosodic factors have been shown to contribute to prosodic end-weight [1]. The present study will focus mainly on the effects of coda size (open CVØ vs. closed CVC), coda sonority and onset size (onsetless ØV vs. onsetful CV). Typologically speaking, closed syllables, more sonorous codas and onsetful syllables induce prosodic end-weight more than open syllables, less sonorous codas and onsetless syllables do.

Motivation: Substantive bias refers to a synchronic learning bias against phonetically unnatural phonological patterns [2, 3]. Support for substantive bias in the literature is scant and is often confounded with complexity bias [4]. Some scholars suggest that pure substantive effects could be uncovered if the lexicon is impoverished [5]. Echo reduplication in Malay provides a potential testing ground for substantive effects because evidence for prosodic end-weight in the lexicon is poor for some prosodic factors. Figure 1 below shows that all the prosodic factors examined in the present study are predominantly natural in that there are more forms that conform to the typology than forms that go against it (prosodic end-weight tendency in the lexicon > 50% for all factors; see Appendix for a formula). For instance, forms whose second member is prosodically heavier than the first member (e.g., coda size: *cucu-cicit* 'descendants'; coda sonority: *sorak-sorai* 'cheer'; onset size: *inca-binca* 'chaotic') are more common in the lexicon. However, the number of forms demonstrating these lexical tendencies is sparse (< 15 forms for each factor), as indicated in Figure 2. The current study aims to investigate the learnability of these poorly instantiated lexical trends. That is, could native speakers of Malay internalize the lexical trends and extend their knowledge to novel contexts?

Method: To this end, 54 native speakers of Malay residing in Malaysia were recruited and completed an online wug test in which they had to choose between two orders for 45 echo-reduplicated wug items each (e.g., *kaju-kajut* vs. *kajut-kaju*; *butat-butam* vs. *butam-butat*; *ipak-kipak* vs. *kipak-ipak*). All the wug items obeyed Malay phonotactics and were created by manipulating the prosodic factors discussed above. Only a subset of the wug items (3 for coda size, 9 for coda sonority, 6 for onset size) is relevant for the present study.

Results: The results for the wug test are given in Figure 3. Overall, the effects of coda sonority and onset size go unlearned. The subjects' responses clustered around the 50% chance level baseline (represented by a horizontal dashed line in the figure), which suggests that they had no strong preference for either order of the wug items for these factors. Interestingly, the effect of coda size is overlearned because the subjects' end-weight responses (73%) were higher than the lexical trend (62%), hinting at the presence of a built-in substantive bias favoring closed syllables in second position. The results mentioned above were confirmed with a mixed-effects logistic regression model using the *glmer* function from the *lme4* package [6].

Implications: The current study uncovers substantive effects for coda size but not for coda sonority nor onset size. This discrepancy can be accounted for if the prosodic factors differ in terms of how robust their phonetic precursor is (cf. [7]). The effect of onset size has been shown to be subordinate to the effect of coda or rimal weight [8]. Likewise, Gordon's [9] extensive survey, as summarized in [10], shows that quantity-sensitive stress systems employing coda sonority as weight distinction (4 out of 86 languages) are much rarer than those using coda size (42 out of 86 languages), suggesting a difference in phonetic robustness of the two factors in shaping the (stress) typology. In light of these facts, the present findings reveal substantive bias in an impoverished lexicon and argue that its effect can be gradient.

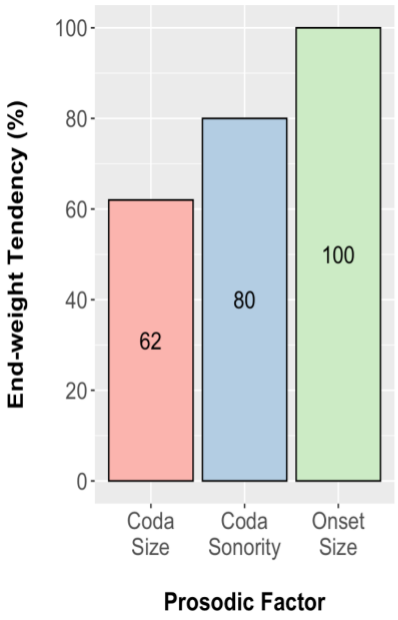


Figure 1. *Prosodic end-weight tendency in the lexicon in percentages (%)*

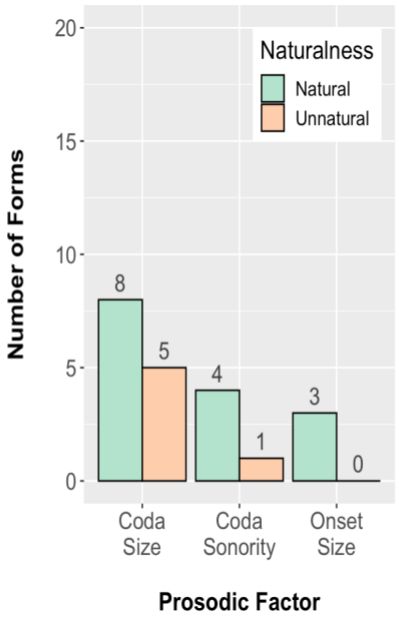


Figure 2. *Prosodic end-weight tendency in the lexicon in raw counts by naturalness*

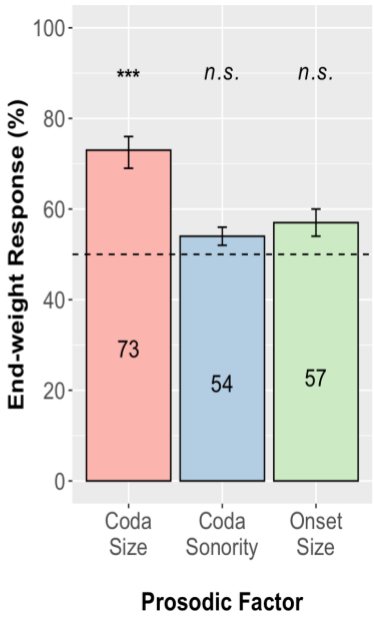


Figure 3. *Experimental results from the wug test (n.s.: non-significant)*

Appendix:

Formula used to calculate prosodic end-weight tendency in the lexicon:

$$\frac{\text{Number of natural forms}}{\text{Number of natural forms} + \text{Number of unnatural forms}} \times 100\%$$

References:

[1] Ryan, K. M. (2019). *Prosodic weight: categories and continua*. OUP Oxford.

[2] Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science,* 30(5), 945-982.

[3] Moreton, E., & Pater, J. (2012). Structure and Substance in Artificial-Phonology Learning, Part II: Substance. *Linguistics and Language Compass*, *6*(11), 702–718.

[4] Glewwe, E. R. (2019). *Bias in Phonotactic Learning: Experimental Studies of Phonotactic Implicationals*. (Doctoral dissertation, University of California, Los Angeles).

[5] van de Vijver, R., & Baer-Henney, D. (2014). Developing biases. *Frontiers in Psychology*, *5*(JUN).

[6] Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

[7] Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*(1), 83–127.

[8] Ryan, K. M. (2014). Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language*, *90*(2), 309–341.

[9] Gordon, M. (2006). *Syllable weight: phonetics, phonology, typology*. Routledge.

[10] Zec, D. (2011). Quantity-Sensitivity. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology*.

Lexical stress may not implicate the foot

Guilherme D. Garcia (Université Laval), Heather Goad (McGill University)

We compare two stress languages, English and (Brazilian) Portuguese, and empirically demonstrate that word-minimality, metrical consistency as well as weight effects suggest that the presence of lexical stress may not always implicate the foot.

In languages where prominence is characterized as "stress", it is computed in the phonological word (PWd) and realized in the foot (Selkirk 1984). English is a language of this type: in non-verbs, left-headed weight-sensitive binary feet (moraic trochees) are built from the right edge of the PWd, coupled with final extrametricality (Hayes 1982): $[ə_{\mu}(ˈd\widehat{ʒ}ɛ_{\mu}n_{\mu})_{\text{Ft}}\langle də_{\mu}\rangle]_{\text{PWd}}$ 'agenda'. Simply put, stress is penultimate if the penult is H(eavy), and antepenultimate if it's L(ight). At first glance, Portuguese looks like English, aside from extrametricality: in non-verbs, moraic trochees are built from the right edge of the PWd: $[pa_{\mu}(ˈpɛ_{\mu}w_{\mu})_{\text{Ft}}]_{\text{PWd}}$ 'paper', $[sa_{\mu}(ˈpa_{\mu}tʊ_{\mu})_{\text{Ft}}]_{\text{PWd}}$ 'shoe'. The languages look even more similar once we note that 12% of Portuguese non-verbs have antepenultimate stress, which, under a footing analysis, suggests that the language permits exceptional final syllable extrametricality: $[pa_{\mu}(ˈtɛ_{\mu}ti_{\mu})\langle kʊ_{\mu}\rangle]$ 'pathetic'.

Comparing the languages more carefully, though, we notice key differences between them. First, in English, binary feet play a role in regulating word size: no subminimal lexical words exist in the language, and truncation, including hypocorization, never results in monomoraic forms: 'chemistry' → [kɛm], *[kɛ]; 'Susan' → [suː], *[sʊ]. Second, there is general consensus that English builds moraic trochees. In Portuguese, in contrast, subminimal words are common in the lexicon and productive in the grammar (e.g., truncation and hypocorization). In addition, the range of patterns attested in the language can't be captured by a single foot type—although weight regulates stress, foot-based analyses have employed trochees (Bisol 1992), iambs (Lee 2007), and/or dactyls (Wetzels 1992), with no single analysis emerging as optimal. These differences may suggest that the foot plays a different role in the two languages: stress assignment in English is uniformly captured by binary trochees; in Portuguese, the foot plays a less important role and may be absent altogether. We explore this further by examining weight effects in antepenultimate position in HLL and LLL words in the two languages.

Predictions. If binary trochees regulate English footing, two logical possibilities are predicted for weight-sensitivity: (a) weight effects are negative, i.e., LLL words are more likely to bear antepenultimate stress than HLL words (ĹLL ≻ H́LL), because H́LL requires an uneven trochee or a medial unparsed syllable. (b) Weight effects are not active in antepenultimate syllables, i.e., a HLL word is just as likely to bear antepenultimate stress as a LLL word (H́LL ∼ ĹLL). If Portuguese does not build feet or if the foot plays a less important role than in English, then we predict positive weight effects in antepenultimate position (H́LL ≻ ĹLL).

Experiment and results. Native speakers of Portuguese ($n = 26$) and English ($n = 25$) listened to pairs of nonce words that differed only in stress location. The stimuli (n = 240 (Pt); n = 180 (En)) were generated based on weight profile (HLL, LHL, LLL, plus LLH for Portuguese). Hierarchical Bayesian regressions found a positive weight effect in antepenultimate syllables for Portuguese ($\hat{\beta} = 0.34$, 95% highest density interval = $[0.04, 0.64]$, replicating the results in Garcia (2019)), but no weight effects were found for English (H́LL ∼ ĹLL). Moreover, we captured a sonority effect for Portuguese (but *not* for English), where sonorant codas appear to be heavier and thus more stress-attracting ($\hat{\beta} = 0.40$, 95% HDI = $[0, 0.82]$). These results strengthen the plausibility of the foot for English and further question its status for Portuguese. Our findings may lend support to studies that challenge the universality of certain prosodic domains (e.g., Pierrehumbert 2003; Blevins 2004; Harris 2007; Schiering et al. 2010; Özçelik 2017), as well as the relationship between feet and stress (e.g., Vaysman 2009).

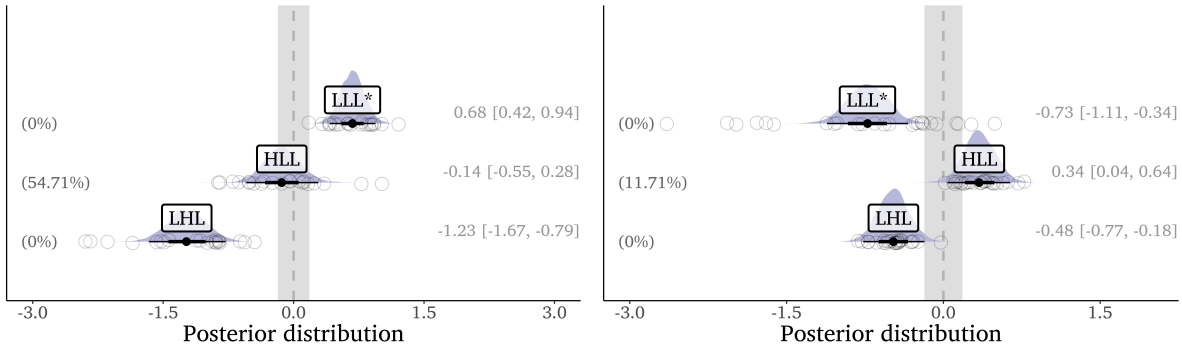


Figure 1: Posterior distributions of effect sizes ($\hat{\beta}$) for English (left) and Portuguese (right) along with means and respective 95% highest density intervals (HDI; on the right in each figure). Percentage of posterior within the region of practical equivalence (ROPE; gray bar around zero) is shown to the left in each figure. HDI of HLL is clustered around zero for English, but entirely positive for Portuguese. LLL* represents the intercept (reference level). Gray circles represent by-speaker offset (random intercepts and slopes).

References

Bisol, L. (1992). O acento e o pé métrico binário. *Cadernos de Estudos Linguísticos 22*, 69–80.

Blevins, J. (2004). *Evolutionary phonology: the emergence of sound patterns*. Cambridge, UK: Cambridge University Press.

Garcia, G. D. (2019). When lexical statistics and the grammar conflict: learning and repairing weight effects on stress. *Language 95*(4), 612–641.

Harris, J. (2007). Representation. In P. de Lacy (Ed.), *The Cambridge handbook of phonology*, pp. 119–137. Cambridge: Cambridge University Press.

Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry 13*(2), 227–276.

Lee, S.-H. (2007). O acento primário no português: uma análise unificada na Teoria da Otimalidade. In G. A. Araújo (Ed.), *O acento em português: abordagens fonológicas*, pp. 120–143. São Paulo: Parábola.

Özçelik, Ö. (2017). The foot is not an obligatory constituent of the prosodic hierarchy: "stress" in Turkish, French and child English. *The Linguistic Review 34*(1), 157–213.

Pierrehumbert, J. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, and S. Jannedy (Eds.), *Probability theory in linguistics*. Cambridge, MA: MIT Press.

Schiering, R., B. Bickel, and K. A. Hildebrandt (2010). The prosodic word is not universal, but emergent. *Journal of Linguistics 46*(3), 657–709.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Vaysman, O. (2009). *Segmental alternations and metrical theory*. Ph. D. thesis, Massachusetts Institute of Technology.

Wetzels, W. L. (1992). Mid vowel neutralization in Brazilian Portuguese. *Cadernos de Estudos Linguísticos 23*, 19–55.

Learning to distinguish morphological categories based on subphonemic detail?

Dinah Baer-Henney and Dominic Schmitz
Heinrich-Heine University Düsseldorf

Recent research has shown that morphological structure leaks into subphonemic detail. One example of this is word-final /s/ which takes several morphological roles in English. While there are words with a non-morphemic final /s/ (e.g., *bu**s***), final /s/ can also denote number and case information (e.g., *two pot**s**, the cat**'s** fur*) as well as a cliticized form of auxiliary verbs (e.g., *it**'s** been a long time, it**'s** me in the picture*). Phonetic differences among morphological distinct types of /s/ have been found for several English varieties in corpus studies [1, 2]: several types of final English /s/ come with a unique duration. Experimental studies have also addressed this question [e.g., 3,4] on production differences between categories, however, mostly with mixed results. Recently, a carefully designed production study [5] confirmed the central finding from corpus data [1,2] with non-morphemic /s/ being the longest in duration, followed by suffix /s/, then followed by clitic /s/.

On a theoretical level, these differences are unexpected when the architecture of language production does not allow for an effect originating from the morphological level to leak down to the subphonemic level [6,7]. More recent experience-based models allow for such an influence and only recently it has been shown that the aforementioned subphonemic differences could be explained as emerging from the lexicon on account of naive or linear discriminative learning [8,9].

The accumulating evidence for the effects in production has raised the question as to whether these durational differences also play a role in comprehension. A recent PhD dissertation [10] addressed this question and investigated in a perception and two comprehension experiments whether subphonemic differences play a role in decoding morphological categories. Indeed, it was found that durational differences cannot only be perceived by English speakers but also significantly affected their comprehension process.

The present study investigates whether language users not only produce, perceive and comprehend durational differences, but also whether these cues are strong enough to guide a learner in morphological learning. We investigate whether the differentiation of morphological categories based on durational cues enables the learner to build up a new representation and whether there is a disadvantage compared to learning morphological categories that differ in phonemes. To avoid native language influences we invented an artificial language with varying final /f/ durations to be learned by adult German native speakers. Participants learn a certain alternation pattern which determines the encoding of singular and plural forms in their artificial language. The alternation pattern varies between experimental groups.-In an ongoing artificial language learning experiment, we are currently collecting data comparing the learning behaviour of these three experimental groups: The 'Phonemic group' learns an artificial language in which plurality is indicated by a phonemic change in the final sound of the word [f~p alternation]. Two 'Phonetic groups' learn an artificial language where plurality is indicated by a shorter or a longer durational difference in the word-final sound [f~f: alternation]. After a short training phase, participants are requested to perform a number decision task to demonstrate what they have learned. In addition to accuracy, we measure mouse tracks to reveal possible fine differences among groups. First results indicate that learners of the 'Phonemic group' have a clear learning advantage over those in the 'Phonetic groups'. Control groups with no specific learning tasks will reveal whether we are dealing with true learning behaviour. Our results will tell us whether information exchange between the domains of phonetics and morphology can be beneficial for language learners as they would be able to use durational cues to identify morphologically relevant units.

References

[1] Plag, I., Homann, J. & Kunter, G. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53, 181–216.

[2] Zimmermann, J. 2016. Morphological status and acoustic realisation: Findings from NZE. In Carignan, C. & Tyler, M.D. (Eds.), *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology,* Parramatta, , 201–204.

[3] Walsh, T. & Parker, F. 1983. The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics* 11, 201–206.

[4] Seyfarth, S., Garallek, M., Gillingham, G., Ackermann F. & Malouf, R. 2017. Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience* 33, 1–18.

[5] Schmitz, D., Baer-Henney , D. & Plag, I. 2021. The duration of word-final /s/ differs across morphological categories in English: Evidence from pseudowords. *Phonetica*, *78*(5-6), 571-616.

[6] Levelt, W. J. M., Roelofs, A. & Meyer, A.S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–38.

[7] Kiparsky, P. 1982. Lexical morphology and phonology. In Yang, I.-S. (Ed.), *Linguistics in the morning calm: Selected papers from SICOL*, Seoul: Hanshin, 3–91.

[8] Tomaschek, F., Plag, I., Baayen R.H. & Ernestus, M. 2019. Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics* 57. 1–39.

[9] Schmitz, D., Plag, I., Baer-Henney, D., & Stein, S. D. 2021. Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology*.

[10] Schmitz, D. 2022. *Production, perception, and comprehension of subphonemic detail: Word-final /s/ in English*. Studies in Laboratory Phonology 11. Berlin: Language Science Press.

Prosody & Individual Differences

Saturday, oral session 5

The effect of musicality on Rapid Prosody Transcription responses

Riccardo Orrico, Jiseung Kim, Stella Gryllia, Amalia Arvaniti
Radboud University, Netherlands

Recent studies indicate that when listeners judge word prominence they weigh cues differently: Some prioritize phonetic and phonological cues, such as accentuation, F0 mean, and F0 shape [1, 2], while others prioritize morphosyntactic cues, such as focus [1], or pragmatic function [2]. Here we tested whether musicality is a predictor of some of these individual differences. Based on earlier studies showing that musicality makes listeners more attentive to pitch and rhythmic cues [3, 4, 5], we hypothesized that individuals with higher musical abilities would be more sensitive to phonetic cues when assessing prominence than those with lower musicality. We used Rapid Prosody Transcription (RPT, [6]), to test this hypothesis in relation to the H* ~ L+H* contrast in British English.

Eighty-two (47 F; 19-50 years old, $\bar{X}$: 33.8) linguistically naïve native speakers of Standard Southern British English (SSBE) took part in a RPT study: They heard 86 SSBE utterances that were elicited from 8 talkers (4 F) and had to select the words that sounded prominent to them. The present results are based on 287 words in the stimuli that were categorized as accented with H* or L+H* based on phonetic criteria: accents were categorized as L+H* if they were not in absolute initial position in the utterance and involved a rise from a low F0 point. [The same accents were also categorized using pragmatic criteria; for reasons of space, this analysis is not reported here.] Generalized Additive Mixed Models (GAMMs) on the accented syllable f0 showed that L+H*s had lower f0 at the accented syllable onset, a larger pitch excursion, and a later peak than H*s (see Fig.1). Linear models revealed that L+H*s also had higher amplitude than H*s; no durational differences between the two were found. After RPT, the participants completed the Mini-PROMS musicality test [7]. The binary RPT output (a word being selected as prominent or not) was analyzed in R [8] using a Generalized Linear Mixed Effect model that included ACCENT (H*, L+H*), MINIPROMS scores, and their interaction as predictors, and with participants, talkers, and items as random intercepts.

The model showed that ACCENT was significant: L+H*-accented words were rated more prominent than H*-accented words [Estimate = 1.056 (0.29), $z = 3.65$, $p < .001$]; see Fig.2 in which the RPT output is presented in terms of p-scores (percentage of participants who marked a word as prominent). MINIPROMS was not significant [Estimate = 0.007 (0.02), $z = 0.35$, $p > .1$], but its interaction with ACCENT was [Estimate = 0.03 (0.009), $z = 3.14$, $p < .01$]. As shown in Fig.3, the higher a participant's MINIPROMS score, the more likely they were to rate L+H*-accented words as prominent; conversely, participants with low MINIPROMS scores were less likely to differentiate H* from L+H* accents in terms of prominence.

Overall, the phonetics of the accents was exploited by all listeners when they assessed prominence, leading to different prominence scores based on accent shape and scaling. However, listeners varied regarding the extent to which they relied on f0 cues: Participants with higher musicality were more sensitive to these cues than those with lower musicality. Thus, the latter group perceived H* and L+H* as more similar in prominence than the former.

These findings have repercussions for our understanding of the H* ~ L+H* contrast in English, the notion of prominence, and the nature of speech processing. First, they indicate that the difference between H* and L+H* is not equally salient to all speakers of English, a finding that may explain the analytical controversy about these accents (see [2] and references therein). Further, our results support previous studies, such as [1, 2], which also show that listeners attend to cues to differing degrees, leading to differences in assessed level of prominence. Finally, our results add to a body of research, e.g., [9, 10], which shows that speech processing is affected by listener characteristics. Such individual variation and its sources, which as shown here include musicality, deserve more in-depth investigation.

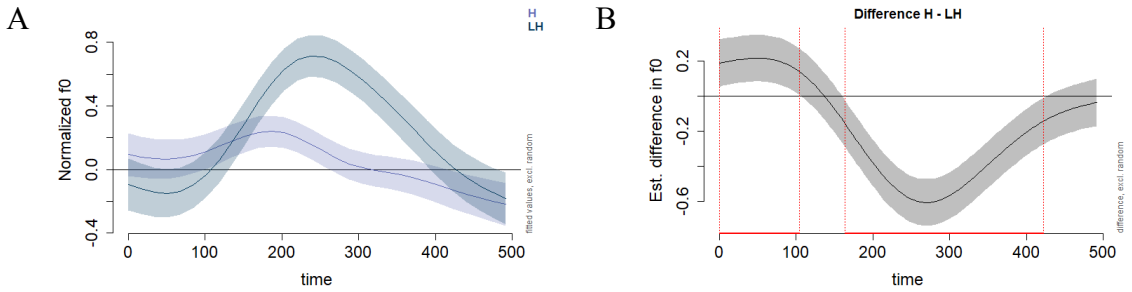


Figure 1. *GAMM results: Predicted f0 values (A) and predicted difference curves (B) between H* and L+H* in the stimuli*

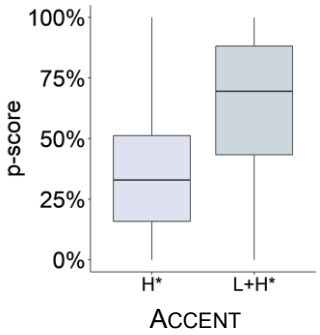


Figure 2. *Effect of ACCENT on p-scores (percentage of participants who marked a word as prominent)*

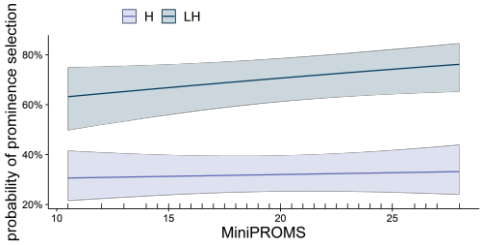


Figure 3. *Predicted RPT response probabilities for the ACCENT and MINIPROMS interaction*

References

[1] Baumann, S., Winter, B. 2018. What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *JPhon* 70: 20-38.

[2] Arvaniti, A., Gryllia, S., Zhang, C., Marcoux, K. P. 2022. Disentangling emphasis from pragmatic contrastivity in the English H*~ L+ H* contrast. *Speech Prosody 2022. https://www.isca-speech.org/archive/speechprosody_2022/arvaniti22_speechprosody.html*

[3] Cui, A., Kuang, J. 2019. The effects of musicality and language background on cue integration in pitch perception. *JASA* 146(6): 4086-4096.

[4] Schön, D., Magne, C., Besson, M. 2004. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341-349.

[5] Cason, N., Marmursztejn, M., D'Imperio, M., Schön, D. 2020. Rhythmic abilities correlate with L2 prosody imitation abilities in typologically different languages. *Lang Speech* 63(1): 149-165.

[6] Cole, J., Shattuck-Hufnagel, S. 2016. New Methods for Prosodic Transcription: Capturing Variability as a Source of Information. *Laboratory Phonology* 7(1): 8. https://www.journal-labphon.org/article/id/6176/

[7] Zentner, M., Strauss, H. 2017. Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS. *Annals of New York Academy of Science* 1400: 33–45.

[8] 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/.

[9] Kidd, E., Donnelly, S., Christiansen, M.H. 2018. Individual differences in language acquisition and processing. *Trends in Cognitive Sciences* 22(2): 154-169.

[10] Yu, A.C., Zellou, G. 2019. Individual differences in language processing: Phonology. *Annual Review of Linguistics*5: 131-150.

Individual differences in lexical stress in Dutch: An examination of cue weighting in production

Giulio G.A. Severijnen[1], Hans Rutger Bosker[1,2], & James M. McQueen[1,2]
[1]Donders Institute for Brain, Cognition, and Behaviour, [2]Max Planck Institute for Psycholinguistics

Different people talk differently, even speakers from the same region. These individual talker differences surface as large acoustic variability in speech, both at the segmental level (vowels and consonants) and the suprasegmental, or prosodic, level (e.g., lexical stress patterns). While individual differences in segmental speech production are well established in the literature, relatively little is known about how individual talkers differ in their prosody.

Evidence from speech perception experiments has shown that listeners learn and use talker-specific prosodic cues to perceive lexical stress in Dutch. Specifically, after exposure to synthesized speech from a talker who uses only one acoustic cue (e.g., intensity, with F0 and duration set to ambiguous values), listeners prioritize that cue in perception of that talker's speech (Severijnen et al., 2021, in press). However, it is unknown how Dutch talkers actually vary in the way they realize lexical stress. The present study therefore examined individual talker differences in lexical stress production.

We recorded 40 native speakers of Dutch (balanced gender; minimal dialectal variation), producing Dutch segmentally overlapping words (e.g., *VOORnaam* vs. *voorNAAM;* 'first name' vs. 'respectable', capitalization indicates lexical stress). The target words were recorded in variable speaking contexts: in isolation and in carrier sentences. An example of the sentences is: (1) *Eerst had Jan met enthouasiasme fiets gezegd,* (2) *toen had Jan het woord VOORnaam gezegd,* (3) *daarna had Koen het woord VOORnaam gezegd* ('(1) First had Jan with enthusiasm bike said, (2) then had Jan the word first name said, (3) afterwards had Koen the word first name said'). In these sentences, the target word appeared once in an accented position (underlined), and once in an unaccented position, allowing us to disentangle acoustic correlates of lexical stress from those of sentence accent. Next, we measured six acoustic cues to lexical stress: mean F0, F0 variation, duration, spectral tilt, intensity, and vowel quality.

Acoustic measurements from each individual participant were analyzed using Linear Discriminant Analyses (LDAs), which predicted whether a syllable was stressed or unstressed based on an optimal linear combination of the acoustic cues. We ran separate LDAs for each individual participant in the different speaking contexts. The analyses provided coefficients for each cue, reflecting the weight of each cue in a given talker's lexical stress productions. This yielded insight into individual phonetic cue-weighting strategies for each participant separately (cf. Schertz et al., 2015).

On average, talkers primarily used mean F0, intensity, and duration (Figure 1), which challenges previous literature emphasizing the importance of spectral tilt in lexical stress in Dutch (e.g., Sluijter & van Heuven, 1996). Moreover, on top of these group-level cue-weighting tendencies, each participant also employed a unique combination of cues to signal lexical stress in a reliable manner as assessed using split-half reliability, illustrating large prosodic variability between talkers (Figure 2). In fact, classes of cue-weighting tendencies emerged, differing in which cue was used as the most important cue. Specifically, the isolation and accented condition contained a large group of F0-weighting talkers and another group of intensity-weighting talkers. In the non-accented condition, this changed to intensity- or duration-weighting groups.

In sum, these results illustrate that, even in a relatively homogeneous participant sample, large prosodic variability is present between individual talkers, which contributes to a more comprehensive acoustic description of lexical stress in Dutch. Together with the perceptual results in Severijnen et al. (2021, in press), this emphasizes the need for Dutch listeners to carefully track these talker-specific tendencies to optimize perception.

Figure 1.
Boxplots of the mean Linear Discriminant Analyses (LDA)-coefficients for different acoustic cues to lexical stress across participants. Cue-weights are displayed in three conditions: isolation, accented, and non-accented. Higher coefficients for duration, F0, and intensity demonstrate that – on average – these cues are the primary cues to lexical stress in Dutch.

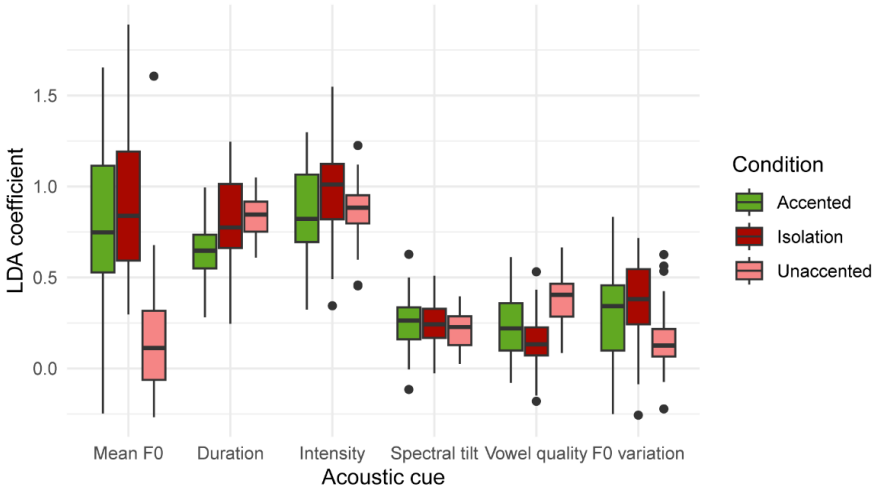
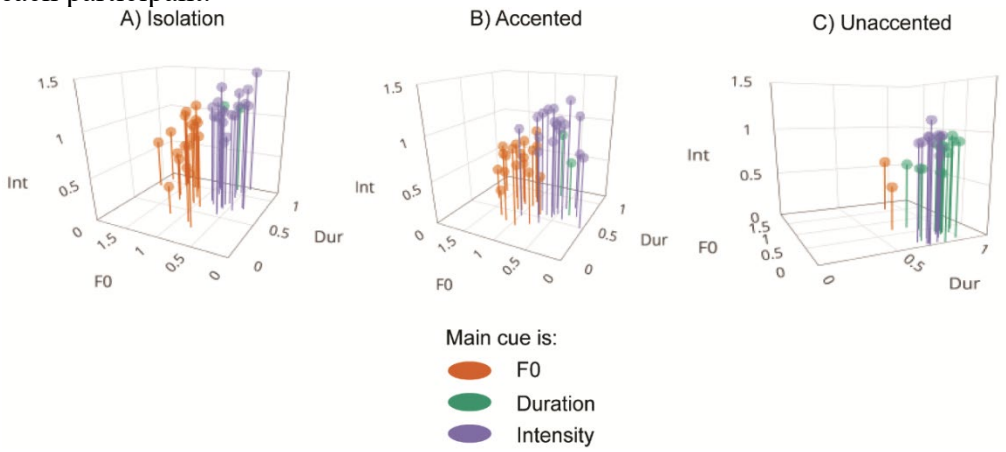


Figure 2.
Scatter plots of Linear Discriminant Analyses (LDA)-coefficients from individual participants. Each data point represents one participant plotted along three acoustic dimensions (mean F0, intensity, and duration), illustrating considerable between-talker variability. The main cue (color coded) is determined as the cue with the highest LDA coefficient of the three dimensions within each participant.



References

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204. https://doi.org/10.1016/j.wocn.2015.07.003

Severijnen, G. G. A., Bosker, H. R., Piai, V., & McQueen, J. M. (2021). Listeners track talker-specific prosody to deal with talker-variability | Elsevier Enhanced Reader. *Brain Research*, *1769*. https://doi.org/10.1016/j.brainres.2021.147605.

Severijnen, G. G. A., Di Dona, G., Bosker, H. R., & McQueen, J. M. (in press). Tracking Talker-Specific Cues to Lexical Stress: Evidence from Perceptual Learning. *Journal of Experimental Psychology: Human Perception and Performance*.

Sluijter, A. M. C., & Van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, *100*(4), 2471–2485. https://doi.org/10.1121/1.417955

Stability of individual differences in convergence across features and tasks

Jessamyn Schertz[1]
[1]*University of Toronto Mississauga, Department of Language Studies*

Background: Various social and cognitive factors have been found to influence the degree of phonetic convergence, or speakers' tendency to approximate the linguistic characteristics of incoming speech [1, 2, 3]. If these non-linguistic factors play a strong role in predicting individual variability in convergence, then individual tendencies should be stable across different features; in other words, we would expect that individuals who show high degrees of convergence in, e.g., stops, should also show greater-than-average convergence in vowels. The few studies that have directly examined this have not found evidence for individual stability across features [4, 5]; however, these all focused on convergence in conversation, where the high degree of inherent variability may obscure correlations. This work examines phonetic convergence to variation in VOT of stops and F2 of /u/ in highly controlled implicit and explicit tasks. Our primary question is whether **individual differences in convergence are stable across the two target features (stops and vowels)** in controlled tasks. We also test the extent to which **individual differences are stable across implicit and explicit imitation tasks**, which would be expected under proposals that they rely on shared processes [6].

Methods: 54 English speakers completed two online tasks (*implicit* and *explicit*) in which they repeated target words of two types: *stop words* began with voiceless stops, and *vowel words* contained the vowel /u/. Stimuli consisted of two versions of each target word, manipulated to vary minimally in a single acoustic dimension: stop words had *low* or *high VOT* (48 ms vs. 148 ms), and vowel words had *low* or *high F2* (1265 Hz vs. 1682 Hz). The implicit task, following [7], was presented as a memory task to disguise the true purpose of the task: participants heard sequences of 1, 2, or 3 words and were asked to repeat them in reverse order. Target words were always presented in isolation to elicit direct repetition. In the explicit task, participants heard the two versions of each word (high and low) in sequence and were explicitly instructed to listen to and imitate the differences. Convergence was assessed in both tasks by comparing the difference between productions following high vs. low levels of each feature, with positive values representing convergence. Statistical analysis was done using linear mixed-effects regression models to assess group-level presence/absence of convergence across tasks and features, and Pearson's product-moment correlation for individual correlations. Results reported below are all based on significance at the $\alpha = .05$ level.

Results: Participants showed convergence in the expected direction for both stop and vowel words in both tasks, as indicated by higher VOT/F2 values following stimuli with high (vs. low) values on each dimension, and the convergence effect was larger for the explicit than the implicit task, as expected based on previous work [6] (Figure 1). In terms of our primary question of whether individual differences in extent of convergence are stable across features, there was no correlation between stop and vowel convergence in the implicit task, although there was in the explicit task (Figure 2a). There was a significant correlation in convergence across the implicit and explicit tasks for stops, but not for vowels (Figure 2b).

Conclusion: These results suggest that the lack of stability in degree of implicit convergence across features [4, 5] holds even in a highly controlled task. Individual patterns did, however, hold across features in the explicit task, likely due to additional processes influencing performance (e.g. attention to and engagement with the task; willingness to modify production norms). We also found limited evidence for a relationship between convergence on explicit and implicit imitation tasks, which held for stops, but not vowels. Overall, the results highlight the need for caution in using tasks targeting a single dimension as a proxy for individual tendency to converge, and they set the groundwork for more rigorous tests of the relationship between implicit convergence and explicit imitation on an individual level.

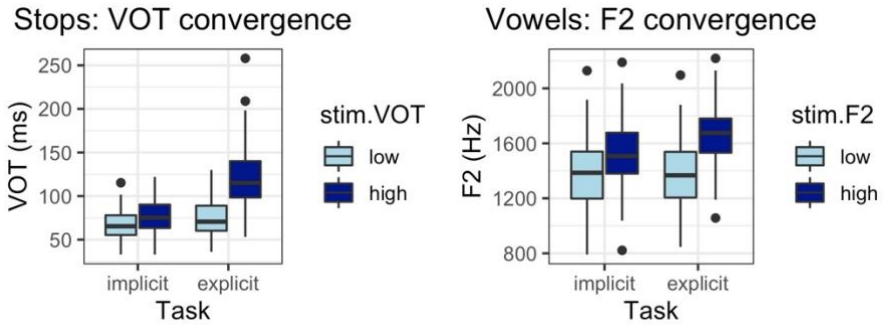


Figure 1. *VOT (for stops) and F2 (for vowels) of participants' productions following stimuli with low (light) or high (dark) of each dimension in the implicit and explicit imitation tasks. Boxplots show distributions of by-participant means.*

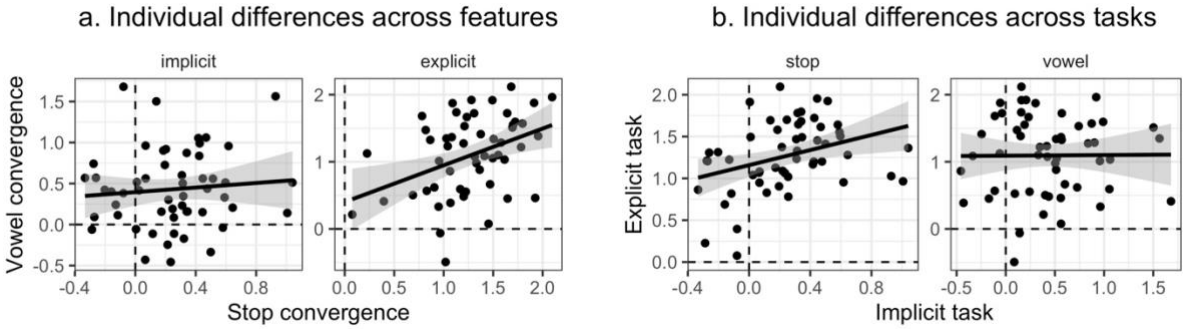


Figure 2. *Correlations between individual imitation scores (a) across features, for implicit vs. explicit tasks separately, and (b) across tasks, for stops vs. vowels separately. Each point represents one participant. Positive scores indicate convergence.*

References

[1] Lewandowski, N., Jilka M. (2019). Phonetic Convergence, Language Talent, Personality and Attention. Frontiers in Communication, 4.

[2] Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. Journal of Phonetics, 40(1), 177–189.

[3] Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and "Autistic" Traits. PloS One, 8(9), e74746.

[4] Cohen Priva, U., & Sanker, C. (2020). Natural Leaders: Some Interlocutors Elicit Greater Convergence Across Conversations and Across Characteristics. Cognitive Science, 44(10), e12897.

[5] Weise, A., & Levitan, R. (2018). Looking for structure in lexical and acoustic-prosodic entrainment behaviors. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2, 297–302.

[6] Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? Frontiers in Psychology, 4, 346–346.

[7] Wagner, M. A., Broersma, M., McQueen, J. M., Dhaene, S., & Lemhöfer, K. (2021). Phonetic convergence to non-native speech: Acoustic and perceptual evidence. Journal of Phonetics, 88, 101076.

Prosodic entrainment in mother-adolescent interaction and relationship quality

Marijke ten Brinke[1], Laura Smorenburg[1], Susan Branje[1], Hugo Quené[1], Stella Gryllia[2], and Aoju Chen[1]

[1]*Utrecht University,* [2]*Radboud University*

Previous research has shown that prosodic entrainment in dialogues, i.e. two or more interlocutors using (increasingly) similar prosody [1], mirrors interpersonal relationship quality [2]; Stronger prosodic entrainment is associated with higher relationship quality [3]. However, little is known on how prosodic entrainment is related to parent-adolescent relationships. To adolescents, a good relationship and high-quality communication with their parents are crucial to the way they develop into adulthood and form new relationships later in life, which in turn is central to their socio-emotional wellbeing [4, 5]. In the current work, we have investigated what prosodic entrainment in mother-adolescent interaction is like and whether it is correlated with measures of relationship quality, communication quality, and conflict resolution.

Method: The Dutch mother-adolescent dyads used in this study were part of the longitudinal project 'Research on Adolescent Development and Relationships' [6]. The longitudinal corpus stemming from the first wave of data collection in this project includes multiple video recordings of home visits and questionnaires on social and communication measures over a period of six years. This study used the first home visit recordings ($N = 15$), in which the mother (hereafter M) (age: Mean = 43.4, $SD = 4.2$) and adolescent (hereafter A) (8 boys, 7 girls; age: Mean = 13.0, $SD = 0.3$) were asked to discuss a topic of recurrent conflict for about 10 minutes.

Speech data were extracted from the video recordings and were annotated for inter-pausal units (IPUs) [1, 3] using Praat [7]. Prosodic measures (mean f_0, max f_0, average syllable duration) [8, 9] were subsequently extracted from the IPUs within speakers and between speakers for the computation of proximity (absolute distance between M and A in adjacent IPUs), synchrony (relative distance between M and A in adjacent IPUs), and convergence (the absolute distance between M and A at conversation level, i.e. in the first versus second half of the dyad, and at turn level, i.e. the absolute distance between M and A in adjacent IPUs as a function of time). After statistically assessing what prosodic entrainment was like in these interactions, correlations were computed between entrainment scores and social variables associated with relationship quality, communication quality, and conflict resolution.

Results: Statistical analysis in prosodic entrainment showed significant convergence only at the turn-level for mean f_0 (Spearman's $r = .14$, $p < .05$). Significant synchrony was found for mean f_0 ($r = .13$, $p < .05$), max f_0 ($r = .20$, $p < .01$), and syllable duration ($r = .14$, $p < .05$). Significant proximity was found for max f_0 (Wilcoxon's $V = 18302$, $p < .01$) and syllable duration ($V = 11338$, $p < .05$). The correlation analyses (Table 1) show that measurements indicating positive relationship quality correlated positively with prosodic entrainment and that the M-A dyads behaved similarly to the adults in [3]. The majority of correlations were based on A's perspective. This might suggest that A's entrain to M's depending on their perception of the relationship quality. In contrast, the measurements indicating frequent negative interaction in the relationship did not correlate with entrainment scores. This may suggest that entrainment mostly reflects communication quality in the conversation at hand, and not so much communication quality in the relationship as a whole over a large span of time.

Conclusions: Prosodic entrainment in mother-adolescent interactions was found in the form of convergence (at turn level), synchrony, and proximity. We showed that the degree of entrainment reflects positive aspects of relationship quality between mothers and adolescents. Future research is needed to better understand the respective role of mothers and adolescents in forming prosodic entrainment and how changes in mother-adolescent relationship throughout puberty influence prosodic entrainment in mother-adolescent interactions.

Table 1. Correlation between social variables and prosodic entrainment. BR = Balanced Relatedness, NRI = Network of Relationships Inventory, A = adolescent, M = mother, n.s. = no significant result, (-) = negative correlation.

	Variable	Reported/ experienced by	Convergence (turn)	Convergence (conv.)	Proximity	Synchrony
BR	Balanced relatedness: Relationship quality	A	n.s.	n.s.	mean F0	max F0
		M	n.s.	n.s.	n.s.	n.s.
NRI	Support	A	mean F0 (-)	n.s.	n.s.	n.s.
		M	n.s.	n.s.	n.s.	n.s.
	Power	A	n.s.	n.s.	n.s.	n.s.
		M	n.s.	n.s.	n.s.	n.s.
	Negative interaction	A	n.s.	n.s.	n.s.	n.s.
		M	n.s.	n.s.	n.s.	n.s.
Conflict resolution	Positive affect	A	n.s.	n.s.	n.s.	n.s.
		M	ASD (-)	n.s.	n.s.	n.s.
	Negative affect	A	n.s.	n.s.	n.s.	n.s.
		M	max F0	n.s.	n.s.	n.s.
	Dominance	A	max F0	n.s.	n.s.	n.s.
		M	n.s.	n.s.	n.s.	n.s.
	Autonomy	A	ASD (-)	n.s.	n.s.	n.s.
		M	mean F0 (-)	n.s.	n.s.	n.s.

References

[1] Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In P. Cosi & R. De Mori (Eds.), *Proceedings of Interspeech, 12* (pp. 3081-3084). https://doi.org/10.7916/D8V12D8F

[2] Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation, 6*(4), 802-813. https://doi.org/10.1007/s12559-014-9261-4

[3] Weidman, S., Breen, M., & Haydon, K.C. (2016). Prosodic speech entrainment in romantic relationships. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of Speech Prosody 8*, Boston, 2016 (pp. 508-512). International Speech Communication Association. https://doi.org/10.21437/SpeechProsody.2016-104

[4] Boele, S. Van der Graaff, J., De Wied, M., Van der Valk, I.E., Crocetti, E., & Branje, S. (2019). Linking parent-child and peer relationship quality to empathy in adolescence: A multilevel meta-analysis. *Journal of Youth and Adolescence, 48*, 1033-1055. https://doi.org/10.1007/s10964-019-00993-5

[5] Rejaän, Z., van der Valk, I. E., Schrama, W. M., … & Branje, S. J. T. (2021). Adolescents' post-divorce sense of belonging. An interdisciplinary review. *European Psychologist.* https://doi.org/10.1027/1016-9040/a000444

[6] RADAR (nd). *Research on Adolescent Development and Relationships*. Retrieved January 31, 2022 from https://www.uu.nl/en/research/radar

[7] Boersma, P. & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.03, retrieved 17 December 2022 from http://www.praat.org/

[8] Xu, Y. (2013). ProsodyPro – A tool for large-scale systematic prosody analysis. In B. Bigi, D. Hirst, J. Lavaud, & C. Pichon-Starke (Eds.) *Proceedings TRASP 2013* (pp. 7-10). Laboratoire Parole et Langage (LPL).

[9] Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America, 123*(2), 1104-1113. https://doi.org/10.1121/1.2821762

L2 Production & Perception

Saturday, oral session 6

Rapid adaptation to unfamiliar lexical tone systems: the effects of dialect and explicit exposure

Liang Zhao[1] and Eleanor Chodroff[1, 2]
[1]*University of York,* [2]*University of Zurich*

Introduction. Though many Mandarin tone systems have four phonological categories, the precise phonetic realization of those tones differs from dialect to dialect (Fig. 1; Hou, 2002; Li, 2002; Zhao & Chodroff, 2022). Relative to the Standard Mandarin (SM) tone system, Jinan Mandarin (JM) has comparable acoustic and especially perceptual phonetic contours, but highly disparate tone–contour mappings (e.g., SM Tone 1 phonetically resembles JM Tone 2; Fig. 1). In contrast, Chengdu Mandarin (CM) has fairly disparate phonetic contours, and critically no level tone. Nevertheless, recent studies on Chengdu Mandarin have found that native SM listeners rapidly adapted to the unfamiliar dialect with less than two minutes of incidental exposure from sentential stimuli in the experiment (Zhao, Sloggett, Chodroff, 2022). Although incidental exposure was sufficient to induce adaptation, we wondered whether explicit exposure to the target dialect could further facilitate adaptation. The present study investigated the effects of dialect and an explicit exposure period on tone adaptation by native SM speakers. Given the *greater dissimilarity* between the CM and SM tone systems than between the JM and SM systems, we expected better adaptation to Chengdu than to Jinan Mandarin (Best & Tyler, 2007; So & Best, 2011). In addition, we expected improved adaptation after two minutes of explicit exposure.

Method. To invoke adaptation, sentential surprisal was manipulated by modifying a single tone on the same segmental sequence that rendered the sentence semantically plausible or implausible (e.g., "the eagle was *flying* in the sky" vs "the eagle was *gaining weight* in the sky"). Twenty-four sentence pairs were created for a spoken sentence semantic plausibility judgment task, where listeners responded "yes" or "no" to the question: "Does this sentence make sense?". The exposure passage was *The North Wind and the Sun* recorded by a Chengdu and a Jinan native speaker. The analysis combined data from two experiments that differed only in the presence of minimal-pair sentence stimuli: previous findings indicated no difference in adaptation between the two types of design (Zhao, Sloggett & Chodroff, 2023). Twenty-seven participants were tested (14 in CM; 13 in JM). The effects of word surprisal, dialect, exposure and tone category were assessed on accuracy and response time.

Results. Accuracy and response time were analyzed respectively with a Bayesian mixed-effects logistic and linear regression model. Accuracy was reliably lower in the high-surprisal condition in all the conditions (Fig. 2), as listeners misjudged high-surprisal sentences as plausible given unfamiliarity with the phonetic tone realization. Accuracy credibly improved after exposure for both dialects, but no between-dialect difference was detected. For specific tones, listeners were more accurate for Tone 1 and Tone 2 compared to Tone 4 in both dialects (Fig. 3). The response-time model revealed a consistent slowdown for the high-surprisal sentences, suggesting successful adaptation to the novel tone system in both dialects. Although responses to Chengdu sentences were generally faster than to Jinan sentences, listeners were more sensitive to the surprisal manipulation in Chengdu than in Jinan before and after exposure (Fig. 4). Moreover, better discrimination between the surprisal conditions was found after exposure for Chengdu sentences relative to Jinan. Response times did not considerably vary across tone categories, except for faster responses for Tone 2 than Tone 4 (Fig. 5).

Conclusion. The present study showed that explicit exposure reliably improved adaptation to a novel tone system in terms of accuracy in lexical judgment. Results in response times strongly indicated better adaptation to the Chengdu tone system than to the Jinan tone system with or without explicit exposure. For tone-specific adaptation, it is likely that tones with more drastically different contours (Tone 1 & 2) from the native ones were perceived better and those with similar contours (Tone 4 as falling or falling-rising in the dialects) were harder to learn.

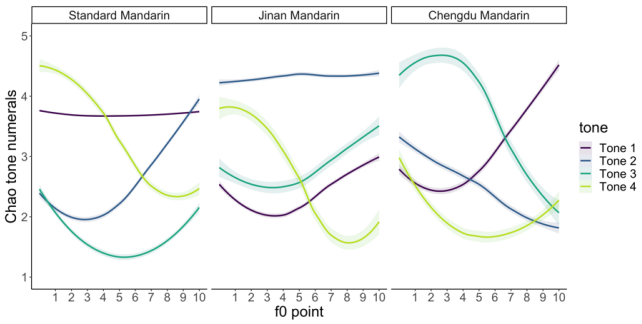


Figure 1. Smoothed lexical tone contours of Standard, Chengdu and Jinan Mandarin converted to Chao Tone numerals (Zhao & Chodroff, 2022). Ribbons reflect ±1 standard error of the mean.

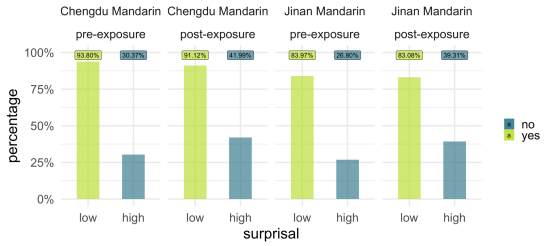


Figure 2. Percentage of "*correct*" responses across dialect, surprisal, and exposure conditions.

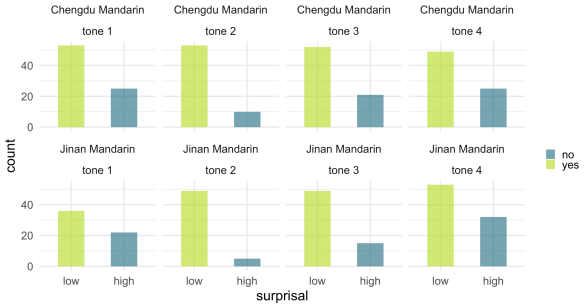


Figure 3. Counts of "*correct*" responses for each tone category in both dialects (tones balanced for each dialect).

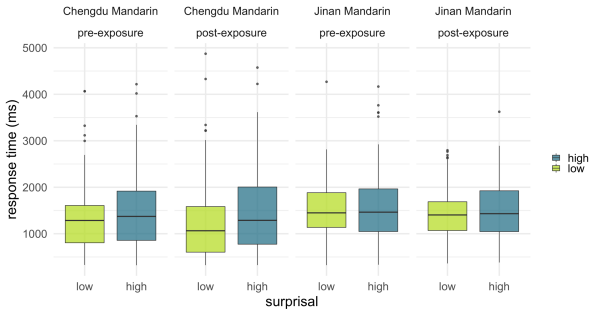


Figure 4. Response times across dialect, surprisal, and exposure conditions.

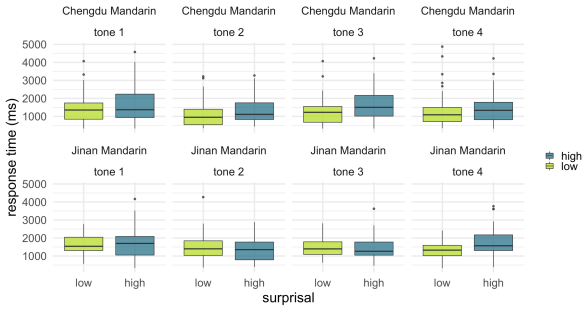


Figure 5. Response times across dialects and tone categories (tones balanced for each dialect).

References

Best, C. T., & Tyler, M. D. "Nonnative and second-language speech. Language experience", in *Second Language Speech Learning: In honor of James Emil Flege*, 2007.

Hou, J.Y. (侯精一). (2002). *The Modern Outline of Chinese Dialects* (现代汉语方言概论). China: Shanghai Education Press (上海教育出版社), 2002.

Li, Rong. (李荣). "Chengdu Dialect Dictionary (成都方言词典)," *The Modern Dictionaries of Chinese Dialects* (现代汉语方言大词典). China: Jiangsu Education Press (江苏教育出版社), 2002.

So, C. K., & Best, C. T. (2011). Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. Poznań Studies in Contemporary Linguistics, 47(1), 133.

Zhao, L., & Chodroff, E. "The ManDi Corpus: A Spoken Corpus of Mandarin Regional Dialects," in *Proc. 13th Language Resources and Evaluation Conference*, May.2022, pp. 1985-1990.

Zhao, L., Sloggett, S., & Chodroff, E. "Top-down and Bottom-up Processing of Familiar and Unfamiliar Mandarin Dialect Tone Systems," in *Proc. Speech Prosody*, May.2022, pp. 842-846.

Zhao, L., Sloggett, S., & Chodroff, E. "Conditions on Adaptation to an Unfamiliar Lexical Tone System: The Role of Quantity and Quality of Exposure," submitted to *ICPhS*, 2023.

Non-native phonological perception in a bilingual community: The influence of Southern Min on the perception of Mandarin fricatives

Caihong Weng,[1] Alexander Martin,[2] and Ioana Chitoran[1]
1. Université Paris Cité, UFR Linguistique, CLILLAC-ARP, 75013 Paris, France
2. University of Groningen, Center for Language and Cognition Groningen, 9712 EK Groningen, The Netherlands
caihong.weng@etu.u-paris.fr alexander.martin@rug.nl ioana.chitoran@u-paris.fr

Background. Much prior work has shown that our first language (L1) has a strong impact on second language (L2) speech perception [1–4]. Yet, most cross-linguistic research has focused on the perception of non-native speech by late L2 learners, with fewer studies having discussed the discrimination abilities of early bilinguals [5]. The present study addresses this question by comparing the discrimination of the [f]~[x] contrast in different phonological contexts by bilingual speakers of L1 Quanzhou Southern Min (QSM) and L2 Mandarin. Neither fricative is part of the QSM inventory, though they form a phonological contrast in Mandarin.

Hypotheses. Based on previous studies [6–8], we predict that the presence of the glide [w] and the context of rounded vowels will influence the perception of L2 [f]. As a theoretical framework for our hypotheses, we adopt the Perceptual Assimilation Model (PAM) [9]. According to the gradient assimilation levels of PAM, we predict that the [f]~[x] contrast in the context of the vowel [a] should form a Two-Category assimilation, in which [fa] will be perceived as [xʷa] (assimilating to the Quanzhou Southern Min /hʷ/ category), while [x] will be assimilated to the native /h/ category. This contrast should be easy to discriminate. However, the contrasts [fa]~[xʷa] and [fu]~[xu] ([i] and [o] was not included as neither [fi] nor [fo] is phonotactically well-formed in Mandarin) are predicted to form a Single-Category assimilation, being assimilated to QSM /hʷa/ and /hu/ respectively and should thus be more difficult to discriminate.

Method. 63 native bilingual speakers of QSM and Mandarin aged from 18 to 58 were recruited through personal connections in China to participate in an ABX discrimination task. All stimuli were recorded by four male and two female L1 Mandarin speakers. A total of 128 trials for 4 contrasts (3 target contrasts +1 control contrast) were created. The entire experiment lasted approximately 20 minutes and was conducted entirely online. Prior to data analysis, we established a strict exclusion criterion: participants who performed below chance level on the control contrast according to a binomial test (fewer than 21/32 correct trials) were considered off task and their data was excluded from analysis ($N = 27$).

Results. Figure 1 shows the mean accuracy of the 36 retained participants for each contrast. Accuracy for the contrast [fa]~[xa] (mean accuracy = 0.832) was generally higher than for [fu]~[xu] (mean accuracy = 0.697) and [fa]~[xʷa] (mean accuracy = 0.704). All analyses were performed using logistic mixed-effect models in R using the lme4 library [10]. We prepared a model including a fixed effect for the factor Contrast and a random intercept for the factor Participant, including random slopes for the factor Contrast. The factor Contrast was found to be a significant predictor ([fa]~[xa] vs. [fa]~[xʷa]: $\beta = -0.68$, SE = 0.12, $z = -5.58$, $p < 0.001$; [fa]~[xa] vs. [fu]~[xu]: $\beta = -0.76$, SE = 0.12, $z = -6.34$, $p < 0.001$). To get a better estimation of individual-level L2 exposure for our participants, we examined their answers to the post-test language use questionnaire. We found significant negative correlations between participants' level of exposure to Mandarin and their accuracy, as calculated by a difference score, for both [fa]~[xa] vs. [fa]~[xʷa] ($R = 0.22$, $p < 0.001$) (Figure 2) and [fa]~[xa] vs. [fu]~[xu] ($R = 0.17$, $p < 0.001$) (Figure 3), indicating that bilingual speakers' individual-level L2 exposure level has an positive impact on the discriminability of L2 contrast (i.e., more exposure = smaller difference between contrast types).

Discussion. In this study, we tested the perceptual assimilation of Mandarin non-sibilant fricatives by bilingual speakers of QSM, in the presence and absence of a labial glide and of rounded vowels. As expected, we found a significant difference in accuracy for the three different contrasts in line with the predictions of PAM's performance levels: Two-Category

assimilation > Single-Category assimilation. We further found that the individual-level exposure to L2 impacts these differences.

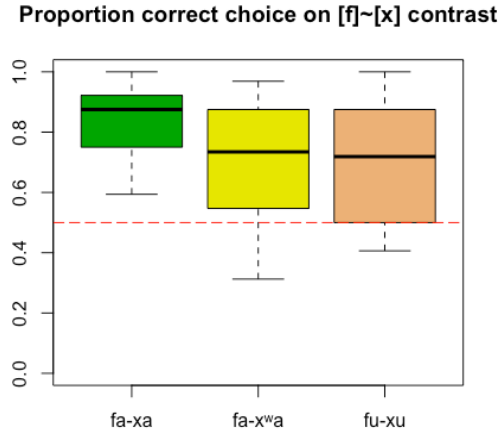


Figure 1. Proportion correct response on three main contrasts

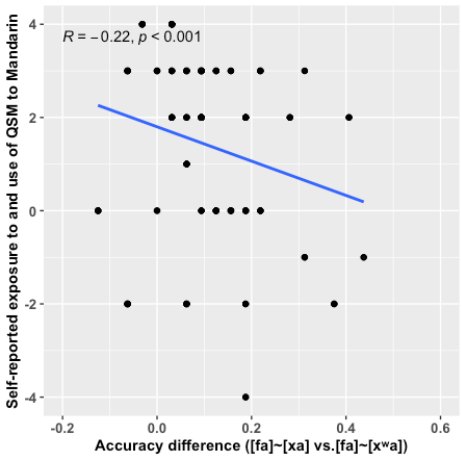


Figure 2. [fa]~[xa] vs [fa]~[xʷa] accuracy difference and L2 use/exposure level

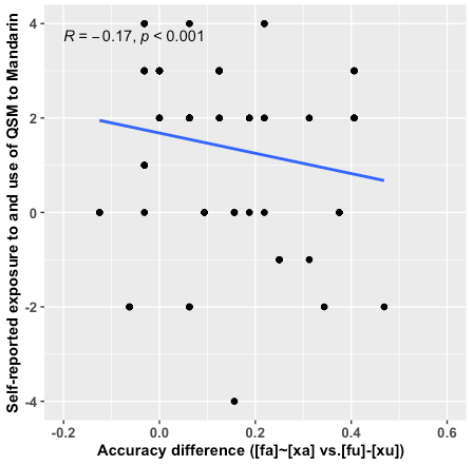


Figure 3. [fa]~[xa] vs [fu]~[xu] accuracy difference and L2 use/exposure level

References

[1] Best, C. T. (1993). "Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development," in Developmental neurocognition: Speech and face processing in the first year of life (Springer), pp. 289–304.
[2] Escudero, P. (2005). Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization (Netherlands Graduate School of Linguistics).
[3] Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," Speech perception and linguistic experience: Issues in cross-language research 92, 233–277.
[4] Robinett, B. W., and Schachter, J. (1983). "Second language learning: Contrastive analysis, error analysis, and related aspects.".
[5] Melguy, Y. V. (2018). "Exploring the bilingual phonological space: Early bilinguals' discrimination of coronal stop contrasts," Language and Speech 61(2), 173–198.
[6] Ohala, J., and Lorentz, J. (1977). "The story of [w]: an exercise in the phonetic explanation for sound patterns," in *Annual Meeting of the Berkeley Linguistics Society*, Vol. 3, pp. 577–599.
[7] Lipski, J. M. (1995). "[round] and [labial] in Spanish and the "free-form" syllable,".
[8] Greenlee, M. (1992). "Perception and production of voiceless Spanish fricatives by Chicano children and adults," Language and speech 35(1-2), 173–187.
[9] Best, C. T. (1995). "A direct realist view of cross-language speech perception," Speech perception and linguistic experience 171–206.
[10] Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). "Fitting linear mixed-effects models using lme4," arXiv preprint arXiv:1406.5823.

Are individual differences in crosslinguistic perceived similarity reflected in L2 vowel identification and discrimination?

Celia Gorba and Juli Cebrian
Universitat Autònoma de Barcelona

According to current second language (L2) speech theories, the likelihood of target-like categorization of L2 speech sounds is determined by L2 learners' ability to detect differences between native and target language sounds [1, 2]. Some studies have found that learners' developmental paths in L2 vowel categorization may not be uniform across individuals and may be affected by individual differences in the perception of similarity between native and non-native sounds [3, 4]. The current study investigated the relationship between perceived cross-language similarity and L2 perception by exploring whether individual differences in perceived similarity between target and native vowels are reflected in the individuals' ability to identify and discriminate L2 vowels. In addition, the study analyzed if perceptual training, which has been found to have a positive effect on L2 identification and discrimination, also affects the perceived similarity between target and native sounds.

A group of 28 L1 Spanish L2 English speakers completed a 6-session high variability phonetic training regime. Before and after training participants completed a series of English vowel identification and discrimination tests as well as a perceptual similarity task (PAT) in which listeners indicated the perceived similarity between native (Spanish) and target (English) vowels. The results of the PAT showed that leaners were fairly consistent in their perception of similarity between L1 and L2 vowels, but some variability was found with English /ɪ/, perceived as closest to either Spanish /i/ or Spanish /e/, and to a lesser extent with English /ɑː/, perceived as closest to Spanish /a/, followed by Spanish /o/. It was then investigated if the identification of English /ɪ/ and /ɑː/ and their discrimination from neighbouring vowels depended on individual perceptual associations.

A series of correlations and multiple regression analyses were conducted to investigate if assimilation patterns predicted identification and discrimination accuracy. The results at pretest indicated that L2 perception was generally unrelated to perceived cross-language similarity, with only a few exceptions. Further, while the 6-session perceptual training regime was effective in improving identification and discrimination of L2 vowels, it was insufficient to affect cross-linguistic similarity relations, as no consistent change in perceived similarity between L1 and L2 vowels was observed from pretest to posttest. Despite this, after training, the perceptual mappings were found to predict L2 perception to a greater extent, particularly for the discrimination of low vowels. These results show little connection between perceived similarity and L2 perception accuracy as the ability to identify L2 sounds and differentiate contrasting L2 sounds does not seem to depend consistently on the perceived similarity between L2 and L1 sounds. Finally, there were some changes in the perceived similarity of the two sounds for which there was more variability in perceived similarity (English /ɪ/ and /ɑː/). This change can be related to the claim that perceived similarity may respond to acoustic-phonetic similarity at initial stages and may become more phonologically driven as learners gain experience with the L2 ([5]).

References

[1] Best, C. T. & Michael D. T. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn & Murray J. Munro (Eds.),

Language Experience in Second Language Speech Learning: In Honor of James Emil Flege, 13–34.

[2] 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, 3–83. New York, NY: Cambridge University Press.

[3] Escudero, P., & Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. Studies in second language acquisition 26(4), 551-585.

[4] Mayr, R., & Escudero, P. 2010. Explaining individual variation in L2 perception: Rounded vowels in English learners of German. Bilingualism: Language and Cognition 13(3), 279-297.

[5] Chang, C. B. 2019. The phonetics of second language learning and bilingualism. In W. F. Katz & P. F. Assmann (Eds.), The Routledge Handbook of Phonetics. Abingdon, UK: Routledge, 427–447.

Socio-phonetic variation in the L1 and its possible effect on the L2: Does the degree of overlap between German /ɛ/ and /e/ (L1) affect the perception and production of /æ/ and /ɛ/ in English (L2)?

Marcel Schlechtweg[1], Jörg Peters[1] and Marina Frank[1,2]
[1]Carl von Ossietzky Universität Oldenburg, [2]Philipps-Universität Marburg

Someone's first language (L1) plays a key role when this person produces and perceives a spoken second language (L2) (see, e.g., Best & Tyler 2007; Flege 1995; Flege & Bohn 2021). For instance, native speakers of German face difficulties with the English vowel /æ/, which does not exist in German, and with the differentiation between the two English vowels /æ/, as in *pan* (/pæn/), and /ɛ/, as in *pen* (/pɛn/) (see, e.g., Hickey 2019). Our objective here is to go beyond this observation relating to L1 German speakers as a whole group and to look at whether the above-named vocalic phenomenon in English is more challenging for some German-speaking individuals than for others. More precisely, since we know little about the impact of L1 socio-linguistic variation on L2 speech perception and production, we consider a piece of socio-phonetic variation from German and examine whether this variation affects the production and perception of the two English vowels /æ/ and /ɛ/. Standard German distinguishes between the vowels /ɛ/, as in *Dänen* (/ˈdɛːnən/, '(the) Danish'), and /e/, as in *dehnen* (/ˈdeːnən/, '(to) stretch'). However, speakers realize this vocalic distinction to different degrees, resulting in the articulation of homophones in the most extreme case. We aim at investigating whether the degree of distinctiveness of /ɛ/ and /e/ in German (L1) has an impact on how individuals perceive and produce the English vowels (L2). Could it be that German speakers who keep the vowels in words like *Dänen* and *dehnen* clearly apart (in production), and are hence aware of an acoustic distinction of two front, mid, and unrounded vowels, have less difficulty in perceiving and producing the English contrast between /æ/ and /ɛ/, in comparison to German speakers who pronounce *Dänen* and *dehnen* in a more similar way?

We tested 56 native speakers of German in two perception (identification, discrimination) and a production task(s). In the perception tasks, we relied on two /æ/-/ɛ/ minimal pairs (*pan*/*pen* and *paddle*/*pedal*). For each pair, a spectral continuum of eleven steps was created (extreme /æ/ = Step 1, extreme /ɛ/ = Step 11). Each spectral step was crossed with the vowel durations short, middle, and long (in English, /ɛ/ is shorter than /æ/). In the identification task, participants saw two pictures, heard a sound file, and pointed via button press to the picture they associated with the sound (2 pairs, 3 durations, 11 steps). In the discrimination task, participants heard two sound files and indicated via button press whether the two were the same or different (2 pairs, 3 durations, distance between the two files was 0, 1, 2, or 3 steps).

The major independent variable was the Pillai score in German, which relies on F1 and F2 values and indicates the degree of overlap of two vowels, in this case between /ɛ/ and /e/, for a specific person (see, e.g., Nycz & Hall-Lew 2013). These values were obtained in a production experiment on German, where words like *Dänen* and *dehnen* were examined. Pillai values range from 0 (absolute overlap of two vowels) to 1 (clear separation).

Our major result is a significant interaction between the German Pillai score and Step (identification task), indicating that German speakers who separate the two German vowels to a greater degree in production select the picture representing the word with the English vowel /æ/ more often at the spectral steps 1 to 5 and less often at the spectral steps 7 to 11 than German speakers who separate the two German vowels less. We will interpret this result, together with several main effects we found in the experiments, against the background of the role the L1 plays in the perception and production of L2 speech and L2 speech perception and production models. Further, we argue for the inclusion and emphasis of socio-linguistic aspects of the L1 in L2 speech perception models, which has so far been neglected, and outline our ideas for future work.

References

[1] Best, C. T. & Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O.-S. & Munro, M. J. (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 13–34.

[2] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium: York, 233–277

[3] Flege, J. E. & Bohn, O.-S. 2021. The revised speech learning model (SLM-r). In Wayland, R. (Ed.), *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press, 3–83.

[4] Hickey, R. 2019. Persistent features in the English of German speakers. In Hickey, R. (Ed.), *English in the German-speaking world*. Cambridge: Cambridge University Press, 208–228.

[5] Nycz, J. & Hall-Lew, L. 2013. Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics* 20 (San Francisco, California).

Prosody Perception

Saturday, oral session 7

Rising shape and duration affect pitch accent categorization in South Kyungsang Korean

Hyunjung Joo[1] & Mariapaola D'Imperio[2]
[1]*Rutgers University*, [2]*Aix-Marseille University*

This study investigates how South Kyungsang Korean (SKK) listeners perceive lexical pitch accents by looking at different theoretical approaches to the phonological representation of f_0 contour. In Intonational Phonology, configurational approach has focused on the direct reflection of f_0 contour shapes onto intonational categories, by viewing the accentual f_0 contour as a whole (i.e. rises, falls) [3]. Autosegmental Metrical (AM) theory, on the other hand, has assumed that phonological primitives of f_0 contour are local f_0 turning points (i.e. H, L targets) and the transitions between the tonal targets are just a derivation from linear interpolation [8, 9]. However, studies have shown that f_0 interpolation shape is also perceptually crucial in determining pitch accent categories [1, 2, 5, 6]. Recently, Tonal Center of Gravity (TCoG) has proposed a weighted timepoint that calculates the distribution of f_0 rising, considering both peak alignment and rising shape [1, 2]. That is, f_0 rising shape, whether it is concave or convex, does matter.

South Kyungsang Korean uses lexical pitch accents to distinguish homophonous lexical items (e.g., /pam/ with H 'night' vs. /pam/ with LH 'chestnut') [7]. Note that f_0 contour for H and LH appears to show different interpolation shapes, with a slightly different timing between the peaks, as shown in Fig.1. However, only one study looked at several factors such as peak alignment and segmental duration, but not rise shape, for lexical pitch accent perception [4]. Therefore, the present study examines which cues SKK listeners use for H vs. LH contrast by looking at three factors: f_0 rise shape, f_0 peak alignment, and segmental duration.

In order to test effects of these factors on SKK listeners' categorization, a two-alternative forced choice task was carried out using three monosyllabic homophone pairs with contrastive pitch accents (H vs. LH: [kan] 'taste' vs. 'liver, [pam] 'night' vs. 'chestnut', and [pal] 'foot' vs. 'shade'). Test words were placed in a phrase-medial position of a carrier sentence. The test words were resynthesized depending on rise shape, peak alignment, and duration in Praat. As for the shape, the onset and the offset of f_0 rise coincided with the vowel onset (190 Hz) and the sonorant coda offset (290 Hz), respectively, while the midpoint of the f_0 rise was adjusted with 10 Hz increments from concave to convex shapes (Fig.2a). Next, the timing of peak alignment differed 10 % of the rime duration for each step from 10% to 100% of the rime (Fig.2b). As for the segmental duration, mean duration was lengthened by 10 ms from the stop release to the sonorant coda offset (Fig.2c). One factor at a time was manipulated, all other factors being ambiguous. 25 SKK listeners (15F, 10M) heard the resynthesized sound stimuli and then were asked to choose one of the two visual stimuli (e.g., for /pam/, 'night' on the left and 'chestnut' on the right) on the screen that would match the sound. A total of 384 stimuli (3 factors (11 shapes + 10 peak alignments + 11 durations) x 3 items x 4 repetitions) were used in the experiment.

Results show that f_0 rise shape and segmental duration were important cues for SKK listeners to distinguish lexical pitch accents, H vs. LH, while f_0 peak alignment did not show such effect. Crucially, listeners identified more concave shapes as LH, while more convex shapes induced an H categorization (Fig.3a). This supports the claim of the configurational approach and the TCoG that the interpolation shape is an integral part of f_0 contour for pitch accent categorization [1, 2, 3]. Moreover, as shown in Fig.3c, shorter duration was also responded as H, while longer duration as LH, confirming the previous finding that longer duration leads to SKK listeners' LH responses [4]. However, despite the significant effect of both f_0 shape and segmental duration on tonal categorization, shape showed a more robust categorical effect than segmental duration. Moreover, as shown in Fig.3b, lack of peak alignment effect shows that tunes cannot be reduced to a sequence of tonal targets without considering interpolation shape [1, 2, 5, 6]. Hence, a mere target-and-interpolation within AM theory seems to have some limitations on characterizing f_0 contour. To conclude, f_0 rise shape information plays an important role in the phonological specification of pitch accent in SKK.

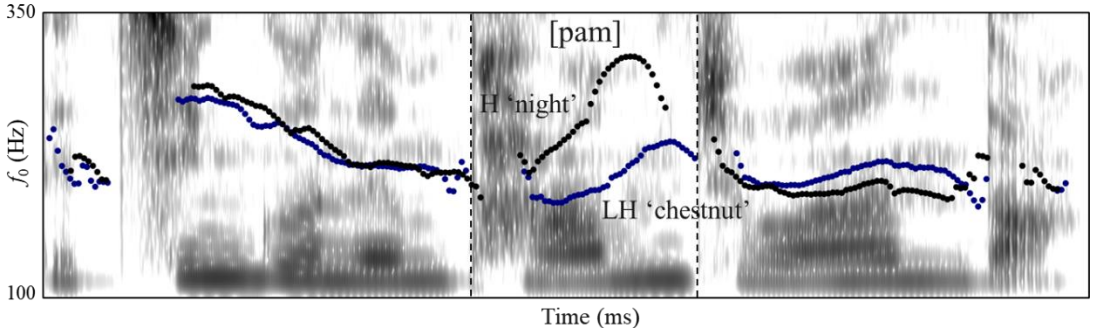


Figure 1. *f_0 contour of the target word ([pam] with H tone vs. [pam] with LH tone) embedded in an IP-medial position. The spectrogram was drawn based on [pam] with LH tone for reference.*

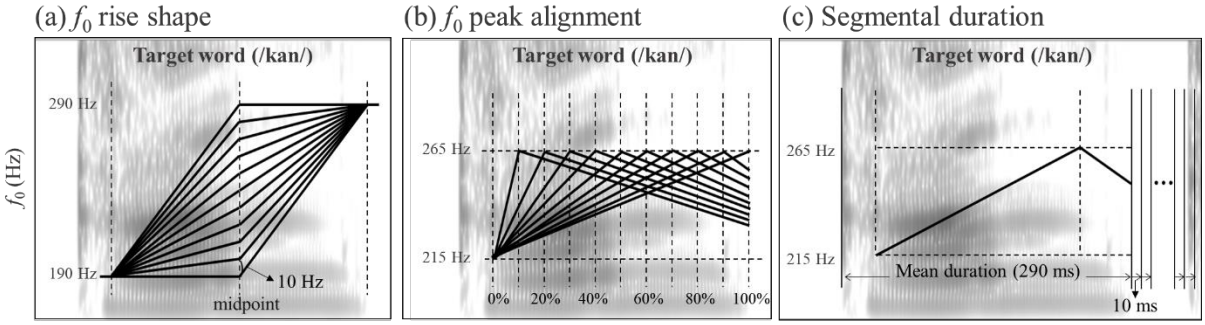


Figure 2. *Resynthesized sound stimuli depending on f_0 rise shape, f_0 peak alignment, and segmental duration.*

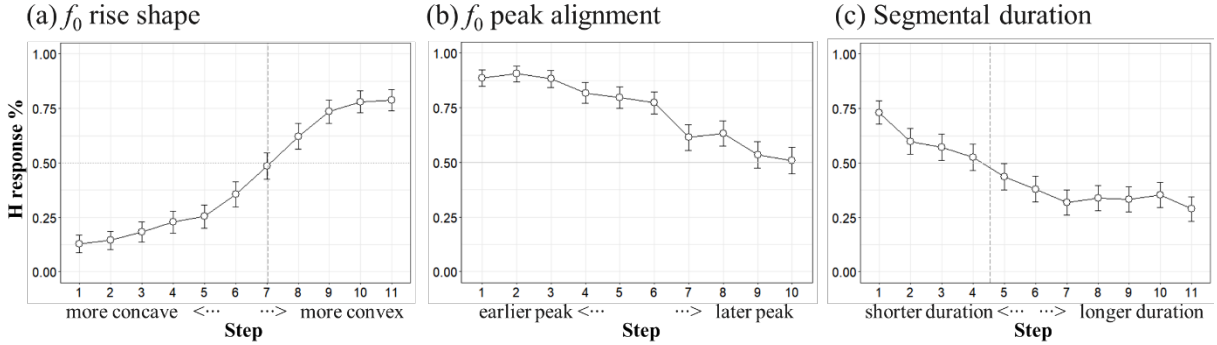


Figure 3. *H response (%) of f_0 rise shape, f_0 peak alignment, and segmental duration. The vertical line indicates the category boundary (50% crossover point).*

References

[1] Barnes, J., Veilleux, N., Brugos, A., & 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.

[2] Barnes, J., Brugos, A., Veilleux, N., & Shattuck-Hufnagel, S. 2021. On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity. *Journal of Phonetics*, 85, 101020.

[3] Bolinger, D. L. 1951. Intonation: levels versus configurations. *Word*, 7(3), 199-210.

[4] Chang, S. E. 2013. Effects of fundamental frequency and duration variation on the perception of South Kyungsang Korean tones. *Language and speech*, 56(2), 211-228.

[5] D'Imperio, M. 2000. *The role of perception in defining tonal targets and their alignment*. Unpublished doctoral dissertation thesis, Ohio State University.

[6] Dorokhova, L., & D'Imperio, M. 2019. Rise dynamics determines tune perception in French: The case of questions and continuations. *In ICPhS 2019*.

[7] Kim, J., & Jun, S. 2009. Prosodic Structure and Focus Prosody of South Kyungsang Korean. *Language Research*, 45.1, 43-66.

[8] Ladd, D. R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[9] Pierrehumbert, J. B. 1980. *The phonology and phonetics of English intonation*. Doctoral dissertation, Massachusetts Institute of Technology.

Phrase-level prosody of Akan-Twi in spoken and whistled modalities

Jonathan Barnes, Andre Batchelder-Schwab, Okrah Oppong
Boston University

Akan-Twi (Kwa, Ghana) is well-known for its musical surrogate languages, most famously perhaps a drummed surrogate in which a pair of drums, one tuned higher than the other, render the two contrasting tone levels (H, L) of the language [1]. In this paper, we provide what is to our knowledge the first substantial documentation of an additional, whistled modality for Akan-Twi. Whistled Akan conforms to the cross-language typology of whistled languages, whereby languages with contrastive lexical tone tend to spawn whistled registers in which the pitch of the whistle corresponds at least primarily to the tone contour of the utterance ([2], [3]).

In this study, we present a first account of the tonal phonetics and phonology of whistled Akan. The overarching question we address is how, if it all, phrase-level prosody is encoded in the whistled implementation of Akan lexical tone. Unlike the two apparently fixed pitches used to implement tone in drummed Akan, the pitch of the whistled signal is, in principle at least, gradiently variable. Do whistlers therefore tend to reproduce fine phonetic detail of the phrase-level F0-scaling patterns found in spoken Akan, or do they instead produce sequences of two stable whistled pitches, akin to those of the drummed signal, representing lexical phonological categories, abstracted away from contextual variability?

To answer these questions, we conducted a lab-based experiment involving six L1 speaker-whistlers of Akan. The task involved replication in the whistled and spoken modalities of Genzel's 2013 investigation of phrase-level prosody in Akan [4]. The relevant portions of Genzel's study focus on four different sentence-level tone patterns (all H, all L, alternating HL, and alternating LH), each realized over utterances of four different lengths. These sentences were constructed to investigate a number of cross-linguistically well-documented patterns of phrase-level prosody in tone languages, including declination, downstep, pre-low raising, look-ahead raising, and final lowering. (See [5] for a review and sources on each of these phenomena in spoken language prosody). We presented participants with written versions of each of these sentences, in Akan orthography (which does not represent tone), in a sequence of randomized slides. Each sentence, furthermore, was presented once as a declarative utterance, and once as a yes-no question with identical word-order. Participants first whistled the entire corpus of sentences, and then went through a second time, to produce spoken versions of each prompt.

One clear result, evident in the productions of all participants, is that while spoken and whistled pitch occupy different, effectively non-overlapping portions of the frequency spectrum, the semitone distances separating tonal targets in the two are broadly comparable. At the same time, however, whistled Akan appears to employ a narrower pitch range than does spoken Akan (e.g., Figure 1). This is true both in terms of the overall distance between High and Low tonal targets, and in the magnitude of the contextual scaling adjustments implemented for the two tone levels. In some ways this is counterintuitive, in that whistled language already has reduced information-carrying capacity compared with speech, and pitch is one of the key channels for encoding information in the whistled modality [6]. In contrast, whistled Sizang Chin (Myanmar) was attested to have a broader pitch range than spoken Chin [7].

Despite this global difference, however, the F0 patterns of our whistled corpus were found to mirror those of the spoken corpus remarkably closely at the level of fine phonetic detail. We present evidence for, among other phenomena, declination and final lowering (Figure 1), pre-low raising or upstep (Figure 2), and look-ahead raising of initial F0 targets in longer utterances (Figure 3) in both spoken and whistled Akan. The Akan whistled modality is therefore quite different from the drummed surrogate in the sense that it goes beyond the encoding of phonologically contrastive lexical tone, to render with great nuance the systematic phrase-level contextual variability characteristic of F0 scaling in the spoken language.

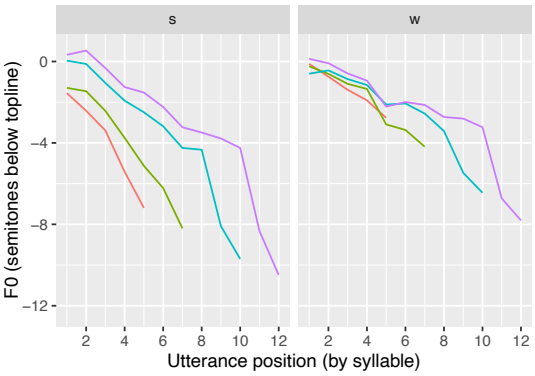


Figure 1. F0 measures for Akan all-High sentences, four lengths, averaged across speakers, speech (s) and whistle (w). Clear overall downtrend (declination) visible in both. Sharpening of downward slope toward the end is indicative of final lowering, which in Akan is quite pronounced. Note narrower pitch range for whistle.

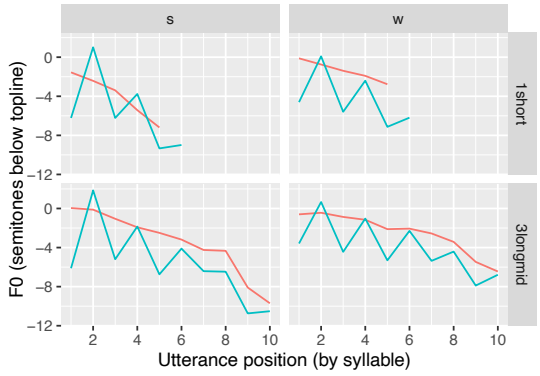


Figure 2. F0 measures for Akan all-H and alternating LH sentences, two lengths, averaged across speakers, speech (s) and whistle (w). Upstep, or prelow raising, clearly visible in both (first peak of LH sentences significantly higher than analogous syllables in all-H sentences). No clear evidence, in either speech or whistle, for downstep (lowering of H after L) beyond the normal slope of declination.

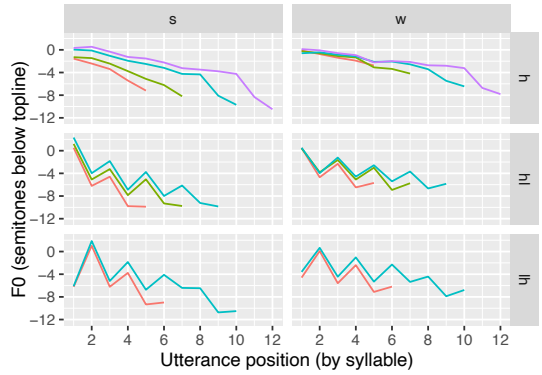


Figure 3. F0 measures for all-H, LH alternating, and HL alternating sentences, four lengths, speech (s) and whistle (w). For all tone patterns, in both speech and whistle, look-ahead raising is apparent. (Shorter sentences tend both to begin lower, and to decline faster, than their longer counterparts.)

References

[1] Kaminski, Joseph S. 2008. Surrogate Speech of the Asante Ivory Trumpeters of Ghana. *Yearbook for Traditional Music* 40. 133.

[2] Meyer, Julian. 2021. Environmental and Linguistic Typology of Whistled Languages. *Annual Review of Linguistics* 7. 493-510.

[3] Busnel, René-Guy & André Classe. 1976. *Whistled Languages*. New York: Springer.

[4] Genzel, Susanne. 2013. Lexical and post-lexical tones in Akan. University of Potsdam: PhD Dissertation.

[5] Barnes, Jonathan, Hansjörg Mixdorff & Oliver Niebuhr. 2020. Phonetic variation in tone and intonation systems. *The Oxford Handbook of Language Prosody*, ed. by Carlos Gussenhoven & Aoju Chen, pp. 125-149. Oxford: Oxford University Press.

[6] Sicoli, Mark. 2016. Repair organization in Chinantec whistled speech. *Language* 92(2).

[7] Stern, Theodore. 1957. Drum and whistle "languages": an analysis of speech surrogates. *American* Anthropologist 59. 487-506. Page 496.

Brain indices for processing rising and falling pitch: An MMN study

Maria Lialiou, Martine Grice, Christine T. Röhr, Petra B. Schumacher
University of Cologne

Attention towards a sound source is fundamentally important not only for survival but also for communication. We investigate whether rises in pitch are special in attention orienting. The idea originates from (neuro)cognitive studies attesting an attentional bias towards sounds with rising as opposed to falling acoustic properties; for example, a sudden increase in loudness or pitch of a sound is experienced by the listener as an approaching sound source, referred to as the auditory looming effect [e.g. 1].

The current study investigates whether this looming effect is triggered by rises in pitch attributable to accents and edge tones in speech. We conducted an EEG study (32 native German speakers: 28f, 4m; mean age 24.5) using the classic oddball paradigm in passive recordings: a sequence of standard repetitive auditory stimulation occasionally interspersed with a deviant sound. Of particular interest are the mismatch negativity (MMN) and a positivity around 300ms (P3) as they are claimed to index activation of pre-attentive and conscious attentional mechanisms respectively [2;3]. To simulate a more natural speech context, participants were presented with rising/falling pitch realised either on the stressed syllable or on the final syllable (see Fig. 1a) of four real CV.**CV**.CV German words (*Banane* "banana", *Limone* "lime", *Marone* "chestnut", *Melone* "melon"), alternating as standards/deviants across four oddball sequences: 1) deviant accentual fall/standard accentual rise, 2) deviant accentual rise/standard accentual fall, 3) deviant boundary fall/standard boundary rise, 4) deviant boundary rise/standard boundary fall. We hypothesize that if the processing of speech is purely signal-based, rising deviants should attract more attention by virtue of their being acoustically more prominent than falling ones. By contrast, if it is more linguistic, falling ones might attract more attention because linguistic processing is highly affected by language-specific expectations [4]. Since each sequence of standard stimuli resembles a list, which in German typically involves rises on non-final items in the list followed by a fall on the final item [e.g. 5;6], the rising sequence might be more natural than a sequence of falls. Listeners might thus habituate better to a sequence of rises, triggering a stronger reaction to falling deviants.

Separate Bayesian hierarchical models (weakly informative priors) were fitted per oddball sequence; Event related potential amplitude (in microvolt) was modelled as a function of condition (standard/deviant) from 0-to-700ms after stimulus onset, in steps of 100ms. Random effects for subjects included full variance-covariance matrices [7]. The results (see Fig. 1b for ERP waves; Fig. 1c for the posterior distributions of the estimated effects for the difference between standard and deviant) show that all deviants evoke an MMN activity relative to their corresponding standard stimulation with an onset in the 200-300ms time window. For all deviants except the accentual fall, the MMN activity lasts for two successive time windows (200-400ms), with the accentual rise exhibiting the most pronounced effect. MMN to falling deviants is followed by an additional P3, at the 400-500ms time window for the accentual falls, and at a later time window (500-600ms) for boundary falls.

Overall, all deviant contour types elicit an MMN activity, indexing an activation of pre-attentive mechanisms detecting regularity violations in the acoustic signal. Accentual rises evoke the largest MMN, indicating that they lead to a greater attraction of preconscious attention. Falling deviants engender a subsequent P3, indicating the use of additional conscious attentional resources. This is due to the naturalness of the standard sequence of rises which led to greater habituation and thus oriented conscious attention towards the deviant. In sum, deviant rises, being acoustically prominent, engender an auditory looming effect at the pre-attentive level, whereas in the case of deviant falls, the processing is affected by the linguistic context of the list, activating conscious attentional mechanisms.

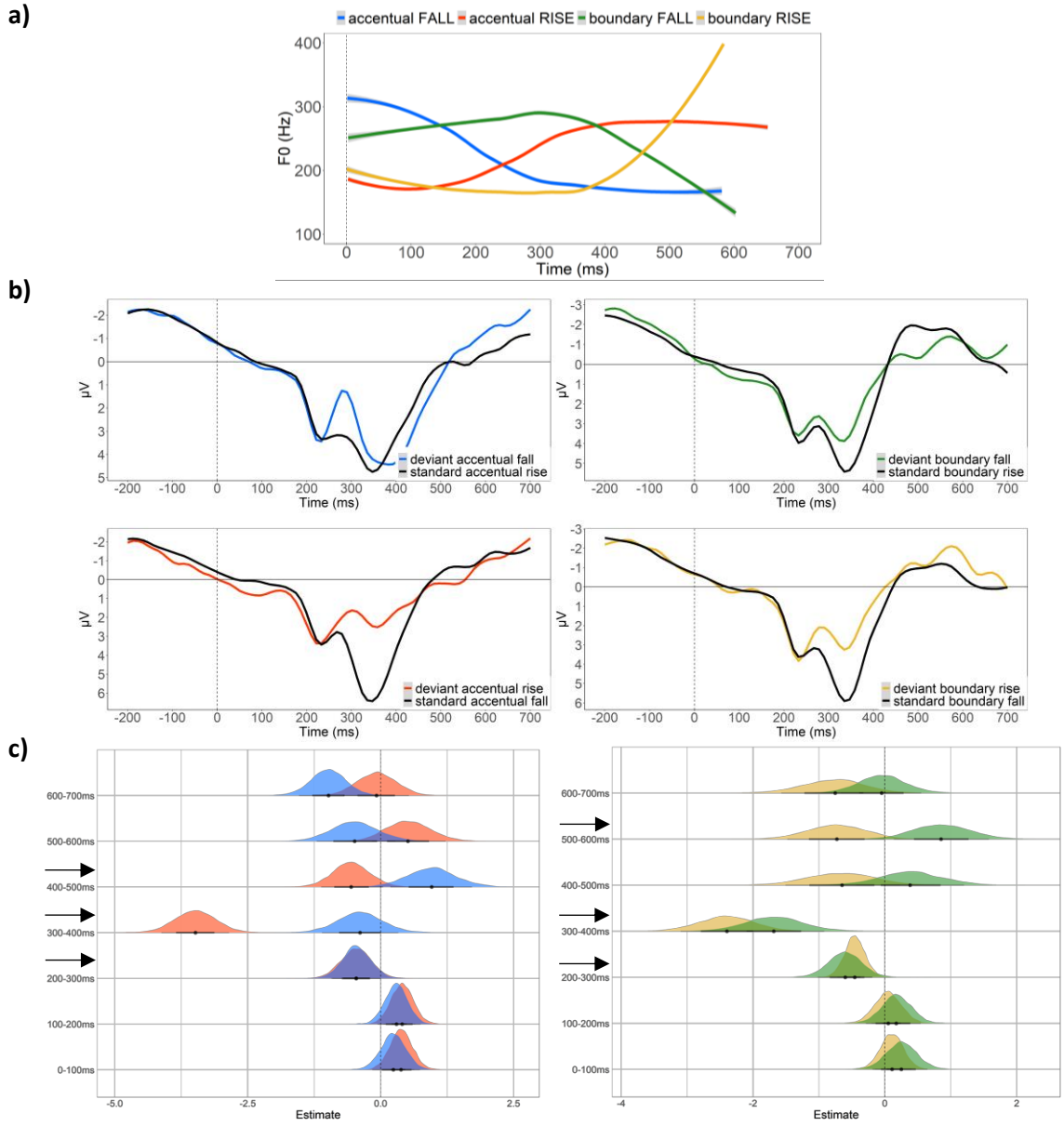


Figure 1. *Top panel (a) depicts the mean f0 contours of the oddball auditory stimulation. Middle panels (b) illustrate the grand average ERPs (per oddball sequence) recorded to the onset of stimulus (illustrated by the vertical dashed line) over time (x-axis) at the AF3, AF4, F3, Fz, F4, FC1, FCz, FC2, Cz electrode sites. Negative voltage is plotted upwards. Bottom panels (c) show the posterior distributions of the estimated effects for the difference between standard and deviant per oddball sequence (red: deviant accentual rise – standard accentual fall; blue: the deviant accentual fall – standard accentual rise; yellow: deviant boundary rise – standard boundary fall; green: deviant boundary fall – standard boundary rise). Error bars around the posterior means represent 66% (thick) and 90% credible intervals. Arrows highlight time windows of interest.*

References

[1] Bach, D., Schachinger, H., Neuhoff, J., Esposito, F., Salle, F., Lehmann, C., Herdener, M., Scheffler, K., & Seifritz, E. 2022. Rising Sound Intensity: An Intrinsic Warning Cue Activating the Amygdala. *Cerebral Cortex*, *18*(1), 145-150. [2] Näätänen, R. 1992. *Attention and brain function.* Lawrence Erlbaum Associates, Inc. [3] Polich J. 2007. Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128-2148. [4] Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. 2017. Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics, 64*(2), 90–107. [5] Baumann, S., & Trouvain, J. 2001. On the Prosody of German Telephone Numbers. Proceedings *7th Conference on Speech Communication and Technology*, 557-560. [6] Peters, J. 2018. Phonological and Semantic Aspects of German Intonation. *Linguistik Online, 88*(1). [7] Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Language Acquisition & Development

Sunday, oral session 8

Language change in Japanese-English bilingual returnee children over the course of five years: evidence from accent rating

Tim Laméris[1], Maki Kubota[2], Tanja Kupisch[3], Jennifer Cabrelli[4], Neal Snape[5], Jason Rothman[2]
[1]*Leiden University,* [2]*UiT,* [3]*University of Konstanz, University of Chicago*[4]*, GPWU*[5]

Only a handful of studies have been concerned with global foreign accent (GFA) in bilingual children [1], and little is known on how GFA develops over time. In this study, we examined the longitudinal development of GFA in bilingual returnees. Returnees are children of immigrant families who spend a significant portion of their formative developmental years (school age) in an L2 majority language (ML) context yet return to their L1 environment as older children or teenagers. During their stay abroad, they are exposed to the ML of the host country and typically acquire this an early L2. At the same time, their L1 becomes a heritage language (HL) which they are only exposed to in the home environment. Yet, upon return to their 'home country', this linguistic environment reverses: their L1 HL once again becomes the majority language, whereas the former L2 ML becomes a minority language. Given this trajectory, returnees provide a fruitful avenue to investigate two potential linguistic consequences: *heritage language reversal* ('re-development'), and *L2 attrition* [2], [3]. In this paper, we examined whether Japanese-English returnee bilinguals exhibit signs of such HL reversal and L2 attrition in their speech, using data collected over the course of five years.

We recorded 17 returnee children at three times: a few weeks after return to Japan (T1); one year after (T2); and five years after return (T3). Mean age at return was 10.02 (*sd* = 1.71). Mean age of onset (AoO) to L2 English was 5.15 (*sd* = 2.59) and mean exposure to L2 English (relative to L1 Japanese) whilst abroad, calculated by [4], was 0.48 (*sd* = 0.14). Recordings were elicited narratives of a picture book from which we created 10-second samples. These samples, in addition to 17 'baseline' samples of Japanese and English monolingual children, were used in two online accent-rating tasks, in which native speakers of American English and Japanese (each $n = 45$, and familiar with child speech) rated the degree of GFA of the samples on a 9-point Likert scale. If raters indicated a '8' or '9' ('very strong foreign accent'), they were additionally asked to indicate what features contributed to their perception of a GFA.

Internal consistency for the ratings (Cronbach's α) was high (0.916 for English and 0.908 for Japanese). A Bayesian model with weakly informative priors investigated the effect of language, time, and two experiential factors of interest (AoO to L2 English and exposure to L2 English) on accent rating. The model suggested that foreign accent in L1 Japanese decreased from T1 to T2, $b = -0.31$ (-0.48, -0.15) and continued to do so from T2 to T3, $b = -0.27$ (-0.43, -0.10), as shown in Figure 1. By contrast, foreign accent in L2 English increased from T2 to T3, $b = 0.31$ (0.13, 0.50). The model also suggested that individuals with a later AoO to L2 English returned to Japan with relatively weak foreign accents in Japanese, but strong foreign accents in English. By contrast, individuals with more exposure to English during their stay abroad returned to Japan with relatively strong foreign accents in Japanese, but relatively weak foreign accents in English. In addition, we found that English raters indicated that both segmental (vowels and consonants) and suprasegmental features (intonation, rhythm, and voice quality) contributed equally to the perception of a strong foreign accent, whereas Japanese raters primarily attributed suprasegmental features to their perception of foreign-accented Japanese (Figure 2). This could suggest that the phonological correlates of perceived global accent may vary cross-linguistically, cf. [5].

Our findings show a swift decrease in foreign accent in the L1 one year after return, and an increase in foreign accent in the L2 five years after return to the majority L1 environment. These may be indicative of HL reversal and L2 attrition in the domain of speech. We discuss how two speech systems develop over time in the bilingual individual in light of existing theories of cross-linguistic influence (CLI) of phonology [6].

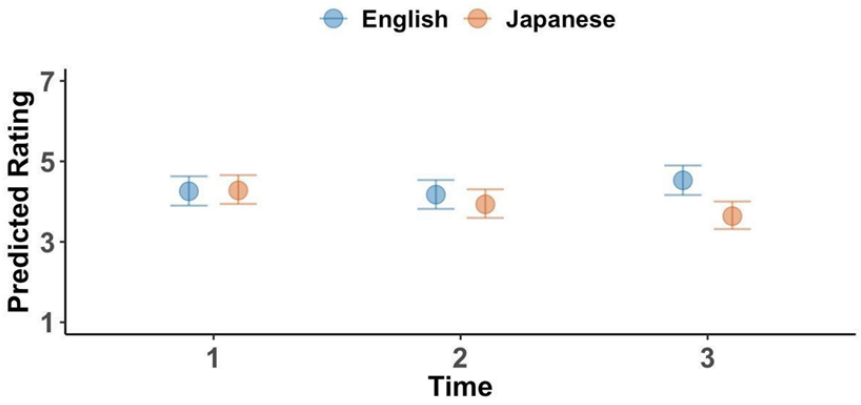


Figure 1. *Predicted accent rating per Language and Time.*

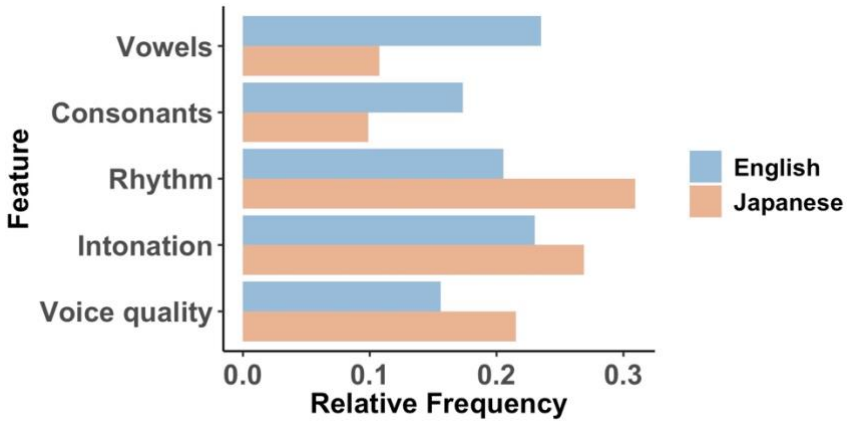


Figure 2. *Distribution of features that contributed to '8' and '9' (very strong) accent ratings per language.*

References

[1] T. Kupisch, N. Kolb, Y. Rodina, and O. Urek, "Foreign accent in pre-and primary school heritage bilinguals," *Languages*, vol. 6, no. 2, 2021, doi: 10.3390/languages6020096.

[2] C. Flores and N. Snape, "Language Attrition and Heritage Language Reversal in Returnees," in *The Cambridge Handbook of Heritage Languages and Linguistics*, Cambridge University Press, 2021, pp. 351–372. doi: 10.1017/9781108766340.016.

[3] M. Kubota, V. Chondrogianni, A. S. Clark, and J. Rothman, "Linguistic consequences of toing and froing: factors that modulate narrative development in bilingual returnee children," *Int J Biling Educ Biling*, vol. 25, no. 7, pp. 2363–2381, Aug. 2022, doi: 10.1080/13670050.2021.1910621.

[4] S. Unsworth, "Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender," *Bilingualism: Language and Cognition*, vol. 16, no. 1, pp. 86–110, Jan. 2013, doi: 10.1017/S1366728912000284.

[5] T. J. Riney, N. Takagi, and K. Inutsuka, "Phonetic Parameters and Perceptual Judgments of Accent in English by American and Japanese Listeners," *TESOL Quarterly*, vol. 39, no. 3, p. 441, Sep. 2005, doi: 10.2307/3588489.

[6] M. Kehoe and M. Havy, "Bilingual phonological acquisition: the influence of language-internal, language-external, and lexical factors," *J Child Lang*, vol. 46, no. 2, pp. 292–333, Mar. 2019, doi: 10.1017/S0305000918000478.

The relationship between phonological representations and coarticulation degree in developing speech

Dzhuma Abakarova[1]
[1]*University of Potsdam*

There is converging evidence across languages that children show greater coarticulation degrees (CD) i.e., greater spatial overlap between consecutive segments than adults [e.g., 1,2,3]. Previous research has suggested that CD reflects the nature of phonological representations, and so related developmental differences in CD to the emergent awareness of phonemes (Nittrouer et al., 1989). Recent findings of a negative correlation between CD and phonological awareness [4] provide support for this view. However, phonological awareness develops in parallel to motor control, vocabulary and various other cognitive abilities, making it difficult to experimentally control for the independent contribution of each of these factors into CD differences. That is why the present study employs dynamic modelling to address the question of whether developmental differences in coarticulation degree and phonological representations are causally related.

Phonological representations are understood here as associations between sensory cues and motor plans. Children's initial motor representations are expected to be based on whole word productions [5] and to be more variable than adults' with less clearly defined boundaries between categories. With the growing vocabulary and practice, children's motor representations for various segments (phonemes or syllables) become more delimited. Over time, the optimal trajectories are selected; the ones that compromise intelligibility or take more effort are discarded [6]. Therefore, in children, the space of motor realizations for a segment is larger than in adults. It is hypothesized here that greater variability in phonological representations allows for greater spatial overlap between co-produced segments and results in the observed age differences in CD.

To test this hypothesis, a series of simulations was conducted with the Task Dynamics Application (TaDA)[7] based on Noiray et al. (2019) dataset. The dataset consisted in ultrasound tongue imaging data collected in 3- to -7-year-old children and adults producing CV syllables (/b/, /d/, or /g/ in six vowel contexts), and preceded by the word /aɪnə/. CD was measured as the relationship between the horizontal position of the tongue body at the acoustically determined temporal midpoint of the consonant and the position of tongue body at acoustically determined temporal midpoint of the subsequent vowel. The age differences in variability of representations were simulated by manipulating the blending parameter for the gestures in TADA. Blending parameter is a part of gestural representation that determines how susceptible a gesture is for influence from other, temporally overlapping gestures. The results suggest that simulations with higher variability in gestural parameters indeed result in higher CD, as was observed for children in the experimental data (Figure 1). However, the explanatory power of higher variability in representation varied depending on the consonant. This implies that higher variability in representations is necessary but not sufficient to account for developmental differences in coarticulation patterns.

This work contributes to an improved understanding of phonological representations and their development. It also helps dissect the relative role of developing phonological representations with respect to the experimentally observed age-related changes in coarticulation patterns.

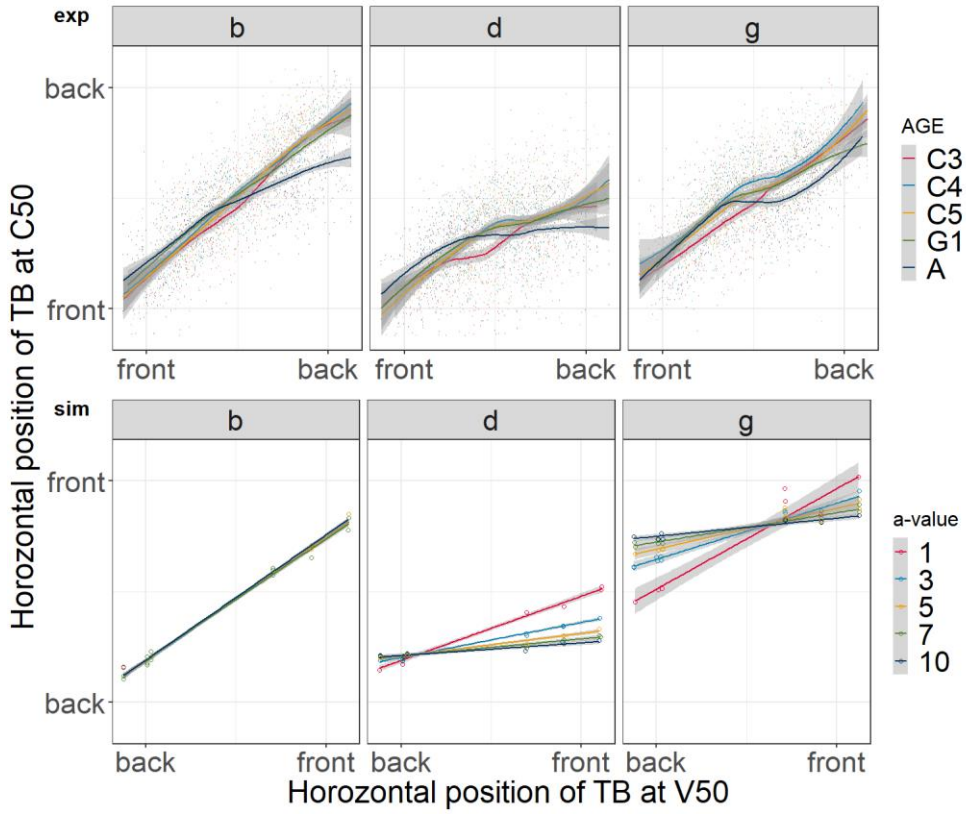


Figure 1. *The regression lines resulting from the regression of the horizontal position of the tongue body at the acoustically determined temporal midpoint of the consonant (C50) on the position of tongue body at acoustically determined temporal midpoint of the subsequent vowel (V50). The slope of the line corresponds to CD. A-value negatively corresponds to variability in representation.*

References

[1] Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. 2019. Back from the future: Nonlinear anticipation in adults' and children's speech. Journal of Speech, Language, and Hearing Research , 62(8S).

[2] Nittrouer, S., Studdert-Kennedy, M., & McGowan Richard, S. 1989. The emergence of phonetic segments. Journal of Speech, Language, and Hearing Research, 32(1), 120–132.

[3] Zharkova et al., 2011.

[4] Noiray, A., Popescu, A., Killmer, H., Rubertus, E., Krüger, S., & Hintermeier, L. (2019). Spoken Language Development and the Challenge of Skill Integration. Frontiers in Psychology, 10.

[5] Vihman, M. M., & Keren-Portnoy, T. 2013. The Emergence of Phonology: Whole-word Approaches and Cross-linguistic Evidence. Cambridge University Press.

[6] Redford, M. A. (2019). Speech Production From a Developmental Perspective. Journal of Speech, Language, and Hearing Research, 62(8S), 2946–2962.

[7] Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. 2004. TADA: An enhanced, portable Task Dynamics model in MATLAB. The Journal of the Acoustical Society of America, 115(5), 2430.

The influence of fundamental frequency on gender perception in children's voices – Results from a longitudinal study

Riccarda Funk, Melanie Weirich, Adrian P. Simpson
Friedrich-Schiller-Universität Jena

Average fundamental frequency (*f0*) of women and men differs significantly. Average adult female *f0* is higher due to shorter and thinner vocal folds and average male resonance frequencies are lower due to a disproportionate lowering of the larynx during puberty [1]. These anatomical and physiological differences play a significant role in correct gender identification of voiced stimuli and can lead to rates close to 100% [2]. In contrast, such differences are marginal in prepubertal voices [3, 4]. However, studies have found different *f0*s of girls and boys before reaching puberty [5, 6, 7]. In addition, listeners can determine gender in children's voices at above-chance levels [8, 9, 10]. Whereas the identification rates for some children are ambivalent, others receive an unambiguous gender assignment [8]. When listeners rate stimuli of such children against a number of perceptual attribute pairs, e.g. high–low, strong correlations between perceptual ratings and acoustic measures can be found [8].

The present study investigates the differences in *f0* in prepubertal girls and boys. Acoustic recordings of a spontaneous picture description and repeated sentences of the same 55 German primary school children (26 girls, 29 boys) were made at three time points: first, second and third grade (6-9 years old). At each time point, over 100 listeners gave their evaluations regarding gender perception on a seven-point scale from 1 = male to 7 = female (experiment 1). Gender ratings were averaged across listeners. Afterwards, $f0_{mean}$, $f0_{min}$, $f0_{max}$ and semitone range of all girls and boys and of the most female- and male-sounding children were compared in two sample t-Tests. To examine the influence of *f0* on the perception, linear regression models were run. In experiment 2, 102 listeners judged the perceived pitch and melodiousness of the five girls and boys who sounded most female/male on a seven-point scale to explore possible relationships between perceptual pitch/range and measured *f0*/semitone range. Due to the longitudinal nature of the project, changes in *f0* and gender perception over time can be investigated.

Not surprisingly, the gender perception ratings of the girls and boys differ significantly at all three time points. If all children are included in the acoustic analysis, no significant differences between the girls and boys of the first and second grade can be found. In the third grade, $f0_{max}$ is significantly higher in the female group. When we compare the ten most female- and male-sounding children, significant differences can be found for $f0_{mean}$, $f0_{min}$ and $f0_{max}$, especially in the first grade. Changes in *f0* over time show a significant decrease of $f0_{mean}$, $f0_{max}$ and semitone range from the first to second grade for all children, especially for boys. This decline also continues from the second to the third grade but does not reach significance. Experiment 2 shows a significant Spearman's ρ correlation between perceived pitch and measured $f0_{mean}$. Furthermore, the female-sounding children are perceived as being significantly higher pitched. No correlation between perceived pitch range and measured semitone range was found, although the female-sounding children are perceived as significantly more melodious than the males. The linear regression models verify a significant influence of $f0_{mean}$, $f0_{min}$ and $f0_{max}$ on gender perception in interaction with gender but not with time point (see Fig. 1). When we split the children by sex, significant relationships between gender perception and $f0_{mean}$, $f0_{min}$ and $f0_{max}$ can only be found for boys. In the female group, gender perception is not predicted by $f0_{mean}$, $f0_{min}$ and $f0_{max}$. It is possible that other parameters like tempo or HNR have a greater influence on gender perception in this case. Semitone range doesn't seem to have any effect on gender perception at all.

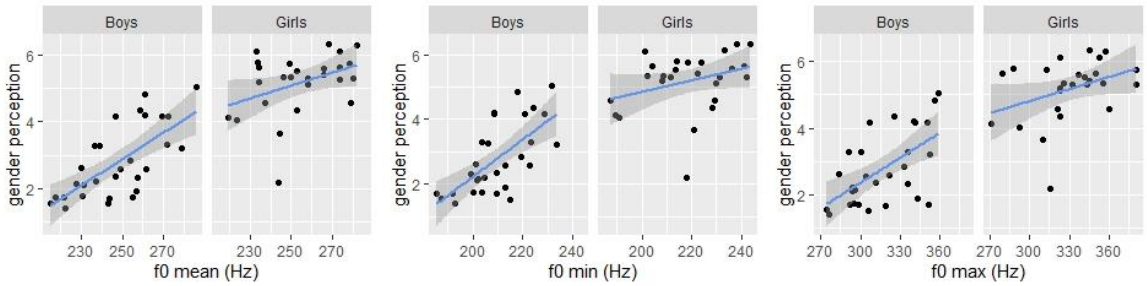


Figure 1. *Relationship between gender perception ratings and $f0_{mean}$, $f0_{min}$ and $f0_{max}$ of boys and girls (mean of all time points).*

References

[1] Simpson, A. P. 2009. Phonetic differences between male and female speech. *Language and Linguistics Compass* 3(2), 621-640.

[2] Whiteside, S. P. 1998. Identification of a speaker's sex: a study of vowels. *Perceptual and Motor Skills* 86(2), 579-584.

[3] Kahane, J. C. 1978. A morphological study of the human prepubertal and pubertal larynx. *Journal of Anatomy* 151, 11-20.

[4] Fitch, W. Tecumseh & J. Giedd. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106(3), 1511-1522.

[5] Hasek, C. S., S. Singh, & T. Murry. 1980. Acoustic attributes of children's voices. *Journal of the Acoustical Society of America* 68, 1262-1265.

[6] Ferrand, C.T. & R. L. Bloom. 1996. Gender differences in children's intonational patterns. *Journal of Voice* 10(3), 284-29.

[7] Glaze, L. E., D. M. Bless, P. Milenkovic, & R. D. Susser. 1988. Acoustic characteristics of children's voice. *Journal of Voice* 2(4), 312-319.

[8] Simpson, A. P., R. Funk, & F. Palmer. 2017. Percep-tual and acoustic correlates of gender in the prepubertal voice. *Interspeech 2017*. Stockholm, 914-918.

[9] Günzburger, D., A. Bresser, & M. ter Keurs. 1987. Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects. *Language and Speech* 30, 47-58.

[10] Kaya, H., A. A. Salahb, A. Karpovc, O. Frolovae, A. Grigoreve, & E. Laykso. 2017. Emotion, age, and gender classification in children's speech by humans and machines. *Computer Speech and Language* 46, 268-283.

Contrastive focus production and perception in 3-5 year-old Swedish children from two regional varieties with and without categorical intonational marking of focus

Gilbert Ambrazaitis[1], Nadja Althaus[2], Charlotte Bertilsson[1], Simone Löhndorf[3], Anna Sara H. Romøren[4] and Susan Sayehli[5]

[1]*Linnaeus University, Sweden,* [2]*University of East Anglia, UK,* [3]*Kristianstad University, Sweden,* [4]*Oslo Metropolitan University, Norway,* [5]*Stockholm University, Sweden*

Despite several studies demonstrating sophisticated prosodic discrimination in infant perception (see [1] for a review), research on the use of prosody for encoding information structure (IS) suggests this skill to be mastered fairly late in children's language development. However, although children's prosodic marking of IS has been studied for various languages using a range of experimental set-ups (e.g., [2]-[11]), we still only have limited knowledge on the relation between children's production and perception of prosodically marked IS [12]. Few studies have conducted parallel production and perception experiments. Furthermore, earlier studies involving perception have made use of offline paradigms (e.g., [3]), while more recent studies using online methods such as eye tracking have usually not included children younger than six years of age and have not been complemented by production data (e.g., [7]).

We also know relatively little about how language-specific aspects of IS coding might impact children's mastering of IS coding. Previous work has indicated that language typology indeed might play a role [9]. For instance, Stockholm Swedish speaking children master the use of a prominence marking H(igh) tone for focus earlier than Dutch speaking children master the use of pitch accents for focus [8][11]. One possible explanation is that the complex contours resulting from the combination of lexical accent + prominence H in Stockholm Swedish make prosodic focus marking particularly salient. Another is that the presence of lexical accents makes Swedish speaking children particularly sensitive to prosodic contrasts. However, these studies have usually had a strict focus on speech production.

In this study we explore the production and perception of intonationally marked contrastive focus in 3-5 year-old children speaking either Scanian or Stockholm Swedish, two dialects which differ crucially in the way focus is encoded phonologically. While both dialects exhibit a lexical accent contrast, focus is phonetically marked more subtle in the Scanian variety [13][14]: instead of adding a prominence H-tone for focus, phrase-level prominence is encoded through phonetic adjustments of the HL accent patterns determined by the lexical accent contrast. By comparing these two Swedish varieties we can thus control for other phonological features (incl. lexical tone), as well as grammar and lexicon, when exploring whether the dialect-specific phonetic realization of contrastive focus affects the way contrastive focus is both perceived and produced by children speaking these dialects.

To this end, we have designed a production and a perception experiment. The production part involves eliciting adjective-noun phrases in three different focus conditions, see (1), using an interactive video/card game (Fig. 1). The task of the participant is to help the experimenter pack a toy suitcase with objects printed on cards, by telling the experimenter which one two objects at a time (displayed on a screen) to put in the suitcase (object marked by a red circle). Focus conditions are elicited by appropriate compositions of objects and colors (e.g., adjective focus: two identical objects with different colors). Production data are analyzed acoustically and auditorily as a function of age and dialect, as well as compared to data from adult controls.

In our visual-word eye-tracking experiment (inspired by [7]), we use the same pictures of colored objects as in the production experiment to investigate whether and how children make use of contrastive intonation for reference resolution (e.g., *Where is the yellow boat? And where is the GREEN boat?* See Fig. 1 (right). Eye-tracking data are analyzed using growth curve analysis [15]. Data from children of both dialects, as well as adult controls, are currently being collected, and preliminary results will be presented at the conference.

(1) focus conditions (examples)
a. broad focus
den gröna båten
the green boat
b. focus on adjective
den GRÖNA båten
the GREEN boat
c. focus on noun
den gröna BÅTEN
the green BOAT

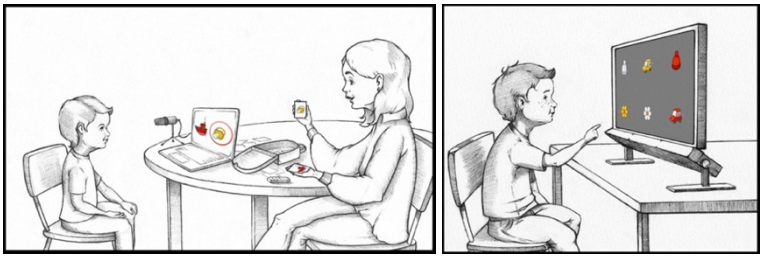


Figure 1. *Illustrations of the experimental set-up. Left: Production experiment; right: Perception (eye-tracking) experiment.*

References

[1] Frota, S., & Butler, J. 2018. Early development of intonation. In Prieto, P., & Esteve-Gibert, N. (Eds.), *The development of prosody in first language acquisition*. Amsterdam: John Benjamins, 145-164.

[2] MacWhinney, B., & Bates, E. 1978. Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behavior* 17, 539-555.

[3] Wells, B., Peppé, S., & Goulandris, N. 2004. Intonation development from five to thirteen. *Journal of Child Language* 31, 749-778.

[4] de Ruiter, L. 2010. *Studies on intonation and information structure in child and adult German*. PhD diss., Max Planck Institute for Psycholinguistics, Nijmegen.

[5] Chen, A. 2011. Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language* 38, 1055-1083.

[6] Grünloh, T., Lieven, E., & Tomasello, M. 2014. Young children's intonational marking of new, given and contrastive referents. *Language Learning and Development* 11(2), 96-127.

[7] Ito, K. 2014. Children's pragmatic use of prosodic prominence. In Matthews, D. (Ed.), *Pragmatic development in first language acquisition*. Amsterdam: John Benjamins, 199-218.

[8] Romøren, A. S. H., & Chen, A. 2015. Quiet is the new loud: Pausing and focus in child and adult Dutch. *Language and speech* 58(1), 8-23.

[9] Prieto, P., & Esteve-Gibert, N. (Eds.). 2018. *The development of prosody in first language acquisition.* Amsterdam: John Benjamins.

[10] Esteve-Gibert, N., Loevenbruck, H., Dohen, M., & D'Imperio, M. 2021. Pre-schoolers use head gestures rather than typical prosodic cues to highlight important information in speech. *Developmental Science* 25, e13154.

[11] Romøren, A. S. H., & Chen, A. 2021. The acquisition of prosodic marking of narrow focus in Central Swedish. *Journal of Child Language*, 49(2), 213-238.

[12] Chen, A. 2010. Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua* 120(8), 1926-1939.

[13] Bruce, G., & Gårding, E. 1978. A prosodic typology for Swedish dialects. In Gårding, E., Bruce, G., & Bannert, R. (Eds.), *Nordic Prosody – Papers from a Symposium*. Lund University, 219-228.

[14] Ambrazaitis, G., Frid, J., & Bruce, G. 2012. Revisiting Southern and Central Swedish intonation from a comparative and functional perspective. In Niebuhr, O. (Ed.), *Understanding prosody – The role of context, function, and communication*. Berlin: de Gruyter, 135–158.

[15] Mirman, D. 2014. *Growth Curve Analysis and Visualization Using R*. Chapman and Hall/CRC.

Multimodal Prosody

Sunday, oral session 9

The processing of prosodic prominence in German

Barbara Zeyer[1], Martina Penke[1]
[1] *University of Cologne*

Prosodic prominence can be used to draw the listeners attention to a particular entity within an utterance. Prosodically prominent entities are defined to stand out from their environment based on their prosodic characteristics such as loudness, pitch accent, or duration [1]. Studies have shown that prosodically prominent words get recalled better compared to deaccented words during recall tasks. For example, Kember and colleagues [2] found out that, inter alia, their participants recalled words in a sentence better when they were presented with a prosodically prominent accent type compared to words that did not receive a prominent accent type. The results indicate that prosodic prominence serves to draw the attention of a hearer to the accented entity, thus, enabling the hearer to recall prominent words better.

To examine if prosodic prominence also has a more immediate effect on ongoing language processing, we conducted a word-monitoring task with reaction time measurement with 50 native speakers of German. The participants were asked to identify a visually presented target word in a sentence that was subsequently presented auditorily either with or without a prosodically prominent accent. We, also, wanted to explore the influence of different levels of prosodic prominence on word identification times. We therefore manipulated the prosodic prominence of the auditorily presented target words. Following the prominence scale established by Baumann and Röhr [3], target words were either presented with the prosodically highly prominent accent types *LH** and *L*H*, with the less prominent accent type *L** or they were deaccented.

We expected a prosodically prominent accent on a target word to draw the attention of a listener to this word, thereby leading to faster word identification when the target word was presented with a prominent accent type (*LH**, *L*H*, *L**) compared to when it was deaccented (*ø*, our baseline condition). Furthermore, we expected identification times to decrease with increasing prosodic prominence of the target word. The more prosodically prominent a word is, the more attention it should draw, leading to shorter identification times. Thus, based on the findings by Baumann and Röhr [3], we expected reaction times to display the following scale: *LH* < L*H < L* < ø.*

Ten sentences were presented for each of the four experimental conditions (target word presented with *LH*, L*H, L** or *ø*). All targets words were bi-syllabic adverbials, controlled for their articulatory duration in ms, their word frequency and their number of phonemes across the experimental conditions. All experimental sentences had the same syntactic structure (Adv V S target word O). In addition, we presented 100 filler sentences in which the prosody of the target word was not manipulated. In these filler sentences the target word was either the subject or the object and occurred in different positions in the sentence.

Our results showed that the condition *L*H* led to the fastest mean reaction time (543.75 ms, *SD* = 158.75), followed by *ø* (544.96 ms, *SD* = 139.01), *LH** (552.78 ms, *SD* = 144.63), and *L** (562.89 ms, *SD* = 161.42). A linear mixed effects model (*lme4* package [4] in R [5]) on the measured reaction times indicated no significant random effects. We then calculated pairwise comparisons (accent types *LH**, *L*H*, and *L** vs. *ø* respectively) using the *emmeans* package [6] in R [5]. Again, the results of the comparisons remained non-significant.

While previous studies have found an effect of prosodic prominence on word recall, our findings suggest that prosodic prominence might not exert an immediate influence on language processing in speeding up word identification in a word-monitoring task. Baumann and Röhr [3] presented words with seven nuclear accent types in German and asked the participants to rate how highlighted (i.e., prominent) the words sounded. They found a graded effect of perceived prosodic prominence (*LH* > L*H > L* > ø*) that we could not replicate in our study.

References

[1] Terken, J., Hermes, D. 2000. The perception of prosodic prominence. In: Horne, M. (eds.), *Prosody: Theory and Experiment*. Kluwer, 89–127.

[2] Kember, H., Choi, J., Yu, J., & Cutler, A. (2021). The Processing of Linguistic Prominence. *Language and Speech*, *64*(2), 413–436.

[3] Baumann, S., Röhr, C. 2015. The perceptual prominence of pitch accent types in German. *18th Proc ICPhS 2015*.

[4] Bates, B., Maechler, M., Bolker, B., Walker, S. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-27.1.

[5] R Core Team. *R: A language and environment for statistical computing* (2021.09.0). R Foundation for Statistical Computing.

[6] Lenth, R., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Migues, F., Riebl H., Singmann, H. 2022. Estimated Marginal Means, aka Least-Square means. Version 1.8.3.

The multimodal marking of information status in French as a foreign language: What can we learn about the use of prosodic and gestural cues in an interlanguage?

Florence Bailly[1,2] & Stefan Baumann[2]
[1]*Universitat Pompeu Fabra,* [2]*Institut für Linguistik-Phonetik, Universität zu Köln*

Speakers of natural languages signal information structure through different linguistic means, i.e., through the choice of words and their order, prosody, and gestures. For instance, it has been shown that prosodic prominence, i.e. the presence of pitch accents, and the production of manual co-speech gestures are both used to mark informationally relevant material in speech, such as new referents and focused or contrastive constituents [see 1 for a review on prosody; 2 for a review on gesture; and 3 for a joint analysis of prosody and gesture]. Regarding non-manual gestures, studies looking at head movements suggest that these play a role as visual prominence markers [4], but their relation to information status marking has not been established yet.

Since languages differ in their dominant strategy to signal information structure, this discrepancy may represent a challenge for language learners. Previous research has shown that L2 speakers tend to transfer L1 prosodic patterns [5] or may even use completely new patterns [6]. As a consequence, adequately signalling information status, for example by deaccenting *given* information, may be difficult for speakers of languages which use this strategy less [7]. As for gestures, there is evidence that learners tend to over-explicitly mark referring expressions such as pronouns [8] but the marking of information status by using gestures in L2 speech has not been studied yet, let alone the role of head movements.

The present study investigates the joint use of prosodic prominence and head movements, i.e., the types of pitch accent and the types of head movement used to mark information status by Catalan learners of French with an intermediate proficiency. Romance languages like Catalan and French may rely more on syntactic movements to encode information structure, however, when sufficient proficiency is not yet attained, learners may express information structure through prosody and non-verbal cues.

An audio-visual corpus of 50 short narrations by 25 Catalan learners of French was obtained by video-recording them giving a short description of a) their best friend and b) their recent stay abroad. The recordings were annotated in terms of information status with the RefLex Scheme [9], pitch accent types [10], perceived prominence, and head movement types (nods, protrusions, tilts, slides) and apexes [11].

Results show that Catalan learners of French marked *new* and *inferable* information more frequently than *given* information with a combination of pitch accents and head movements (see Figure 1). Accordingly, *given* information was marked - and perceived - as less prominent than *new*(er) information (more initial accents, fewer accentual rises and head movements, lower mean prominence score) but still received a large proportion of pitch accents. Head nods were produced significantly more often with *new* referents. Furthermore, the high pitch accent H* was by far the most frequently used but did not play a disambiguating role in terms of information status. Currently, the annotation and analysis of comparable speech by ten French native speakers who performed the same narrative task is being annotated, and the result of their analysis will be contrasted with the present results in order to detect and describe potential differences between L1 and L2 speech.

References

[1] Kügler, F., & Calhoun, S. 2020. Prosodic encoding of information structure: A typological perspective. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 454-467). Oxford Academic.

[2] Holler, J. & Bavelas, J. 2017. Multimodal communication of common ground: a review of social functions. In R. B. Church, M. W. Alibali & S. D. Kelly (Eds.), *Why Gesture? How the hands function in speaking, thinking and communicating* (pp. 213-240). John Benjamins.
[3] Rohrer, P. 2022. A temporal and pragmatic analysis of gesture-speech association: A corpus based approach using the novel MultiModal MultiDimension (M3D) labelling system [PhD dissertation, Universitat Pompeu Fabra].
[4] Ambrazaitis, G., House, D. 2022. Probing effects of lexical prosody on speech-gesture integration in prominence production by Swedish news presenters. *Laboratory Phonology, 13*, 1–35.
[5] van Maastricht, L., Krahmer., E., & Swerts, M. 2016. Prominence patterns in a second language: Intonational transfer From Dutch to Spanish and vice versa. *Language Learning, 66*, 124–158.
[6] Santiago, F. & Delais-Roussarie, E. 2015. What motivates extra-rising patterns in L2 French: Acquisition factors or L1 Transfer? *Proceedings of the 18th Congress on Phonetic Sciences*. Glasgow, UK.
[7] Rasier, L. & Hiligsmann, P. 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux cahiers de linguistique française, 28*, 41–66.
[8] Yoshioka, K. 2008. Gesture and information structure in first and second language. *Gesture, 8*, 236–255.
[9] Riester, A. & Baumann, S. 2017. The RefLex Scheme – Annotation Guidelines. *SinSpeC. Working Papers of the SFB 732, vol. 14*. University of Stuttgart.
[10] Hualde, J.I. & Prieto, P. 2016. Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology* 7, 5.
[11] Rohrer, P., Vilà-Giménez, I, Florit-Pons, J., Esteve-Gibert, N., Ren, A., Shattuck-Hufnagel, S., Prieto, P. 2020. *The MultiModal MultiDimensional (M3D) labeling system*.

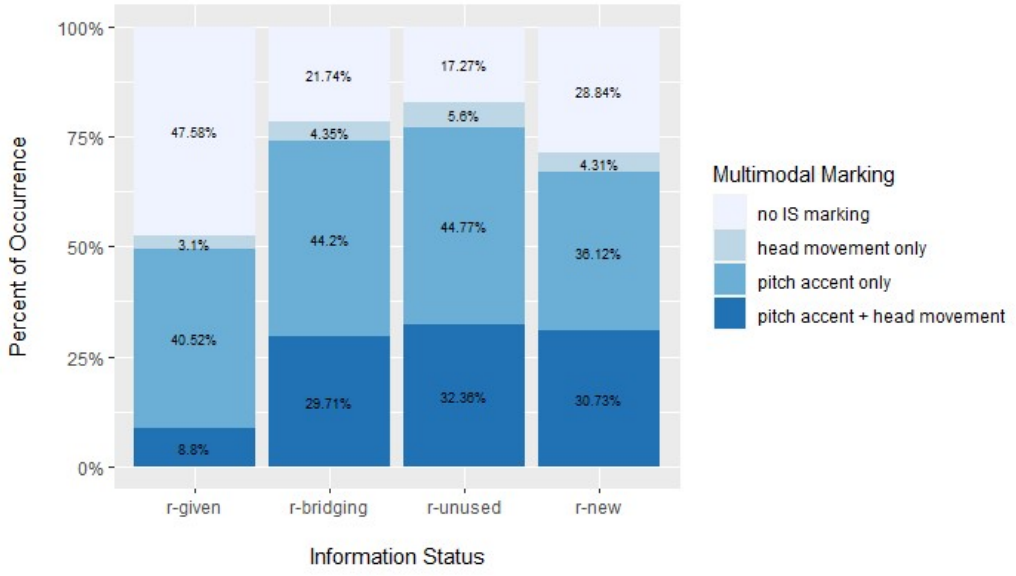


Figure 1. *Multimodal marking of information status (referential level of* RefLex*; Riester & Baumann 2017). R-given refers to the referents present in the textual or non-textual previous context; r-bridging refers to discourse-new referents derivable from or dependent on the previous context; r-unused refers to identifiable, discourse-new referent (indefinite expressions); and r-new refers to non-identifiable, discourse-new referent (indefinite expressions)*

Phrase initial strengthening effects in gesture production

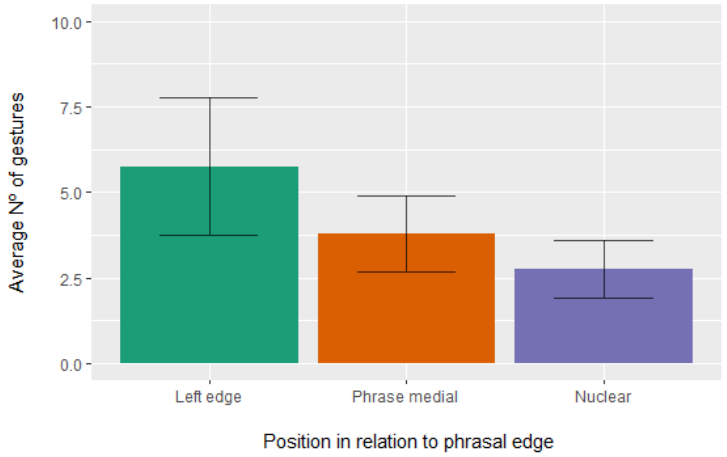
Patrick Louis Rohrer[1,2], Elisabeth Delais-Roussarie[2], Pilar Prieto[1,3]

[1]*Grup d'Estudis de Prosòdia, Universitat Pompeu Fabra, Barcelona, Catalonia*
[2]*Laboratoire de Linguistique de Nantes (LLING) – UMR 6310, Université de Nantes, Nantes, France*
[3]*Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia*

Prosodic structure consists of not only prosodic heads (e.g., pitch accentuation) but also of prosodic edges (loosely understood here as initial and final positions within a prosodic phrase). Importantly, some researchers have described phrase-initial strengthening effects, calling for models of speech production where speakers regularly aim to place a prenuclear pitch accent as early in an intermediate phrase as possible [1][2]. In the field of gesture studies, much research has shown how gesture and pitch accentuation are closely temporally coordinated (see [3] for a recent review). While some studies have investigated how prosodic phrasing indeed plays a role in the temporal execution of gesture [4][5][6], no previous study to our knowledge has explicitly assessed the value of prosodic edges (in terms of nuclear vs. prenuclear pitch accentuation) in the attraction of manual gestures while at the same time controlling for the relative degree of prominence associated with the pitch accents and it's structural position within the phrase in an independent manner. The current study adds to our knowledge of how gesture production is integrated with phrasal prosodic structure by assessing the following questions, namely (a) whether gesture strokes align more with nuclear than prenuclear pitch accents at the intermediate phrase level; and (b) whether this relationship is driven by prominence relations or by phrasal position.

A prosodic and gestural analysis of the English M3D-TED corpus was carried out, which contains a total of 5 academic lectures with over 23 minutes of multimodal speech. Gesture was coded according to the M3D gesture labeling system [7], and speech prosody was annotated following MAE-ToBI [8]. Results revealed that at the phrasal level, strokes tend to align with phrase-initial prenuclear pitch accents over phrase-medial or nuclear accents (Fig. 1), and this relationship is not driven by prominence relations between the pitch accents. All in all, these findings show that not only prosodic heads, but also prosodic edges (referring to the first prenuclear pitch accent), act as strong attractors of manual gestures, and highlights how a phrase-initial strengthening effect may be present both in prosody as well as gesture.

Figure 1. *Gesture association as a function of phrasal position of the pitch accent.*



References

[1] Bolinger, D. 1985. Two views of accent. *Journal of Linguistics* 21(1), 79–123.

[2] Shattuck-Hufnagel, S., Ostendorf, M. & Ross, K. 1994. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22(4), 357–388.

[3] Shattuck-Hufnagel, S. & Ren, A. 2018. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology* 9.

[4] Loehr, D. P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3(1), 71–89.

[5] 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), 3.

[6] Esteve-Gibert, N. & Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research* 56(3), 850–864.

[7] Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Gibert, N. E., Ren, A., Shattuck-Hufnagel, S., & Prieto, P. (2021, February 24). *The MultiModal MultiDimensional (M3D) labeling system*. https://doi.org/10.17605/OSF.IO/ANKDX

[8] Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. B., & Hirschberg, J. 1992. TOBI: A standard for labeling English prosody. *ICSLP-1992*, 867–870.

POSTER SESSIONS

(in order of appearance in the programme)

Poster Session 1

Saturday, 10:00 - 11:20

Prosody-driven word segmentation in infants acquiring Czech

Kateřina Chládková[1,2], Václav Jonáš Podlipský[3], Nikola Paillereau[1,2], Natalia Nudga[1,2], Šárka Šimáčková[3]

[1]Czech Academy of Sciences, [2]Charles University, [3]Palacký University

Humans employ statistical learning mechanisms to uncover regularities in their environment [1]. To segment words from continuous speech, infants track the transitional probabilities (TPs) between syllables [2, 3]. Besides the TP cues, infants and adults also segment words on the basis of prosody [4] and phonotactics [5]. Which of those cues is primary is language-specific: for instance, our recent work showed that adult speakers of different dialects of Czech could segment words in a novel artificial language only when word boundaries were cued by prosody and native phonotactics [6]. Here, we tested the reliance on TPs and prosody in Czech-learning 8-month olds.

Czech is a fixed-stress language. Its Central-Bohemian (CB) variety has word-initial stress, which in connected speech is realized only at the phrasal level as is the case in e.g. French [7]. In contrast, the Moravian-Silesian variety of Czech has phonetically prominent stress fixed to the penultimate (longer and louder) syllable. We hypothesized that the placement of word stress and its phonetic saliency in the L1 dialect would predict how infants use TPs and prosody cues to segment words: MS infants would rely on prosody to a greater extent than CB infants and would parse words according to penultimate rather than word-initial stress.

119 infants, 7.5 to 8.5 months old, took part in the experiment (23 additional infants were excluded; data collection is still ongoing aiming at n = 176, i.e. 44 per condition and per dialect). They were born full-term, had normal hearing and no familial risk of dyslexia. 72 infants were learners of MS Czech (44 of them were tested in condition 1, 28 in condition 2), and 47 were learners of CB Czech (37 tested in condition 1, 10 in condition 2). During training, they were exposed to a 3-minute stream of syllables, following the design of [8]. In condition 1, the only cue to trisyllabic word boundaries were TPs. In condition 2, prosody cues were added: the initial syllables of statistical words (which were also the penultimate syllables of partwords) were made longer and with a raised F0. After training, infants' looking times to trials with statistical words or part-words (always without prosody cues) were assessed in a central-fixation paradigm: a trial contained a maximum of 12 repetitions of the target word for a duration of 15 s, or ended when an infant looked away for 2 s. There were 12 test trials in total, pseudorandomized. If infants recover the TP-cued statistical word boundaries, they should look longer to part-words (i.e. to novelty) during test trials.

Infant looking times (log-transformed) were analyzed with a linear mixed-effects model in R. Table 1 shows the model formula and the fixed-effects output. The triple interaction of Word type, L1 dialect, and Condition is relevant to our hypothesis. Pairwise comparisons, plotted in Figure 1, indicate that learning effects were found only in the TP-plus-Prosody condition, and these were different between MS and CB infants. MS infants preferentially looked to statistical words that had word-initial stress in training than to part-words that had penultimate stress in training, which suggests that they parsed the syllable stream according to penultimate prosodic prominence. An opposite trend was seen in CB infants: i.e., longer looking times to part-words that had penultimate stress during training than to statistical words that had initial stress in training. (But note the small sample size in this condition as data collection is still ongoing, and expected to complete before the conference.)

Our preliminary results suggest that 8-month-olds acquiring a fixed-stress language segment trisyllabic words from a novel speech stream only when prosodic cues are present. Whether they parse words according to word-initial or penultimate stress depends on stress placement in their L1 dialect.

Model: log(Looking time)~Word type*Cond.*L1 dialect*Fam.lang.+Trial nr.+(1+Word type|Subject)+(1|Word item)					
Fixed effects	*Estimate*	*SE*	*df*	*t*	*p*
Intercept	2.051	0.035	32.560	58.553	<0.001
Word type (-part.word +stat.word)	0.005	0.015	111.000	0.352	0.725
Condition (-TPonly +TPplusProsody)	-0.066	0.033	110.900	-1.984	0.050
L1 dialect (-MS +CB)	-0.067	0.033	110.900	-2.013	0.047
Familiarization language (-A +B, swapping the role of stat.words vs part words)	0.050	0.033	110.900	1.500	0.136
Trial number (mean centered)	-0.210	0.019	1,205.000	-11.010	<0.001
Word type : L1 dialect	-0.037	0.015	111.000	-2.479	0.015
Word type : Condition : L1 dialect	**-0.030**	**0.015**	**111.000**	**-1.995**	**0.048**

Table 1. The model (first row) and its fixed-effects output: all the main effects and significant interaction effects (with alpha = 0.05) are shown. Bold-face interaction unpacked in Fig 1.

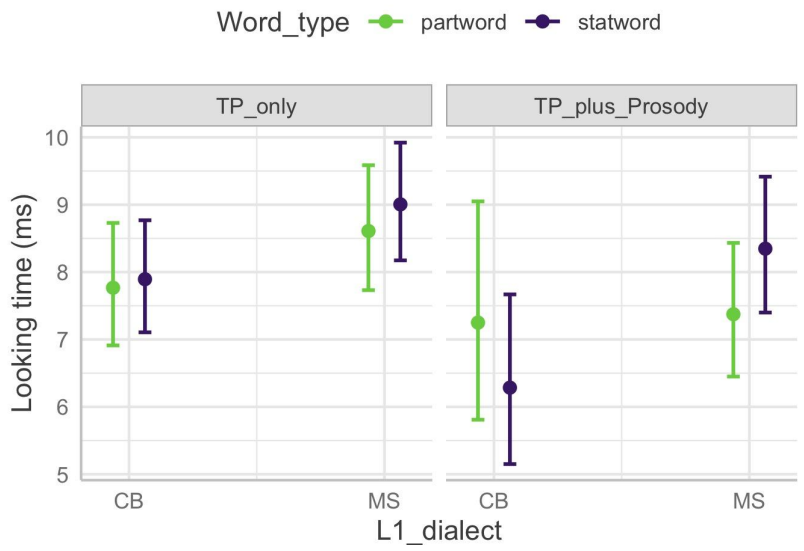


Figure 1. Estimated looking times (means and 95% CIs). Learning effects were observed after training with TP plus Prosody (right graph): CB infants (current n = 10) tended to look longer to part words that had penultimate stress in training, MS infants (current n = 28) looked longer to statistical words that had initial stress in training.

References:

[1] Krogh, L., Vlach, H.A., Johnson, S.P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, 3, 598.

[2] Saffran, J.R., Aslin, R.N., Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.

[3] Isbilen, E.S., Christiansen, M.H. (2022). Statistical Learning of Language: A Meta-Analysis Into 25 Years of Research. *Cognitive Science*, 46(9), e13198.

[4] Johnson, E.K., Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4), 548-567.

[5] Dal Ben, R., Souza, D.D.H., Hay, J.F. (2021). When statistics collide: The use of transitional and phonotactic probability cues to word boundaries. *Memory & Cognition* 49, 1300-10.

[6] Podlipský, V.J., Chládková, K., Paillereau, N., Šimáčková, Š. (2022). Native-variety influence on speech segmentation in a novel language. Talk at *New Sounds*, Barcelona.

[7] Skarnitzl, R., Eriksson, A. (2017). The acoustics of word stress in Czech as a function of speaking style. In *Proceedings of Interspeech* 2017, pp. 3221-3225.

[8] Aslin, R.N., Saffran, J.R., Newport, E.L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321-324.

‘Sounding Asian’ in native German: an experiment on speaker voice and ethnolinguistic identification

Thanh Lan Truong[1] and Andrea Weber[1]
[1]English Department, University of Tübingen, Germany

Starting with Labov’s seminal study on New York City English [1], small differences in segmental and suprasegmental features have been found to facilitate the identification of a speaker’s ethnolinguistic background [2]. For example, speech from European Americans can be reliably distinguished from African American speech based on vowel and voice quality [3]. While previous literature often focused on ethnolects in American English, the current study is the first to investigate the identification of an Asian heritage background in native German. Specifically, we asked native German listeners, who either had a bicultural East Asian heritage background (hereafter, Asian heritage Germans) or a monocultural German background, to identify speakers who were either Asian heritage Germans or monocultural Germans. Based on previous findings, we predicted that German listeners with an East Asian heritage would identify Asian heritage speakers more correctly than monocultural German listeners would [7, 8]. This is because listeners with an Asian heritage are likely to be familiar with Asian heritage speech from their community and thus may be more sensitive to detect speech by other Asian heritage speakers.

Twenty-five Asian heritage Germans and 25 monocultural Germans between the ages of 18 and 35 years participated in the study as listeners (mean age: 25.1; 29 females, 2 undisclosed). Stimuli were recorded from 16 speakers: 8 monocultural Germans (4 male, 4 female) and 8 Asian heritage Germans (4 male, 4 female, born in Germany with Vietnamese parents). All speakers grew up in Southern Germany and were recorded reciting one sentence: *Flöhe können um das hundertfache ihrer eigenen Körperlänge in die Höhe springen* (‘flea can jump a hundredfold time their body length’). We opted for this sentence as its length would provide listeners with enough information for speaker heritage identification and also provide ample data for a variety of phonetic analyses. Participants were presented with all sixteen sentence recordings in random order, and after each sentence they had to indicate if the speaker was either Asian heritage German or monocultural German. R [6] and lme4 [5] were used to perform linear mixed effects analyses (see Fig. 1) with correct identification rates as dependent variable. *Speaker group* and *listener group* were entered as fixed effects and their interaction term was added. *Participants* and *items* were included as random factors with random slopes. There was a main effect of *listener group* (b = -0.82, SE = 0.29, p = .01) and an interaction between *speaker group* and *listener group* (b = 2.06, SE = 0.58, p < .001). While German speakers without an Asian heritage were identified correctly more often by listeners without an Asian heritage (b = 1.2, SE = 0.44, p = .01), German speakers with an Asian heritage were identified correctly more often by listeners with an Asian heritage (b = -0.8, SE = 0.29, p = .01). In fact, listeners without an Asian heritage performed only marginally above chance for the Asian heritage speakers (p > .08). While we found higher accuracy rates for Asian heritage listeners when identifying other Asian heritage speakers, in line with previous research on American English [7, 8], the present results also indicate that a similar advantage exists for monocultural listeners and speakers. That is, all listeners performed more accurately when the speakers matched their heritage background, but Asian heritage listeners could also identify heritage-mismatching German speakers with above chance accuracy, while non-Asian German listeners failed to do the same with Asian heritage speakers. Since the Asian heritage speakers and the monocultural speakers grew up in Germany with German as their native language, the perceptible differences in their speech are rather subtle. To better understand which acoustic cues make the heritage speakers sound Asian, features suggested in previous studies [7], such as a breathier voice, will be analyzed and correlated with the identification results.

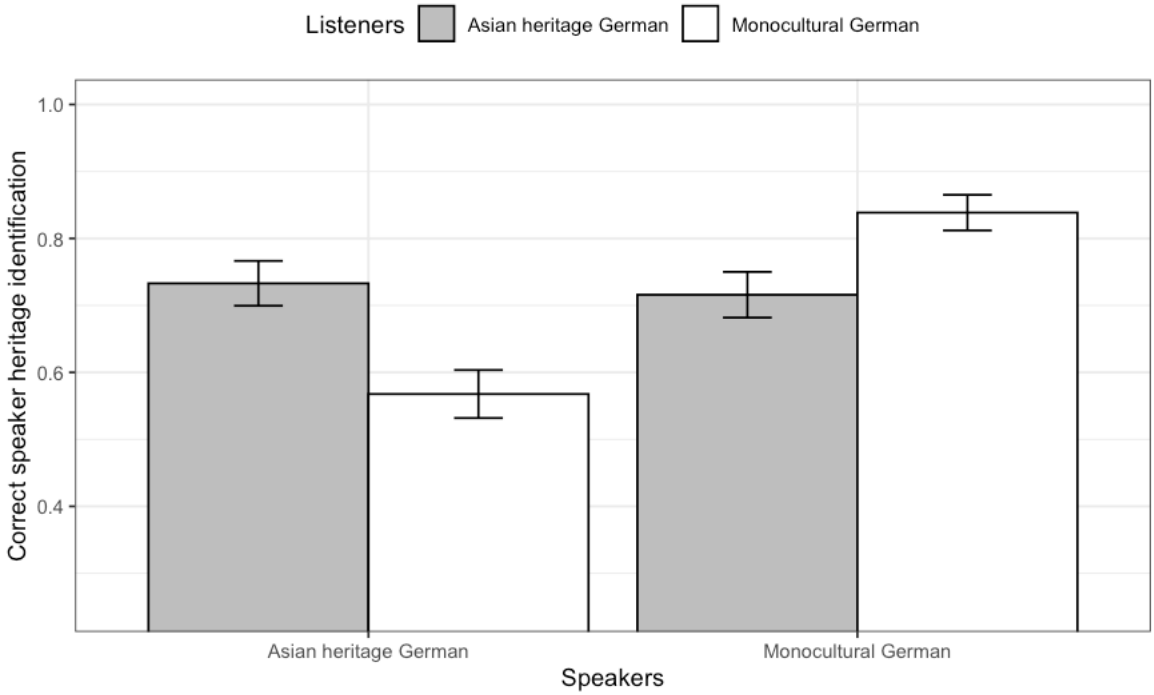


Figure 1. *Average correct identification scores in the heritage identification test.*

References

[1] Labov, W. 1966. *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics.

[2] Kushins, E. R. 2014. Sounding like your race in the employment process: An experiment on speaker voice, race identification, and stereotyping. *Race and Social Problems*, *6*(3), 237–248.

[3] Thomas, E., Lass, N. & Carpenter, J. 2010. Chapter 12. Identification of African American Speech. In D. Preston & N. Niedzielski (Ed.), *A Reader in Sociophonetics* (pp. 265-288). Berlin, New York: De Gruyter Mouton.

[4] Jaeger, T. F. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language, *59,* 434-446.

[5] Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1-48.

[6] R Core Team. 2021. R: A language environment for statistical computing, https://www.r-project.org/ (Last viewed April 2021) [version: 4.0.5, 6].

[7] Newman, M., & Wu, A. 2011. Do you sound Asian when you speak English? Racial identification and voice in Chinese and Korean American English. *American Speech, 86*(2), 152-177.

[8] Hanna, D. B. 1997. Do I sound “Asian” to you? Linguistic markers of Asian American identity. In A. Dimitriadis, L. Siegel, C. Surek-Clark & A. Williams (Eds.), *University of Pennsylvania working papers in linguistics* (pp. 141-153). Philadelphia, PA: University of Pennsylvania Department of Linguistics.

Production of prosodic phrase boundaries in Dutch

Jorik Geutjes, Caroline Junge, Aoju Chen

In spoken language, major prosodic boundaries can be indicated by three types of prosodic cues: *pitch change, final lengthening*, and *pause* [1]. Crucially, although these cues appear cross-linguistically, the weighting of the individual cues in signaling these boundaries is considered to be language-specific. Research on production of prosodic boundaries in Dutch is both scarce and fragmented. Although there are some perception studies suggesting a major role for pauses as boundary cue in Dutch [2, 3], these studies did not include all three types of boundary cues. Moreover, there have been no language production studies that systematically investigate the realization of prosodic boundaries in Dutch. In language acquisition literature, however, it has been suggested that the Dutch prosodic phrasing system may lead to a delay in word segmentation in Dutch-learning infants [4], indicating an urgent need for more research on production of prosodic phrasing in adult- and infant-directed Dutch.

The present study has set out to examine the production of intonational phrase (IP) boundaries in syntactically different constructions in adult-directed Dutch, as part of a large project on development in prosodic phrasing. Two types of constructions containing the same sequence of two names were included: phrases (examples 1a, 1b) and compound sentences (examples 2a, 2b). In the former structure prosodic phrasing serves as the only means to disambiguate (1a) and (1b), whereas in the latter the placement of an IP boundary after 'Demi' in (2b) follows from the syntactic structure of (2b).

(1a) Bella en Demi en Vera
'Bella and Demi and Vera'

(1b) Bella en Demi / en Vera
'Bella and Demi, and Vera'

(2a) Bella speelt met Demi en Vera en ze zitten in hetzelfde team.
'Bella plays with Demi and Vera and they are on the same team.'

(2b) Bella speelt met Demi / en Vera zit in het andere team.
'Bella plays with Demi and Vera is on the other team.'

Target utterances like the examples were elicited from 16 native speakers of Dutch (12 female) using pictures displaying three girls wearing either the same or differently colored shirts to indicate team membership in a fictitious game called 'Teamball' (Table 1). The participants were instructed to indicate the different teams in each picture, using name sequences (e.g. 1a or 1b) in one round and using a template coordinate clause (e.g. 2a or 2b) in another. The order of these rounds was counterbalanced between participants.

The recordings were annotated for analysis on all three types of prosodic boundary cues in *Praat* [5]. In line with the literature, pitch change was operationalized as suspension of declination, determined by measuring the final pre-boundary F_0. Final lengthening was determined by measuring the duration of the final pre-boundary syllable and the duration of the penultimate pre-boundary syllable. Pauses were defined as a between-word silence longer than 20 ms and were treated as a gradient dependent variable. Mixed effects regression modelling yielded a main effect of prosodic boundary (present vs. absent) for all cues (Figure 1), indicating that the speakers used all cues to mark the IP boundary. It also revealed an interaction of Prosodic boundary x Construction type ($ß = -257.89$, $p < 0.000$), indicating that speakers used the pause cue to a larger degree in phrases (1b) than in compound sentences (2b). Furthermore, relative weight analysis [7] was used to determine the weighting of individual cues. It was found that final lengthening had the highest relative weight in predicting the presence of a prosodic boundary, followed by pause duration and pitch change.

Our results thus challenge the notion of pause as the most dominant cue in Dutch prosodic phrasing. They suggest that final lengthening may be the most important cue, at least, from a production perspective. A follow-up EEG study is planned to examine cue weighting in the processing of prosodic phrasing in adult-directed Dutch.

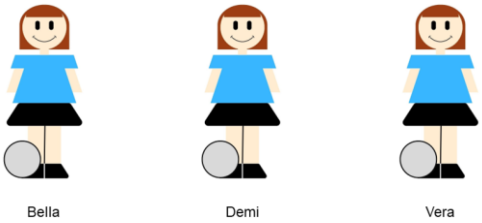
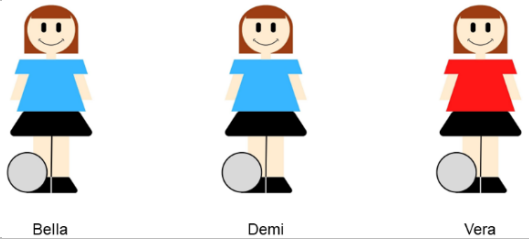
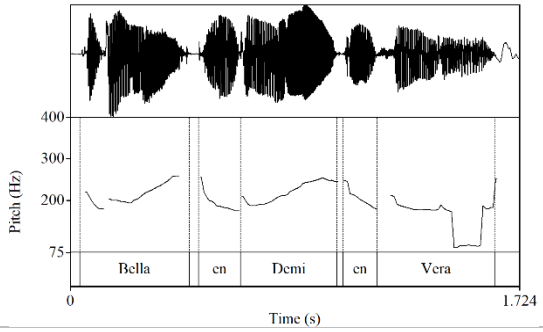
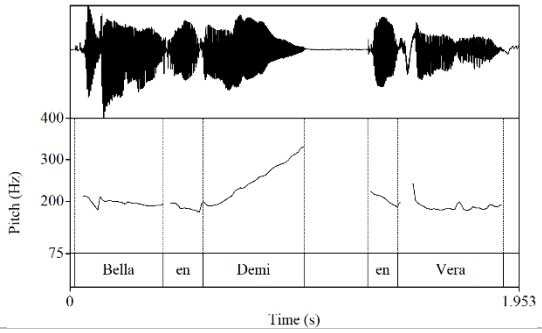
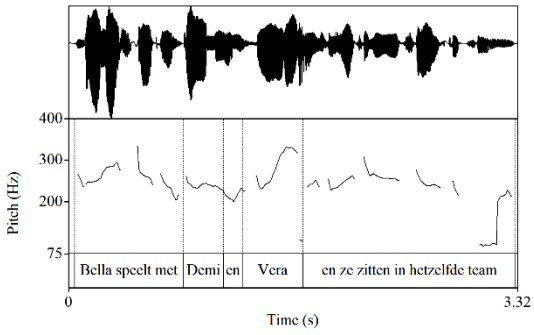
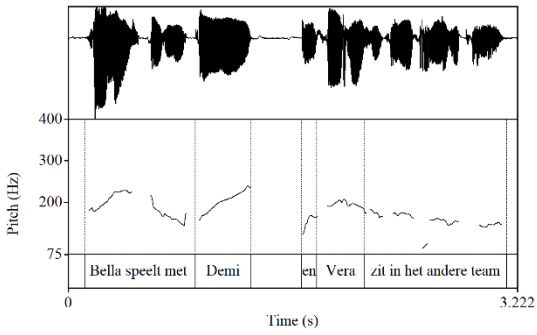
	No-boundary condition	Boundary condition
Picture shown		
Name sequence response	Q: Who make up a team here? A: *Bella and Demi and Vera* 	Q: Who make up the teams here? A: *Bella and Demi / and Vera* 
Coordinate clause response	Q: What is happening here? A: *Bella plays with Demi and Vera and they are on the same team.* 	Q: What is happening here? A: *Bella plays with Demi / and Vera is on the other team.* 

Table 1. Overview of test items and responses in boundary and no-boundary conditions.

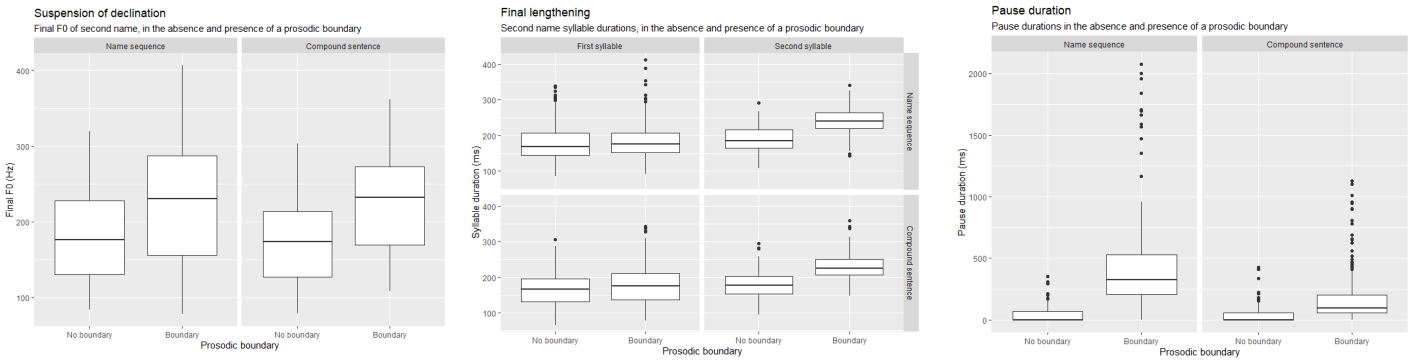


Figure 1. Boxplots of Final F0, syllable durations and pause durations in Boundary and No-boundary conditions

References:

[1] Wagner, M., & Watson, D. 2010. Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*, 25(7-9), 905-945.

[2] De Pijper, J. & Sanderman, A. 1994. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *JASA*, 96(4), 2037-2047.

[3] Swerts, M. 1997. Prosodic features at discourse boundaries of different strength. *JASA*, 101(1), 514-521.

[4] Johnson, E., & Seidl, A. 2008. Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*, 13(5), 440-455.

[5] Boersma, P. & Weenink, D. 2023, *'Praat: doing phonetics by computer'*. http://www.praat.org/.

[6] Cambier-Langeveld, T. 1997. The domain of final lengthening in the production of Dutch. *Linguistics in the Netherlands*,14, 13-24.

[7] Zhang, X. 2012 *A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese*, Doctoral dissertation, University of Michigan.

Production of L2 German stops by Georgian speakers

Nato Sulaberidze
Friedrich Schiller University, Jena

In contrast to German, which exhibits a binary contrast between fortis and lenis stops, Georgian stops exhibit a three-way phonological contrast at the labial, alveolar, and velar places of articulation with voiced (b, d, g), voiceless aspirated (p, t, k), and ejective stops (p', t', k'). In German, however, there is evidence that the final fortis plosives preceding glottalised vowels are commonly produced acoustically and auditorily as ejectives [1, 2]. Thus, ejectives in this language occur only epiphenomenally, i.e., are context-specific and lack phonological status.

It has been suggested that ejectives are produced with laryngeal involvement to achieve the intraoral pressure increase required for the ejective release of the plosive [3, 4]. Yet, in the production of German epiphenomenal ejectives there is no motivation for any vertical movement of the larynx during articulation, as they result from the overlap of a fortis stop and a glottalised vowel [1].

As part of a larger study (funded by DFG) of the production and perception of ejectives in German, Georgian and English, the present analysis investigates how Georgian speakers produce stops in their L2 German in the read speech (a) and in the imitation task (b) of the same elicitation material. The aim is to examine whether Georgian speakers' L2 stop productions differ between the reading and imitation tasks, with the expectation of sounding more native-like in the latter. Furthermore, the study will investigate whether and how (i.e., with which articulatory mechanisms) Georgians produce epiphenomenal ejectives in their L2 German.

In the speech material, the target condition was defined as a word-final stop preceding a word-initial open vowel, which in German is routinely produced with a glottalised onset, e.g.: [hat ʔaʊ̯x]. In the control condition, the stop is followed by a schwa, as in: [hatə ʔaʊ̯x].

For the acoustic and aerodynamic analysis, as well as analysis of larynx activity during stop production, multi-channel recordings were made (microphone, dual-channel electroglottography, intraoral pressure measurement).

In the first task (a), subjects were asked to read German material, which was displayed on a screen in randomised order. In the imitation task (b), performed separately from the task (a), the same subjects had to pronounce the same sentences as in task (a), but in this case, they were asked to imitate the recorded audio stimuli played in the headphones, produced by native German speakers. The audio stimuli were selected according to the following criteria: target stimuli with clear auditory and acoustic ejective production and control stimuli with pulmonically fuelled release of the fortis plosive preceding a schwa.

Preliminary acoustic analysis shows VOT differences in target (fortis stop preceding open vowel) and control (preceding schwa) conditions in both read and imitation tasks, with shorter values in the target position in both German and Georgian speakers (Figure 1). There is a visible tendency for the values in the imitation task to be more similar to the productions of native German speakers, compared to the reading task. There is also evidence that subjects produce more epiphenomenal ejectives in the imitation task rather than in read speech (Figure 2).

Thus, in the imitation task, participants were able to replicate the fine phonetic details of the L2, including those that are not phonologically classified in a language and moreover are unlikely to be learned through traditional didactic methods. This finding supports the notion that imitation can be an effective approach for L2 learners to improve their pronunciation, even to produce very subtle phonetic differences.

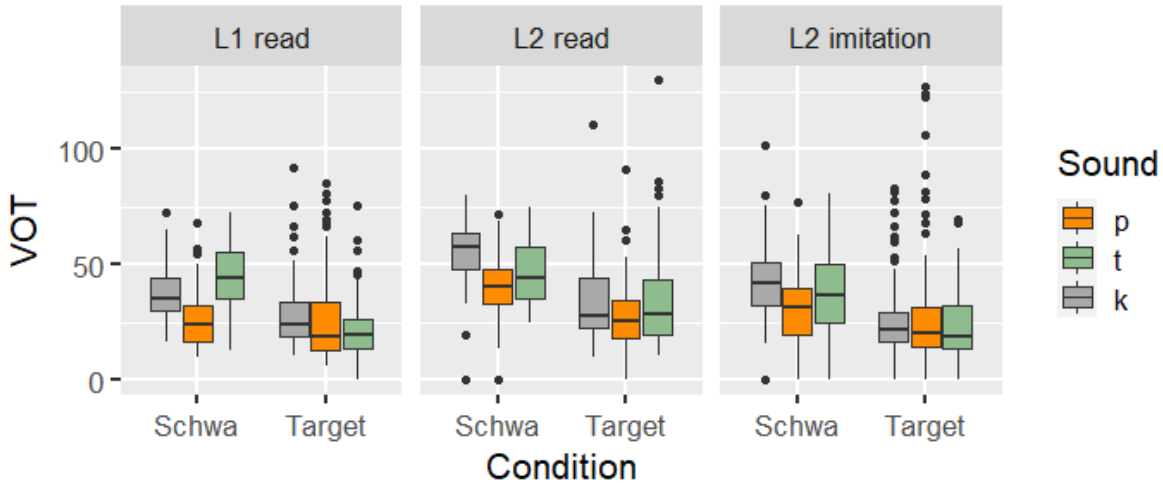


Figure 1. *Voice onset time of three fortis stops in target (preceding open vowel) and control (preceding schwa) conditions produced by German native speakers in read speech (left) and by Georgian learners of German in read speech (centre) and in the imitation task (right).*

Figure 2. *Production frequency of German epiphenomenal ejectives by 5 Georgian speakers in the imitation and read tasks.*

References

[1] Brandt, E. & Simpson, A.P. 2021. The production of ejectives in German and Georgian, *Journal of Phonetics,* Vol. 89

[2] Simpson, A. P. 2014. Ejectives in English and German: Linguistic, sociophonetic, interactional, epiphenomenal? In Celeta, C. Calamai, S. (Eds.), *Advances in Sociophonetics.* Amsterdam: Benjamins, 189-204

[3] Catford, J. 1988. A Practical Introduction to Phonetics. *Oxford: Clarendon Press*

[4] Maddieson, I. 2013. The World Atlas of Language Structures Online. In M. S. Dryer & M. Haspelmath (Eds.). Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/7

The interplay between syllable-based predictability and voicing during closure in intersonorant German stops

Omnia Ibrahim, Ivan Yuen, Bistra Andreeva, Bernd Möbius
Department of Language Science and Technology, Saarland University, Germany
omnia@lst.uni-saarland.de

Contextual predictability has pervasive effects on the acoustic realization of speech [1,2,3]. Generally, duration is shortened in more predictable contexts and conversely lengthened in less predictable contexts. There are several measures to quantify predictability in a message [4]. One of them is surprisal, which is calculated as $S(Unit_i) = -\log_2 P(Unit_i|Context)$. In a recent work, Ibrahim et al. [5] have found that the effect of syllable-based surprisal on the temporal dimension(s) of a syllable selectively extends to the segmental level, for example, consonant voicing in German. Closure duration was uniformly longer for both voiceless and voiced consonants, but voice onset time was not. The voice onset time pattern might be related to German being typically considered an 'aspirating' language, using [+spread glottis] for voiceless consonants and [-spread glottis] for their voiced counterparts [6,7]. However, voicing has also been reported in an intervocalic context for both voiceless and voiced consonants to varying extents. To further test whether the previously reported surprisal-based effect on voice onset time is driven by the phonological feature [spread glottis], the current study re-examined the downstream effect of syllable-based predictability on segmental voicing in German stops by measuring the degree of residual (phonetic) voicing during stop closure in an inter-sonorant context.

Method: Data were based on a subset of stimuli (speech produced in a quiet acoustic condition) from Ibrahim et al. [8]. 38 German speakers recorded 60 sentences. Each sentence contained a target stressed CV syllable in a polysyllabic word. Each target syllable began with one of the stops /p, k, b, d/, combined with one of the vowels /a:, e:, i:, o:, u:/. The analyzed data contained voiceless vs. voiced initial stops in a low or high surprisal syllable. Closure duration (CD) and voicing during closure (VDC) were extracted using in-house Python and Praat scripts. A ratio measure VDC/CD was used to factor out any potential covariation between VDC and CD. Linear mixed-effects modeling was used to evaluate the effect(s) of surprisal and target stop voicing status on VDC/CD ratio using the lmer package in R [9]. The final model was: VDC/CD ratio ~ Surprisal + Target stop voicing status + (1 | Speaker) + (1 | Syllable) + (1 | PrevManner) + (1 | Sentence).

Results: In an inter-sonorant context, we found a smaller VDC/CD ratio in voiceless stops than in voiced ones (p=2.04e-08***). As expected, residual voicing is shorter during a voiceless closure than during a voiced closure (Figure 1). This is consistent with the idea of preserving a phonological voicing distinction, as well as the physiological constraint of sustaining voicing for a long period during the closure of a voiceless stop. Moreover, the results yielded a significant effect of surprisal on VDC/CD ratio (p=.017*), with no interaction between the two factors (voicing and surprisal). The VDC/CD ratio is larger in a low than in a high surprisal syllable, irrespective of the voicing status of the target stops (Figure 1). That is, the syllable-based surprisal effect percolated down to German voicing, and the effect is uniform for a voiceless and voiced stop, when residual voicing was measured. Such a uniform effect on residual voicing is consistent with the previous result on closure duration. These findings reveal that the syllable-based surprisal effect can spread downstream to the segmental level and the effect is uniform for acoustic cues that are not directly tied to a phonological feature in German voicing (i.e. [spread glottis]).

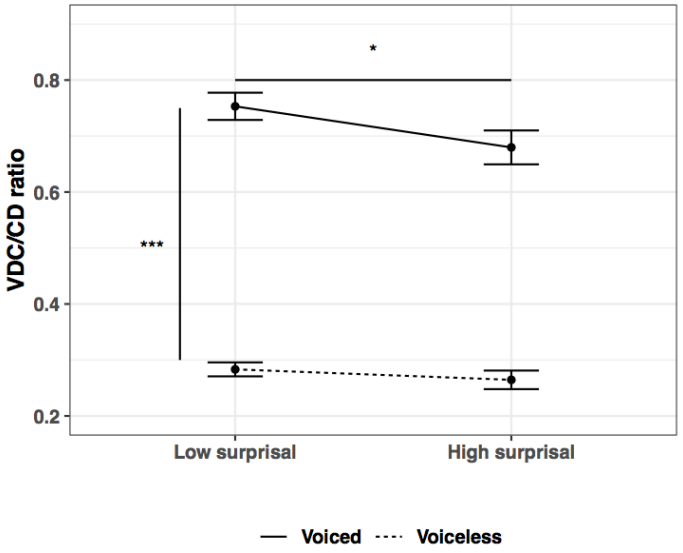


Figure 1: *Mean VDC/CD ratio as a function of surprisal and stop voicing status (with ±SE).*

References

[1] Aylett, M., and Turk, A. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech, *Language and Speech*, vol. 47, no. 1, pp. 31–56.

[2] Frank, A. F. and Jaeger, T. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production,in *Proceedings of the Annual Meeting of the Cognitive Science Society*.

[3] Crocker,M. W. , Demberg, V. and Teich, E. 2016. Information Density and Linguistic Encoding (IDeaL), KI - K¨unstliche Intelligenz ,vol. 30, no. 1, pp. 77–81.

[4] Hale, J. 2016. Information-theoretical complexity metrics. *Language and Linguistics Compass*, pp. 1–16.

[5] Ibrahim, O., Yuen, I., Andreeva, B., & Möbius, B. 2022. The effect of predictability on German stop voicing is phonologically selective. *Proc. Speech Prosody 2022*, 669-673.

[6] Jessen, M., Wiesbaden, B., and Ringen, C. 2002. Laryngeal features in German, *Phonology*, pp. 1–30.

[7] Beckman,J., Jessen, M., and Ringen,C. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages," *Journal of Linguistics*, vol. 49, no. 2, p. 259–284.

[8] Ibrahim, O., Yuen, I., van Os, M., Andreeva, B., & Möbius, B. 2022. The combined effects of contextual predictability and noise on the acoustic realisation of German syllables. *The Journal of the Acoustical Society of America*, 152(2), 911-920.

[9] Bates, D., Mächler, M. , Bolker, B., and Walker, S. , 2015, Fitting linear mixed-effects models using lme4, *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48.

Disfluency patterns in Italian parkinsonian speech of early-stage patients

Marta Maffia[1], Loredana Schettino[2], Rosa De Micco[3], Alessandro Tessitore[3]
[1]*Università degli Studi di Napoli "L'Orientale",* [2]*Università degli Studi di Napoli "Federico II",* [3]*Università degli Studi della Campania "Luigi Vanvitelli"*

Parkinson's Disease (PD) affects more than 2-3% of people aged over 65 in the world [1]. This neurodegenerative disorder consists of progressive deterioration and loss of dopaminergic neurons located in the substantia nigra pars compacta and basal ganglia. Hypokinetic dysarthria is one of the earliest symptoms of the disease and includes impairments both at the segmental and the suprasegmental level of speech such as imprecise articulation of vocalic and consonantal sounds and reduced vowel space area, narrow pitch variability and reduced tonal range, alteration of speech rate and rhythm [2-5]. In particular, parkinsonian speech has been described as "disfluent", however, the specific characteristics of disrupted PD speech have not been well documented [6-8]. Based on the evidence that phenomena like pauses, fillers, repetitions and self-repairs are commonly used in spontaneous speech for effectively managing and monitoring the own speech production, i.e. taking extra time for planning or editing something already uttered [9], this study aims at investigating the characteristics that distinguish disfluency phenomena patterns in Italian early-stage PD subjects' speech.

To reach the goal, 40 Italian native speakers were involved in the study, all residing in the Campania region (South of Italy): 20 participants with idiopathic non-demented PD (12 males, 8 females; aged 51–81, M=64) and 20 age-matched controls (8 males, 12 females; aged 54–77, M=65) were examined. The PD patients have been recruited from a longitudinal cohort of patients and they have all been diagnosed with PD in the last 2 years, with no history of previous language and speech disorders. These patients are prospectively followed and underwent extensive motor and non-motor clinical assessments every 12 months, with a clinical follow-up every 6 months. This study concerns the analysis of disfluency patterns in monologic speech (approximately a 1-2 minutes description of their neighborhood per speaker) using the ELAN software [10]. Disfluencies are defined as linguistic tools at speakers' disposal to monitor the online processes of speech planning, coding, articulation, and reception [11] by gaining extra time for planning or editing something. In particular, the objects of this analysis are Forward-Looking Disfluencies (FLDs), such as silent pauses, lexical and non-verbal fillers and prolongations that suspend the message delivery due to speech processing [12]. PD speech is compared with Health Control (HC) speech with reference to the following parameters: the frequency of disfluent items; the syntactic positioning of the items (within words, within phrases, between phrases, between clauses); the duration of silent pauses, filled pauses and prolongations. To test the statistical significance of the results, (generalized) linear mixed models are fitted [13], considering the health condition (PD or HC) as the independent variable and the speakers as the random effect.

The preliminary results from a pilot conducted considering 8 speakers (4 PD and 4 HC) show that FLDs are, actually, more frequent in HC speech (25 per minute) than in PD speech (17 per minute). In particular, the group of parkinsonian patients use more prolongations and fewer silent pauses than the HC group. As for the disfluent item positioning, in PD speech, a higher number of disfluent items is found to occur after word fragments, meaning that words are more often interrupted. Considering the duration of filled pauses, silent pauses and prolongations, on average, it is longer in PD than in HC speech (respectively, mean durations are: for filled pauses (FP), 506.6 ms vs. 333.2 ms; for silent pauses (SP), 616.5 ms vs. 241.8 ms; for prolongations (PRL), 366.6 ms vs. 265.3 ms). However, it is worth noticing that the durations of prolongations, filled and silent pauses show greater variability in PD (Figure 1).

These preliminary results highlight the relevance of investigating the specific uses, i.e., types and thereof characteristics, of disfluency phenomena rather than just considering their frequency of occurrence to gain insight into the features that characterize PD speech.

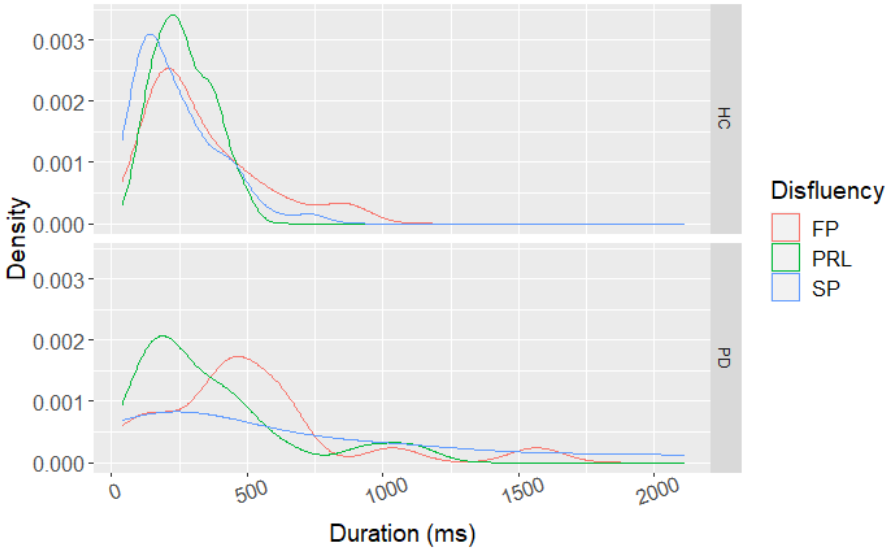


Figure 1. *Density of the duration values grouped by type of FLD.*

References

[1] de Lau, L. M. & Breteler, M. M. B. 2006. Epidemiology of Parkinson's disease. The Lancet. Neurology 5(6), 525-535.

[2] Darley, F. L., Aronson, A. E., & Brown, J. R. 1969. Clusters of deviant speech dimension in the dysarthrias. Journal of Speech and Hearing Research 12(3), 462-469.

[3] Goberman, A. M. & Coelho, C. A. 2005. Prosodic characteristics of Parkinsonian speech: The effect of levodopa-based medication. Journal of Medical Speech-Language Pathology 13, 51-68.

[4] Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M. & Caviness, J. N. 2009. Quantifying speech rhythm abnormalities in the dysarthrias. Journal of Speech, Language, and Hearing Research 52, 1334-1352.

[5] Skodda, S., Visser, W. & Schlegel, U. 2011. Vowel articulation in Parkinson's disease. Journal of voice: official journal of the Voice Foundation 25(4). 467-472.

[6] Goberman, A. M., Blomgren, M., & Metzger, E. 2010. Characteristics of speech disfluency in Parkinson disease. Journal of Neurolinguistics 23(5). 470-478.

[7] Juste, F. S., Sassi, F. C., Costa, J. B., & de Andrade, C. R. F. 2018. Frequency of speech disruptions in Parkinson's Disease and developmental stuttering: A comparison among speech tasks. Plos one 13(6). e0199054.

[8] Półrola, P. J., & Góral-Półrola, J. 2015. Speech disfluencies in Parkinson's disease. Medical Studies/Studia Medyczne 31(4). 267-270.

[9] Lickley, R. J. 2015. Fluency and Disfluency. In M. A. Redford (Ed.), The Handbook of Speech Production. Chichester: Wiley Online Library. 445-474.

[10] Sloetjes, H., & Wittenburg, P. 2008. Annotation by category-ELAN and ISO DCR. Proceeding of the 6th international Conference LREC 2008. 816-820.

[11] Levelt, W. J. 1989. Speaking: From intention to articulation. Cambridge, MIT Press.

[12] Schettino, L., Betz, S., Cutugno, F., Wagner, P. 2021. Hesitations and individual variability in Italian tourist guides' speech. In C. Bernardasci, D. Dipino, D. Garassino, S. Negrinelli, E. Pellegrino e S. Schmid (Eds.), Proceedings of the XVII Convegno Nazionale AISV. Studi AISV 8. 243-262.

[13] Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67(1). 1-4.

Acoustic duration and typing timing – same, same… but different?

Julia Muschalik[1], Dominic Schmitz[1], Akhilesh Kakolu Ramarao[1] and Dinah Baer-Henney[2]
[1]*Heinrich Heine University Düsseldorf, Department of English Language and Linguistics,*
[2]*Heinrich Heine University Düsseldorf, Department of Linguistics*

In recent years, evidence has been accumulated that both response latencies and within-word interkeystroke intervals (IKI), i.e., the time that elapses between the pressing of two keys, are influenced by lexical and sub-lexical variables. Comparable to acoustic duration is speech, IKIs appear to be susceptible to manipulations of, for example, word-, constituent-, bi- and trigram-frequencies (e.g., Baus et al. 2013; Bertram et al. 2015; Bonin et al. 2002; Sahel et al. 2008; Scaltritti et al. 2016), semantic transparency (e.g., Gagné & Spalding 2016; Libben & Weber 2014), prosodic boundaries (e.g., Fuchs & Krivokapic 2016), syllable structure (e.g., Nottbusch et al. 2005; Weingarten et al. 2007; Will et al. 2006), and morphological structure (e.g., Gagné & Spalding 2016; Will et al. 2006). In other words, IKIs appear not to be determined solely by random variation or by non-linguistic factors such as typing experience or location of keys on the keyboard. Instead, existing evidence suggests that typing as a peripheral process might be comparable to articulation in being a window into the processing architecture involved in language production and the interaction of central and peripheral production stages in general. Despite the obvious commonalities, however, research on durational differences in typing has remained largely independent of research on durational differences in pronunciation. This lack of direct comparison has left unanswered many questions regarding the similarities – and also differences – of the two language production modes.

This paper presents such a direct comparison. Our approach tests the generalizability of results from the articulatory domain to the domain of written language production with a well-researched phenomenon: word-final /s/ in English. Recent research has repeatedly demonstrated that word-final /s/ in English differs in duration depending on its morphological status (Zimmermann 2016; Plag et al. 2017; Plag et al. 2020; Schmitz et al. 2021; Tomaschek et al. 2019). In an extensive online typing study using the experimental design of Schmitz et al. (2021), we test their results for transferability to the written domain. Specifically, our study investigates whether language users type word-final /s/ in English pseudowords at different internal boundaries – non-morphemic, plural, auxiliary has-clitic and auxiliary is-clitic – with differing speeds and how our results compare to those found by Schmitz et al. (2021). For acoustic duration, the authors report that non-morphemic /s/ is longer than plural /s/, which in turn is longer than the auxiliary clitic /s/.

Analyzing our data with generalized additive mixed models (Wood 2017), we find that the influence of morphological structure on articulation and typing timing does not follow an identical principle. Participants in our experiment type non-morphemic /s/ and plural /s/ at almost identical speed. A significant difference emerges, however, for the typing of auxiliary clitics. Our results suggest that typing timing might be influenced by processing units other than morphemes. We discuss our results in relation to current theories of (written) language production.

References

[1] Baus, Cristina, Kristof Strijkers, and Albert Costa. 2013. "When Does Word Frequency Influence Written Production?" *Frontiers in Psychology* 4 (December): 963. https://doi.org/10.3389/fpsyg.2013.00963.

[2] Bertram, Raymond, Finn Tonnessen, Sven Strömqvist, Jukka Hyönä, and Pekka Niemi. 2015. "Cascaded Processing in Written Compound Word Production." *Frontiers in Human Neuroscience* 9 (April): 207. https://doi.org/10.3389/fnhum.2015.00207.

[3] Bonin, Patrick, Marylène Chalard, Alain Méot, and Michel Fayol. 2002. "The determinants of spoken and written picture naming latencies." *British Journal of Psychology* 93: 89.

[4] Fuchs, Susanne and Jelena Krivokapić. 2016. "Prosodic Boundaries in Writing: Evidence from a Keystroke Analysis." *Frontiers in Psychology*. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01678.

[5] Gagné, Christina L. and Thomas L Spalding. 2016. "Written Production of English Compounds: Effects of Morphology and Semantic Transparency." *Morphology* 26 (2): 133–55. https://doi.org/10.1007/s11525-015-9265-0.

[6] Libben, Gary and Silke Weber. 2014. "Semantic transparency, compounding, and the nature of independent variables." *Morphology and meaning* 327: 205.

[7] Nottbusch, Guido, Angela Grimm, Rüdiger Weingarten, and Udo Will. 2005. "Syllabic Sructures in Typing: Evidence from Deaf Writers." *Reading and Writing* 18 (6): 497–526.

[8] Plag, Ingo, Julia Homann, and Gero Kunter. 2017. "Homophony and morphology: The acoustics of word-final S in English." *Journal of Linguistics* 53. 181–216.

[9] Plag, Ingo, Arne Lohmann, Sonia Ben Hedia, and Julia Zimmermann. 2020. "An <s> is an <s'>, or is it? Plural and genitive-plural are not homophonous." In Lívia Körtvélyessy and Pavol Stekauer (eds.), Complex words: Advances in Morphology. Cambridge: CUP.

[10] Sahel, Said, Guido Nottbusch, Angela Grimm, and Rüdiger Weingarten. 2008. "Written production of German compounds: Effects of lexical frequency and semantic transparency". *Written Language & Literacy* 11(2): 211.

[11] Scaltritti, Michele, Barbara Arfé, Mark Torrance, and Francesca Peressotti. 2016. "Typing Pictures: Linguistic Processing Cascades into Finger Movements." *Cognition* 156: 16. https://doi.org/https://doi.org/10.1016/j.cognition.2016.07.006.

[12] Schmitz, Dominic, Dinah Baer-Henney, and Ingo Plag. 2021. "The duration of word-final /s/ differs across morphological categories in English: evidence from pseudowords." *Phonetica* 78.5-6 (2021): 571-616.

[13] Tomaschek, Fabian, Ingo Plag, R. Harald Baayen, and Mirjam Ernestus. 2019. "Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning." *Journal of Linguistics* 57. 1–39.

[14] Weingarten, Rüdiger, Guido Nottbusch, and Udo Will. 2007. "Morphemes, Syllables and Graphemes in Written Word Production." *Multidisciplinary Approaches to Language Production*, January. https://doi.org/10.1515/9783110894028.529.

[15] Will, Udo, Guido Nottbusch, and Rüdiger Weingarten. 2006. "Linguistic Units in Word Typing: Effects of Word Presentation Modes and Typing Delay." *Written Language & Literacy* 9 (1): 153–76. https://doi.org/10.1075/wll.9.1.10wil.

[16] Wood, Simon N. 2017. Generalized Additive Models: An Introduction with R, 2nd ed. Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

[17] Zimmermann, Julia. 2016. "Morphological status and acoustic realisation: Findings from NZE." In Christopher Carignanand & Michael D. Tyler (eds.), Proceedings of the sixteenth Australasian international conference on speech science and technology, 201–204. Parramatta.

Pyrlato, a novel methodology to collect real-world acoustic data

Giuseppe Magistro[1] and Claudia Crocco[1]
Ghent University

Background The use of real-world speech in prosodic research is necessary to capture the complexity and variability of natural interactions. In fact, widely used elicitation techniques such as games and tasks may fail to represent the full range of speech variation [1]. We aim at contributing to the study of prosody "in the wild" by presenting a tool, Pyrlato, which builds a corpus of non-controlled data extracted from YouTube.

Methodology We developed Pyrlato in Python. Pyrlato has the following pipeline: first, the user selects the language to be surveyed, among the ones where YouTube can provide automatic captions. At this stage, the user can input specific keywords to narrow the research of videos: specific speakers, channels, genres, and so forth. Pyrlato will generate a Python object for each video found in the research, which contains a method to obtain the automatically generated subtitles. The user can then specify which string to look for in the subtitles: specific words, constructions, or patterns matched with regular expressions. If the program finds correspondence in the subtitles, it will extract the audio portion containing the desired item observing a time window customizable by the user (we set a span of 7 seconds for our case study and adjusted manually when necessary). When the execution is completed, all the relevant extracts are available in the working directory in the format selected by the user. The entire code will be soon available on GitHub. To showcase the usefulness of Pyrlato, we used it to corroborate hypotheses elaborated in previous studies. We do so by exploring a corpus of real-world speech created in a fast and convenient way.

Case study Italian employs an optional reinforcer of negation, *mica,* that corrects and denies a contextually activated proposition [2], (ex. 1). While a pragmatic and syntactic description of Italian *mica* was conducted thoroughly in many works [2,3,4], a description of its prosodic realization is missing, except for [3] that cursorily hinted at a prominence found on this corrective particle. In addition, [5] described *miga* in the Venetian dialect as a metrically strong element, bearing a rising pitch accent. Pyrlato was used to explore the prosodic realization of *mica* in Italian, in particular to verify if the claim in [3] about the presence of a prominence was verified. By using Pyrlato, we scraped instances of *mica* in unscripted TV interviews and talk shows (to ensure better recording quality) on YouTube. A total amount of 43 hours was scraped by the software, finding 434 examples of Italian *mica*. After sanity-checking the data and removing hesitations and instances that could not be analyzed for different reasons, we obtained a dataset of 290 usable instances, which were inspected using Praat [6]. Confirming the previous impressionistic observations [3], *mica* is stressed and bears the most prominent pitch-accent of the contour, in the shape of a rising tone (fig.1). However, the real-world data obtained with Pyrlato also highlighted the presence of variation in the prosodic realization of *mica.* Indeed, 7% of the corpus show cases where *mica* is not prominent and does not display a pitch-accent (fig.2). While this minority may be interpreted as the result of phonetic undershooting, the contexts where these latter instances of *mica* appeared did not evoke any explicit correction of a previous turn, but rather seemed to encode the denial of an expectation or implicature (ex.2). The real-world data scraped with Pyrlato, therefore, indicate that the prosodic realization of *mica* may vary "in the wild". These data call for further investigations to identify the source of the observed variation. In conclusion, Pyrlato can be used as a faster manner to explore the variability of "wild data" (in our case, the pitch accent marking possibly in relation to pragmatic differentiation), so as to spot possible dimensions of variations for more specific and targeted studies.

1. A: *Non andarci se è pericoloso!* ‘Don’t go there if it is dangerous!’
   B: *Non è mica pericoloso.* ‘It is not dangerous at all’

2. *Ora parliamo dell’uomo della luce [elettrica], che non è mica un fisico.*
   Now let’s talk about the power man, who is not a physicist.

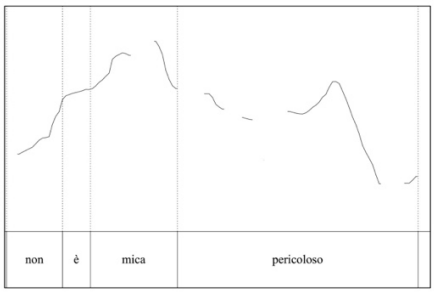


Figure 1. *Statement with mica in an explicit correction*

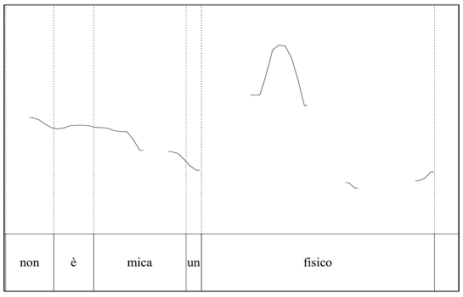


Figure 2. *Statement with mica in the correction of an implicature*

References

[1] del Mar Vanrell, M., Feldhausen, I., & Astruc, L. (2018). The Discourse Completion Task in Romance prosody research: Status quo and outlook. In I. Feldhausen, J. Fliessbach, M. del Mar Vanrell (Eds.) Methods in prosody: a Romance language perspective. Language Science Press.

[2] Cinque, G. (1976). Mica. In *Annali della Facoltà di Lettere e Filosofia dell’Università di Padova*, 101–12.

[3] Frana, I., & Rawlins, K. (2019). Attitudes in discourse: Italian polar questions and the particle mica. Semantics and Pragmatics, 12, 16.

[4] Magistro, G. (2022). Mica preposing as focus fronting. Glossa- A Journal of General Linguistics, 7(1).

[5] Magistro, G., Crocco, C., & Breitbarth, A. (2022). Information structure and Jespersen’s cycle : the dialects of Veneto as a window on processes of language change. In N. Catasso, M. Coniglio, & C. De Bastiani (Eds.), Language change at the interfaces : intrasentential and intersentential phenomena (Vol. 275, pp. 35–59). John Benjamins.

[6] Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer.

Temporal and/or prosodic interaction between head and eyebrow peaks in statements and yes-no questions?

Marisa Cruz and Sónia Frota
University of Lisbon

It is well-known that gestures and speech interact in a synchronic way [1], although not necessarily in strict simultaneity [e.g., 2], being, instead, constrained by gesture type [3] and function [4]. We examined whether the interaction between gestures and speech is temporally and/or prosodically driven, by exploring the kinematics of head and eyebrow movements in statements and yes-no questions in European Portuguese (EP). Specifically, we analysed: (i) the temporal alignment of head and eyebrow peaks relative to the nuclear prosodic word (PW); (ii) prominence (IP head, non-head; PW head, non-head); and (iii) prosodic position in the IP, PW (initial, medial, and final) and syllable (onset, nucleus, and coda) of the spoken material aligned with head and eyebrow peaks. In addition, peak amplitude was inspected to determine whether it differed across sentence types and depending on the prosody.

Audiovisual data from 3 female native speakers of Standard EP was used. Speakers were videotaped within the InAPoP project [5], while performing a Discourse Completion Task [6] adapted for EP. For the analysis, 11 neutral statements and 34 neutral yes-no questions were selected, which respectively exhibit the falling H+L* L% and falling-rising H+L* LH% nuclear contours [7]), as well as different visual cues: head falling movement in statements *versus* head falling and eyebrow raising movements in yes-no questions [8]. We also inspected a set of less frequent yes-no questions produced by the same speakers: 10 falling-rising yes-no questions involving head movement only (without eyebrow raising); 7 yes-no questions with both visual cues but a different melodic pattern (H*+L L%, hereafter *yes-no other*). The kinematic analysis was performed using *Kinovea* [9], by tracking the vertical displacement of the head and eyebrows in pixels (px), along the time series (ms).

We found that head (N=48) and eyebrow (N=50) peaks are motorically coordinated with intonation, and with each other, on a regular (but not concurrent) timing distribution: head peaks mainly co-occur with the nuclear PW, regardless of sentence type (Fig.1), whereas eyebrow peaks, although exhibiting a varying time lapse (av. 222.8ms-648ms), precede the head peaks and mainly occur before the nuclear word (Fig.2). This supports Loehr's (2012) suggestion that *synchrony* cannot be interpreted as strict co-occurrence (not even between concurrent gestures). Prosodic alignment is more stable than time-alignment: head peaks mostly align with the IP head, across sentence types, whereas eyebrow peaks mainly align with the IP head in statements, but with the head of a non-nuclear PW in yes-no questions (Fig.3). GLMM analyses on peak amplitude revealed: for head peak amplitude, a significant effect of position in the IP ($F(2, 28) = 5.20$, $p = .012$), prosodic prominence ($F(2, 28) = 7.23$, $p = .003$), and the interaction prominence *sentence type ($F(6, 28) = 8.52$, $p = .000$), with higher peaks on IP initial position, and on non-nuclear PW heads in statements, although peaks mainly align with IP heads (i.e., IP final position); for eyebrow peak amplitude, a significant effect of the position in the PW ($F(3, 32) = 32.92$, $p = .000$), prominence ($F(2, 32) = 8.22$, $p = .001$), sentence type ($F(3, 32) = 32.87$, $p = .000$), and the interaction prominence*sentence type ($F(5, 32) = 15.67$, $p = .000$), with higher peaks on PW medial position (usually a PW head) and on non-heads, and lower/no peaks on yes-no questions without eyebrow raising, although peaks align to both prominent positions and edges.

In sum, the prosodic loci and the amplitude of head and eyebrow peaks display a synergy whereby the first cues prominence and the latter prosodic edges for head peaks, while eyebrow peak position and amplitude may cue either (Fig.4). This points to a complementary role between head and eyebrow, and between gestures and intonation, and suggests that the interaction between gestures and speech is prosodically (rather than temporally) driven.

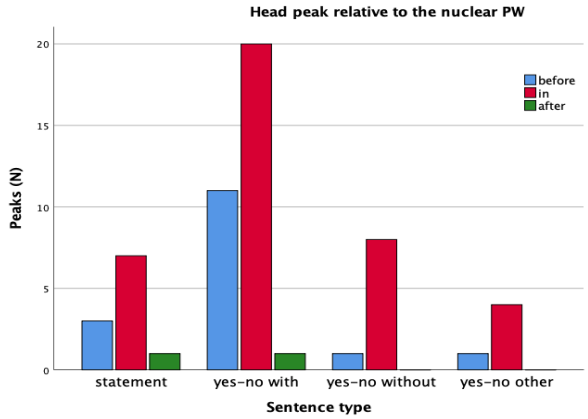


Figure 1. *Temporal alignment of the head peak relative to the nuclear PW.*

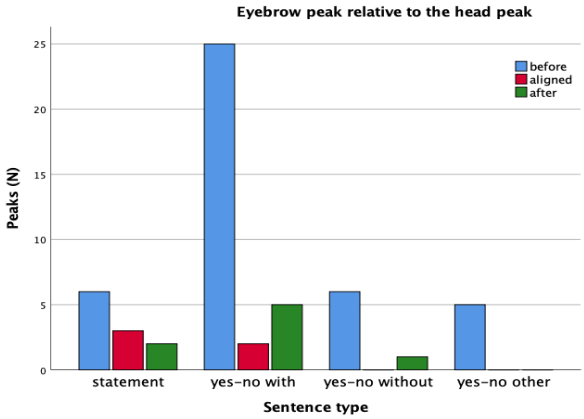


Figure 2. *Temporal alignment of the eyebrow peak relative to the head peak.*

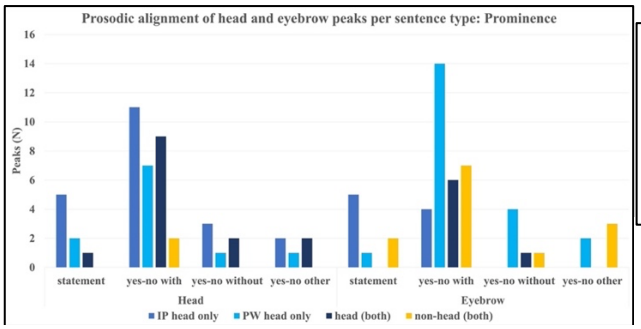


Figure 3. *Prosodic alignment of head and eyebrow peaks per sentence type: Prominence considering the IP and PW prosodic levels.*

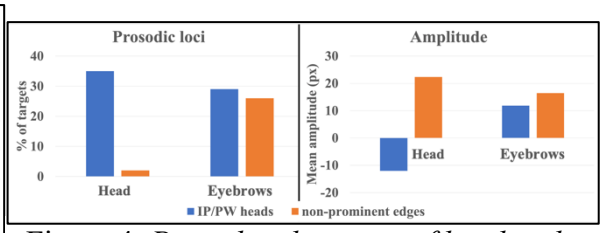


Figure 4. *Prosodic alignment of head and eyebrow peaks: Loci (left panel) and amplitude (right panel) considering IP and PW heads (prosodic prominence) versus prosodic edges (not coinciding with heads) .*

References

[1] McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

[2] Loehr, D. P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab. Phonol*. Special Issue Gesture as Language, Gesture and Language 2, 3(1), 71–89. doi: 10.1515/lp-2012-0006

[3] Leonard, T., Cummins, F. 2010. The temporal relation between beat gestures and speech. *Lang. Cognitive Proc.* 26(10), 1457–1471. doi: 10.1080/01690965.2010.500218

[4] Wagner, P., Malisz, Z., Kopp, S. 2014. Gesture and speech in interaction: An overview. *Speech Commun*. 57, 209–232. doi: 10.1016/j.specom.2013.09.008

[5] Frota, S. (Coord.) (2012-2015). *Interactive Atlas of the Prosody of Portuguese* Project (PTDC/CLE-LIN/119787/2010), funded by FCT – Fundação para a Ciência e a Tecnologia, Portugal.

[6] Félix-Brasdefer, J. C. 2009. Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In: Juán, E. U., Martinéz-Flor, A. (eds), *Speech act performance: Theoretical, Empirical, and Methodological Issues*. John Benjamins Publishing, 41–56.

[7] Frota, S. 2002. Nuclear falls and rises in European Portuguese: A phonological analysis of declarative and question intonation. *Probus* 14(1) (Special Issue on Intonation in Romance, ed. by Hualde, J. I.), 113–146. doi: 10.1515/prbs.2002.001

[8] Cruz, M., Swerts, M., Frota, S. 2015. Variation in tone and gesture within language. In: The Scottish Consortium for ICPhS 2015 (ed), *Proc. of the 18th International Congress of Phonetic Sciences*, University of Glasgow.

[9] Kinovea 0.9.4. https://www.kinovea.org

Lexical stress in language contact: the interplay of acoustic correlates

Natália Brambatti Guzzo[1], Guilherme D. Garcia[2]
[1]*Saint Mary's University,* [2]*Université Laval*

We examine the acoustic manifestation of stress in Portuguese-Veneto contact in Brazil. We propose that the representation of stress involves not only metrical features, but also the specification of acoustic cues, and that these acoustic specifications may be transferred in a bilingual situation. In both Brazilian Portuguese (BP) and Brazilian Veneto (BV; Romance), stress is mostly penultimate, but it is typically realized in the final syllable if this syllable is heavy (i.e., ends in a coda or diphthong; Garcia, 2017; Guzzo, 2022; Mateus & d'Andrade, 2000). However, previous investigations indicate that the acoustic manifestation of stress and stress-related properties in the two languages is somewhat different. While stress is cued with duration in both languages, they differ regarding vowel reduction. Unstressed final vowels in BP exhibit substantial vowel reduction and often undergo devoicing and deletion (Massini-Cagliari, 1992; Walker & Mendes, 2019), whereas reduction of unstressed vowels in BV is marginal (Guzzo, 2022).

Assuming (a) that the representation of stress includes the acoustic cues employed in its manifestation, and (b) that stress is cued differently in BP and BV to some extent, two possibilities for BP-BV bilinguals' productions arise: (i) they produce stress differently in the two languages, mirroring BP monolinguals in their BP productions, or (ii) they exhibit an overlap of stress cues in the two languages. We hypothesize that possibility (ii) applies to the BP-BV contact situation: their representations for BP and BV stress interact, similar to what has been observed with other phonological phenomena in contact (Newlin-Lukowicz, 2014; Sundara et al., 2006). To test this, we conducted a production experiment where participants named figures using carrier sentences. Participants were BP-BV bilinguals (n = 21) and BP monolinguals (n = 9); no BV monolinguals were included since virtually all speakers of BV in Brazil also speak BP. Bilingual participants completed two versions of the experiment, one in each language. In both languages, the target vowel was /a/, which was found in final, penultimate or antepenultimate position in three-syllable nouns with penultimate stress. Participants were instructed to place the target words in either phrase-medial or phrase-final position in the carrier sentences.

Duration, F1, F2 and f0 (from three points) of all the target vowels (n = 2,602) were extracted using Praat scripts (Boersma & Weenink, 2022). We found systematic differences in duration, f0 and F1. The data were analysed with three mixed-effects linear models (one per correlate). Models included by-speaker and by-item random intercepts (all coefficients below have $P < 0.01$ given their t values). No differences were found for position of the target word in the carrier sentence. Bilinguals produced stressed vowels with similar duration in BP and BV, but significantly longer than those produced by monolinguals (see Figure 1 (left); $\hat{\beta}$(Bil-BV)=37.84, t=6.15; relative to Mon-BP in syllable 2). Word-finally, vowels were significantly longer in BV productions. For f0, BV vowels had substantially higher values word-finally relative to both bilingual and monolingual (see Figure 2; $\hat{\beta}$=61.55, t=4.98) BP productions. Finally, F1 in word-final vowels was significantly lower in monolingual than bilingual productions (see Figure 1 (right); BV: $\hat{\beta}$=56.05, t=3.95; BP: $\hat{\beta}$=63.07, t=3.95), which patterned together, indicating more reduction in monolingual BP.

These results suggest an overlap in some of the cues (i.e., duration, F1) used to signal stress in bilingual BP and BV, as per our hypothesis. At the same time, f0 is used differently in bilingual BP and BV, indicating that their acoustic specifications are not identical and that, although interdependent, the two systems are not merged. In summary, contact promotes cue interaction in the manifestation of stress, suggesting that such cues must be representationally encoded.

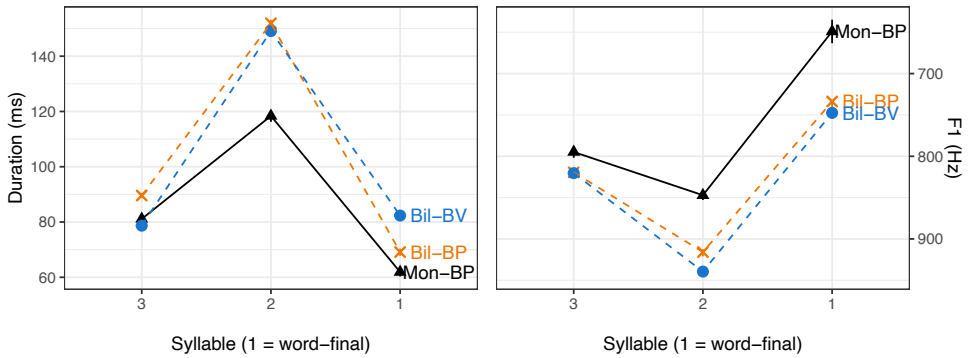


Figure 1: Results for duration (left) and F1 (right) in bilingual BV and BP, and monolingual BP.

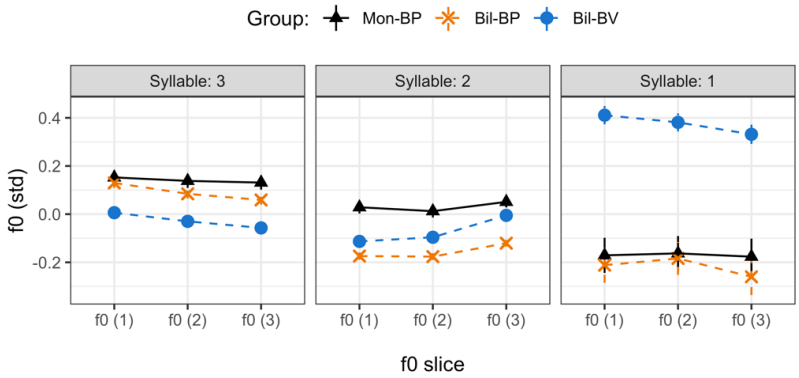


Figure 2: Results for f0 in bilingual BV and BP, and monolingual BP.

References

Boersma, Paul & David Weenink. 2022. *Praat: Doing phonetics by computer* [computer program]. Version 6.2.14.

Garcia, G. D. 2017. Weight gradience and stress in Portuguese. *Phonology*, *34*(1), 41–79. https://doi.org/10.1017/S0952675717000033

Guzzo, Natália Brambatti. 2022. Brazilian Veneto (Talian). *Journal of the International Phonetic Association – Illustrations of the IPA*. Available online.

Massini-Cagliari, Gladis. 1992. *Acento e ritmo* [Stress and rhythm]. São Paulo: Contexto.

Mateus, Maria Helena & Ernesto d'Andrade. 2000. *The phonology of Portuguese*. Oxford: Oxford University Press.

Newlin-Lukowicz, Luiza. 2014. From interference to transfer in language contact: Variation in voice onset time. *Language Variation and Change* 26: 359-385.

Sundara, Megha, Linda Polka & Shri Baum. 2006. Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition* 9: 97-114.

Walker, James A. & Ronald Beline Mendes. 2019. Lower your voice: Vowel devoicing and deletion in Brazilian Portuguese. *Proceedings of the ICPhS 2019*.

Uncover articulatory correlates of acoustic duration with analysis-by-synthesis: the case of diphthongs

Eoin O'Reilly[1], Christopher Geissler[1] and Kevin Tang[1,2])
[1]*Heinrich Heine University Düsseldorf,* [2]*University of Florida*

Acoustic reduction is widely attested in speech, but how this takes place in articulation is not as well known. The aim of this study is to examine how reduction takes place in articulation, taking the example of the English PRICE /a͡ɪ/ diphthong. We identify four articulatory mechanisms that could potentially result in similar reductions in acoustic duration: increased gestural **overlap**, **undershoot**, **shortening** of gestures, and increase in **stiffness** (resulting in faster movement).

Previous research has found evidence suggesting the nature of temporal coordination in diphthongs, but has not directly tested the predictions of such coordination patterns using simulation. Acoustic reduction of Spanish diphthongs across task conditions was studied by [1], who constructed a continuum of reduction as hiatus→diphthong→monophthong, but the articulatory manifestation of this process is not clear. Differences in articulation of the /a͡ɪ/ diphthong (in English and German) have been studied in terms of the Euclidian distance travelled between the targets [2][3]. Gestural timing has been identified as a key difference between diphthongs and hiatus sequences in Romanian [4][5]. Changes in the timing of gestures play an important role in diachronic sound change, as has been shown in Romance [6] and English, including the PRICE vowel which is the focus of this study [7]. In sum, previous work suggests that diphthong reduction could involve changes in the timing of gestures as well as their targets.

The simulation procedure used the Task Dynamics Application (TADA) [8], with an analysis-by-synthesis approach similar to [9]. Unlike [9], the present study focuses solely on articulation rather than matching acoustics, and adds undershoot and changes to stiffness. After initially generating a gestural score based on the Coupled Oscillator Model of Syllable Structure [10], TADA uses this gestural score as input and simulates the trajectories of the vocal tract organs. We successively modify gestural score files according to a specified set of reduction rules. The "reduced" versions of these utterances produced by our script are then used as input for the TADA trajectory simulation algorithm.

This procedure was used to generate 65,536 variations (combinations of 16 parameters each with two values) of the English word *five*, which were compared with 425 examples from 48 speakers in the X-Ray Microbeam Database [11]. Analysis was performed using Dynamic Time Warping (DTW), [12], as implemented by the DTAIDistance Python package [13], with an additional penalty for durations that differ substantially from the XRMB data. Examples of good and bad fits between simulated and actual gestures are shown in Figures 1 and 2. To understand the reduction strategies, regression analyses were used to predict acoustic duration with the best fit articulatory parameters, controlling for speaker, text, task type and giveness.

Results show that the best-fit simulations tended to use gestural shortening and overlap of both [a] and [ɪ] components of the diphthong, but all four strategies were observed. However, a correlation analysis between each dimension of reduction and acoustic duration showed the strongest correlation with gestural shortening. Overlap resulting from earlier phasing of [ɪ] was also correlated with duration, but was even more strongly correlated with shortening of [ɪ]. We interpret this as evidence that gestural shortening is the most important articulatory correlate of overall acoustic duration, but that the other forms of reduction contribute to the shape of an articulatory trajectory, and likely to other aspects of acoustic detail.

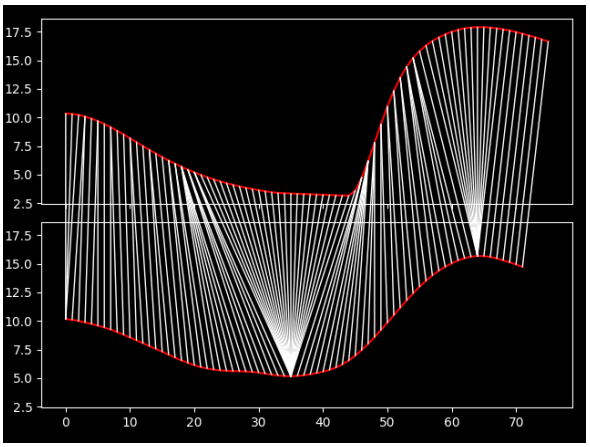


Figure 1. *A good match between real and simulated tongue dorsum trajectories* (*the lower half is the real utterance).*

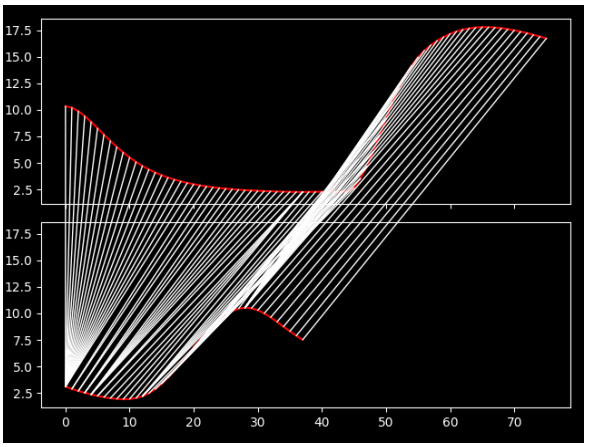


Figure 2. *A poor match between real and simulated trajectories.*

References

[1] Aguilar, L. 1999. Hiatus and diphthong: Acoustic cues and speech situation dierences. *Speech Communication* 18.

[2] Simpson, A. 2002. Gender-specific articulatory–acoustic relations in vowel sequences. *Journal of Phonetics* 30(3). 417–435.

[3] Weirich, M. & Simpson, A. 2018. Individual differences in acoustic and articulatory undershoot in a German diphthong – Variation between male and female speakers. *Journal of Phonetics* 71. 35–50.

[4] Marin, S. & Goldstein, L. 2012. A gestural model of the temporal organization of vowel clusters in Romanian. In Philip Hoole, Lasse Bombien, Marianne Pouplier, Christine Mooshammer & Barbara Kühnert (eds.), *Consonant Clusters and Structural Complexity*, 177–204. De Gruyter.

[5] Marin, S. 2014. Romanian diphthongs /ea/ y /oa/: an articulatory comparison with /ja/-/wa/ and with hiatus sequences. *Revista de Filología Románica* 31(1). 83–97.

[6] Chitoran, I. & Hualde, J.I. 2007. From hiatus to diphthong: the evolution of vowel sequences in Romance. *Phonology*. Cambridge University Press 24(1). 37–75.

[7] Sóskuthy, M., Hay, J. & Brand, J. 2019. Horizontal diphthong shift in New Zealand English. In *Proceedings of the 19th International Congress of Phonetic Sciences*.

[8] Nam, H., Goldstein, L., Saltzman, E. & Byrd, D. 2004. TADA: An enhanced, portable Task Dynamics model in MATLAB. *The Journal of the Acoustical Society of America*. Acoustical Society of America 115(5). 2430–2430.

[9] Nam, H., Mitra, V., Tiede, M., Saltzman, E., Goldstein, L., Espy-Wilson, C. & Hasegawa-Johnson, M.. 2010. A Procedure for Estimating Gestural Scores from Natural Speech. *Interspeech 2010*.

[10] Nam, H. & Saltzman, E. 2003. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of the Phonetic Sciences*.

[11] Westbury, J.R., Turner, G. & Dembowski, J. 1994. X-ray microbeam speech production database user's handbook. *University of Wisconsin*.

[12] Sakoe, H. & Chiba, S. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*. IEEE 26(1). 43–49.

[13] Meert, W., Hendrickx, K., Van Craenendonck, T, Robberechts, P. 2020. wannesm/dtaidistance v2.0.0. Zenodo.

Learning words while playing with virtual peers in multiple accents: school-aged children benefit from experience with variability

Adriana Hanulíková and Helena Levy
University of Freiburg

We present a novel paradigm to examine the effect of language exposure and variable input on the acquisition of words in primary school–aged children. The bulk of research on word learning has focused on infants and young children, whereas research regarding vocabulary acquisition in heterogeneous groups at primary school age is much more limited. Children growing up with different languages and foreign or regional accents might benefit from their experience with variability when learning new words from peers with unfamiliar accents. The aim of the current study was therefore to examine the extent to which language and accent experience predicts monolingual and bilingual children's success at word learning through play in the context of accent variability. Experience here refers to exposure to different languages and foreign/regional accents, measured as the number of hours per week spent with each language or accent according to a parental and teacher questionnaire.

German-speaking children (aged 7–11 years, 45 girls, 43 boys, 43 monolinguals, and 45 bilinguals with various language backgrounds and with age of acquisition of German 0 to 7 years) played a computerized card game with virtual peers that resembles natural advanced lexical acquisition, during which new words are learned from child speakers and are produced actively in peer-group interactions. During the game, children learned six words, technical names of objects usually unknown to children this age (e.g. *Amboss* "anvil"; *Hippe* "pruning knife"), from six peers speaking a familiar accent (Standard German) or unfamiliar foreign (Hebrew-accented German) and regional (Swiss German) accents. Children's task was to find two identical objects on two cards and to name the object. Children first heard the names of the unfamiliar objects that were introduced by virtual peers in different accents and were then able to actively produce these words in simulated interactions with the virtual peers. Children's short-term and working memory was assessed via both nonword repetition and digit span tasks (forward and backward), and we used a standardized vocabulary test (SET 5-10) to assess German vocabulary size.

We used logistic mixed-effects regression models to analyse the data. In line with the hypothesis, successful learning was predicted by the amount of input in regional and foreign accents but not by exposure to other languages than German (i.e. bilingualism). This suggests that experience with variable speech trumps bilingualism when it comes to coping with heterogeneous input. Differences between previous studies that found effects of bilingualism on word learning [1, 2] and our study could be related to exposure being assessed on a continuum rather than as a binary (yes/no) variable. We also found that increased working memory capacity and age (older better than younger) predicted better performance at the task.

Future studies could examine whether similarities between a familiar accent and the test accent can explain the benefits of accent experience for learning from unfamiliar accents. The children in our sample had experience with diverse accents, and it was impossible to determine whether children with better performance were growing up with accents that were perceptually closer to the test accents. Alternatively, the positive effect may be due to increased perceptual flexibility through experience with variability in general [3], that is, due to a better ability to process accented speech that extends beyond similarities with familiar accents [4]. We will discuss how accent experience affects word learning under variable input conditions and how this paradigm can be extended to further test situations. Given that peer groups gain importance during the primary school years, and interactions between children are often characterized by linguistic heterogeneity, the paradigm offers a realistic learning context.

References

[1] Kaushanskaya, M., Gross, M., & Buac, M. 2014. Effects of classroom bilingualism on task-shifting, verbal memory, and word learning in children. *Developmental Science* 17, 564–583.

[2] Menjivar, J., & Akhtar, N. 2017. Language experience and preschoolers' foreign word learning. *Bilingualism: Language and Cognition* 20, 642–648.

[3] Levy, H., Konieczny, L., & Hanulíková, A. 2019. Processing of unfamiliar accents in monolingual and bilingual children: Effects of type and amount of accent experience. *Journal of Child Language* 46, 368–392.

[4] Kaushanskaya, M., & Marian, V. 2009. The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, 16, 705–710.

The prosodic realization of Spanish focus in nuclear position by proficient Chinese learners of Spanish: The case of broad, corrective and contrastive foci

Peng Li[1] and Xiaotong Xi[2]

[1] *Center for Multilingualism in Society across the Lifespan, University of Oslo, Norway*

[2] Universitat Pompeu Fabra, Spain

Given the dynamic nature and abstractness of prosody, crosslinguistic influence (CLI) in speech prosody is less investigated than in speech sounds. According to the prediction of L2 Intonation Learning theory (LILt) [1], prosody in learners' first language (L1) may predict the learning outcome in L2. Therefore, the prosodic acquisition of L2 intonation language by L1 tone language speakers will offer an ideal window for observing CLI on the prosodic level. For instance, Chinese learners showed more pitch variants than Spanish natives on lexical stress [2] and translated L1 tonal patterns to Spanish stress [3] and intonation [4]. However, few studies have investigated Chinese students' prosodic realization of narrow focus in Spanish and the role of lexical stress position on the contours of focus. This study thus aims at filling the research gap. Based on the literature reviewed, we hypothesized that Chinese students manipulated pitch contours to a larger degree than Spanish natives to contrast broad and narrow focus and that the lexical stress of the focused word would affect the contour shapes.

Sixteen Chinese speakers (female = 12) with proficient level of Spanish and nine native Spanish speakers (female = 6) participated in a discourse completion task, which elicited nine sentences varied in focus type (broad, corrective and contrastive) and lexical stress pattern (initial-stressed "**vi**no", medial-stressed "Ma**ri**na", and final-stressed "Mi**lán**"). We extracted 10 regularly spaced pitch points from each syllable and generated a time-normalized pitch contour using *z*-scored F0. To compare the pitch contours, we built three Generalized Additive Mixed Models for the three target words. The fixed effects included time (normalized pitch points), gender (m vs. f), focus (broad vs. corrective vs. contrastive), group (Chinese students vs. Spanish natives), and a two-way interaction of Focus × Group. The smooth terms included a smooth curve for the Focus × Group interaction and a by participant random smooth. The post-hoc comparisons by group across focus types are plotted in figures 1-3. We found that Spanish natives showed similar pitch contours for broad and contrastive foci (L* L%), which contrasted with corrective focus (L+H* HL%) regardless of stress pattern. By contrast, Chinese students showed similar contour patterns for narrow focus (both corrective and contrastive: L+¡H* HL%) but varied broad focus contours by stress pattern (L+H* HL% for medial vs. H* L% for initial and final stress patterns).

The results have the following implications. First, although the pitch contours of corrective focus were similar between Chinese students and Spanish natives, Chinese students showed a significantly higher pitch peak than Spanish natives. Similar patterns appeared in the broad focus contours, with Chinese students showing higher pitch accent than Spanish natives. These results confirmed our hypothesis that Chinese students manipulated pitch contours to a larger extent than Spanish natives to mark Spanish focus. Second, lexical stress significantly affected the nuclear contour shape of Chinese students' Spanish speech. In broad focus, the medial stress position showed a rising pitch accent (L+H*), which was inconsistent with other positions (H*). It is therefore in line with previous research that showed that Chinese students tended to use high or rising tones to replace Spanish lexical stress [3]. Third, different from Spanish natives, Chinese students used similar prosodic strategies to mark Spanish narrow focus (corrective and contrastive). This suggests that even proficient learners still show some nonnative prosodic patterns.

All in all, this study added more evidence for CLI in speech prosody to support the predictions of LILt by showing that Chinese students tend to produce high or rising tones in L2 prosody. L2 teachers may thus pay more attention to the prosodic realizations of focus in Spanish pronunciation instruction.

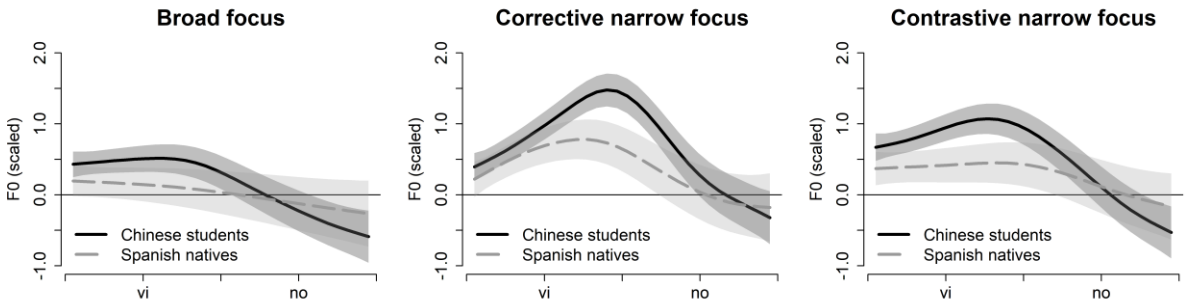


Figure 1. *Pitch contours of initial-stressed "vino" plotted by group across focus condition*

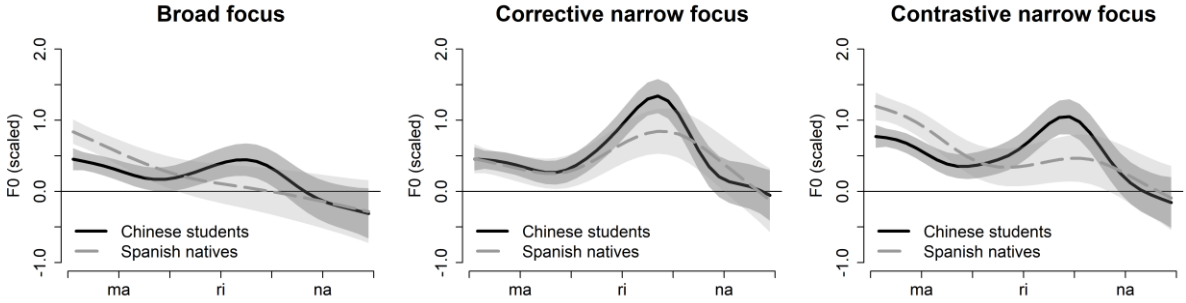


Figure 2. *Pitch contours of medial-stressed "Marina" plotted by group across focus condition*

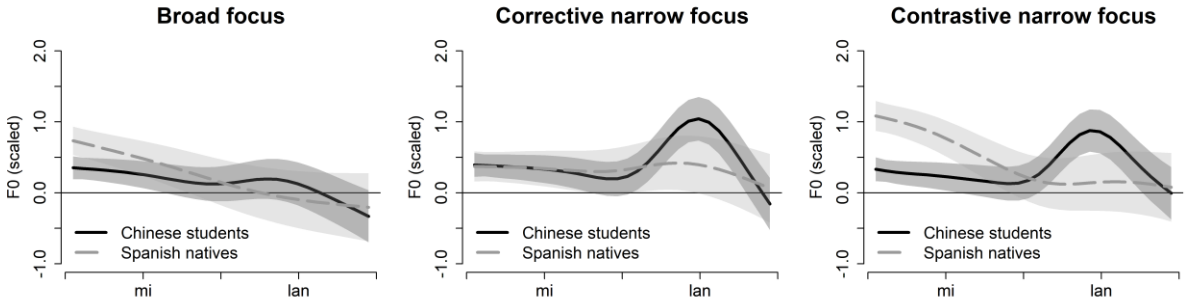


Figure 3. *Pitch contours of final-stressed "Milán" plotted by group across focus condition*

References

[1] 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). Springer. https://doi.org/10.1007/978-3-662-45168-7_9

[2] Li, P., & Xi, X. (2022). Spanish lexical stress produced by proficient Mandarin learners of Spanish. *Proceedings of the 4th International Symposium on Applied Phonetics*.

[3] Chen, Y. (2007). From tone to accent: The tonal transfer strategy for Chinese L2 learners of Spanish. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Science* (pp. 1645–1648). Saarbrücken Univ. des Saarlandes.

[4] Shang, P., & Elvira-García, W. (2022). Second language acquisition of Spanish prosody by Chinese speakers: Nuclear contours and pitch characteristics. *Vigo International Journal of Applied Linguistics*, 19, 129–176. https://doi.org/10.35869/vial.v0i19.3762

The Effectivity of Explicit Pronunciation Training in Dutch EFL Learners' Phoneme Perception and Production in English

Jasmijn Stolvoort[1], Margreet Pieper[1], Aoju Chen[1]
[1]*Utrecht University*

Previous studies have suggested that the word-final /t-d/ pair in English tends to pose persistent difficulty for Dutch learners as this word-final contrast is absent in Dutch [e.g., 2, 4, 9]. Conflicting findings have been reported about the effectiveness of pronunciation training for phoneme production and perception of Dutch EFL learners, even in the case of the same learner groups (e.g., pupils of senior general secondary education or HAVO) [e.g., 3, 7]. Further, some studies suggest that language proficiency positively affects learning outcomes [e.g., 5], while others have found no such an effect [e.g., 10]. Therefore, the current study aims to shed new light on whether (short-term) explicit pronunciation training improves English sound production and perception in Dutch EFL learners, and whether its effectiveness is influenced by their English proficiency.

To this end, adopting the method used in [7], we conducted a training study focusing on the word-final /t-d/ pair in English. It involved an experimental group (n = 49) and a control group (n = 24), consisting of HAVO 4th-graders (mean age: 16.1 years; SD = 0.8). They took the LexTALE test [6] to determine their level of English prior to the experiment. Their perception and production were examined in a pre- and post-test before and after the intervention (experimental group) or unrelated regular lessons (control group). The perception test was conducted in Qualtrics and contained 12 randomized recordings of nonwords produced by a native speaker of English. The participants indicated for each nonword whether they heard word-final /t/ or /d/. The production test consisted of slides that presented 12 nonwords containing word-final /t/ or /d/ in IPA with pronunciation tips (i.e., real words that contain the same sounds). The participants recorded their productions. The intervention consisted of two lessons that explicated phonetics in general and the role of vowel duration in the articulation of word-final /t/ and /d/ with examples and pronunciation exercises.

The participants' LexTALE scores (43.75 ~ 86.25) indicated that their proficiency ranged between beginner and lower advanced levels [6]. Their perception was coded as either correct or incorrect. Their production was presented to seven native speakers of English, who judged whether they heard /t/ or /d/ in each production. Each production was judged by one rater and subsequently coded as correct (or correctly perceived by the rater) or incorrect.

The data were analyzed in RStudio [8] by using generalized linear mixed models in the *lme4* package [1]. The fixed factors consisted of phase (pretest, posttest) and group (experimental, control), and proficiency (gradient scores). Participant and item were included as random factors. Separate analyses were run for the outcome variables perception and production accuracy. Data produced by the participants who did not speak Dutch as their native language or had dyslexia, DLD, or hearing impairments (8 and 15 for the control and experimental group respectively) were not analyzed to increase homogeneity of the data. The results showed that explicit pronunciation training did not alter participants' pronunciation. However, it improved their perception: only the participants in the experimental group achieved a higher score on the post-test (0.89) compared to the pre-test (0.84) (B = 0.79, p < .01). In addition, the impact of the intervention was not affected by English proficiency.

To conclude, our study has shown that short-term explicit pronunciation training can improve Dutch EFL learners' sound perception, in line with [7], but not necessarily their production, in line with [3] but contra [7], irrespective of language proficiency. This finding suggests that more extensive or longer training is required to improve sound production in Dutch intermediate EFL learners and improvement in phoneme perception may serve as a prelude to improvement in phoneme production.

References

[1] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

[2] Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206). York Press.

[3] Brouwer, C. P. M. (2020). *Using explicit phonetic instruction and performative output training for improving speaking ability in an EFL classroom* [Master's thesis, Utrecht University]. UU Theses Repository. https://studenttheses.uu.nl/handle/20.500.12932/39112

[4] Flege, J. E. (2011). English vowel production by Dutch talkers: More evidence for the "similar" vs "new" distinction. In A. James & J. Leather (Eds.), *Second-language speech* (pp. 11-52). De Gruyter Mouton. https://doi.org/10.1515/9783110882933.11

[5] Isbell, D., Park, O., & Lee, K. (2019). Learning Korean pronunciation: Effects of instruction, proficiency, and L1. *Journal of Second Language Pronunciation*, *5*(1), 13-48. https://doi.org/10.1075/jslp.17010.isb

[6] Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325-343.

[7] Pieper, M. G. (2017). *Bad or bet? The effect of pronunciation teaching on Dutch secondary school pupils' production and perception of English phonemes* [Master's thesis, Utrecht University]. Utrecht University Repository. http://dspace.library.uu.nl/handle/1874/357043

[8] RStudio Team. (2020). *RStudio: Integrated development for R*. RStudio, PBC, Boston, MA. http://www.rstudio.com

[9] Tops, G. A. J., Dekeyser, X., Devriendtand, B., & Geukens, S. (2001). Dutch speakers. In M. Swan, & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (pp. 1-20). Cambridge University Press. https://doi.org/10.1017/CBO9780511667121.003

[10] Wong, J. (2014). The effects of high and low variability phonetic training on the perception and production of English vowels /e/-/æ/ by Cantonese ESL learners with high and low L2 proficiency levels. *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 524-528.

Perception and interpretation of L2 Italian intonation patterns by L1 Italian listeners

[1]Andrea Pešková, [2]Linda Bäumler
[1]*University of Osnabrück,* [2]*University of Vienna*

Intonation is a crucial part of phonological competence and important for the speaker's intelligibility. However, it is said to be very "difficult if not impossible" to be acquired in L2 ([1]). One reason behind this challenge is the fact that learners, especially those at home courses, are mostly confronted with unnatural intonation patterns, a lack of pronunciation training and a large range of regional variation (e.g., [2]). Moreover, L2 speech is characterized by features transferred from the first language (L1) (e.g., [3]). Whereas research on L2 intonation production has grown considerably in recent years (e.g., [4]), we still know very little on how natives perceive and interpret foreign intonation. The aim of this explorative study is to fill this gap, by testing how L1 Italian listeners evaluate L2 Italian utterances.

For this, we created an online perception test using the platform SoSci Survey [5]. Each listener heard 15 short sentences —(non)neutral yes-no questions, wh-questions, statements of the obvious, vocatives (see Fig. 1)— and 6 fillers. All tested items were produced by ten L2 Italian learners (B/C level) with L1 Czech by means of a Discourse Completion Task ([6]). The items were selected according to one criterion: whether they resemble an Italian pattern or not. Hence, half of the items showed a target-like pattern (Italian-like) and half of the items a non-target-like pattern (Czech-like). This accuracy was established according to the realization of nuclear configurations (NC) annotated with AM terms. We limited to the NC, because this part is considered crucial for meaning (e.g., [7]). The two languages differ in several ways. For example, whereas L1 Italian vocatives are realized with L+H* H!H% ([8]), Czech vocatives usually end with L+H* L% ([9]). Hence, if the vocative was produced with L+H* L% in L2 Italian, it was labelled as a Czech-like pattern. On the contrary, the vocative produced with L+H* !H% was labelled as Italian-like. Since intonation conveys both emotions and linguistic meaning, listeners had to judge all items with several linguistic (e.g., the speaker is Italian) and paralinguistic attributes (e.g., the speaker is polite) on a five-point Likert scale.

The results reveal that the L1 listeners (N=204 from all over Italy) identified all items as non-Italian and in most cases had no problems with the semantic interpretation of the utterances. Italian-like NCs were judged towards "more Italian" only with imperative wh-questions, echo yes-no questions and vocatives (see Fig. 1). The attributes *angry, polite, friendly, bored* were evaluated in a very similar fashion for Czech-like vs. Italian-like patterns. *Surprise* was the only attribute showing large differences between the Czech-like vs. Italian-like group: For example, the statement of the obvious with the Italian-like pattern H*+L L% was qualified with "more surprise", whereas its Czech-like counterpart (L*+H L%) was qualified with "no surprise" at all. Furthermore, imperative wh-questions were perceived as less polite and less friendly in comparison to other sentences. This was expected, since they were produced in a context, where the speaker was supposed to be upset with the interlocutor. Interestingly, vocatives showed a tendency towards "less polite" and "less friendly", albeit they were embedded in a neutral situation, in which the speaker had to call a friend on the street.

Since the results show —at least partially— no differences between Czech-like vs. Italian-like patterns, the nuclear configuration might not be the only determinative cue for interpreting foreign patterns and its role might just be secondary. We assume that other prosodic and segmental phenomena are involved in perceiving "otherness" too. In a next step, we will verify regional varieties of the listeners, which might also play a role in perception and interpretation. And finally, we will talk about the possibility to embed the "paralinguistic dimension" into L2 speech models such as L2 Intonation Learning Theory [4] and discuss implications the findings have for foreign language learning and teaching.

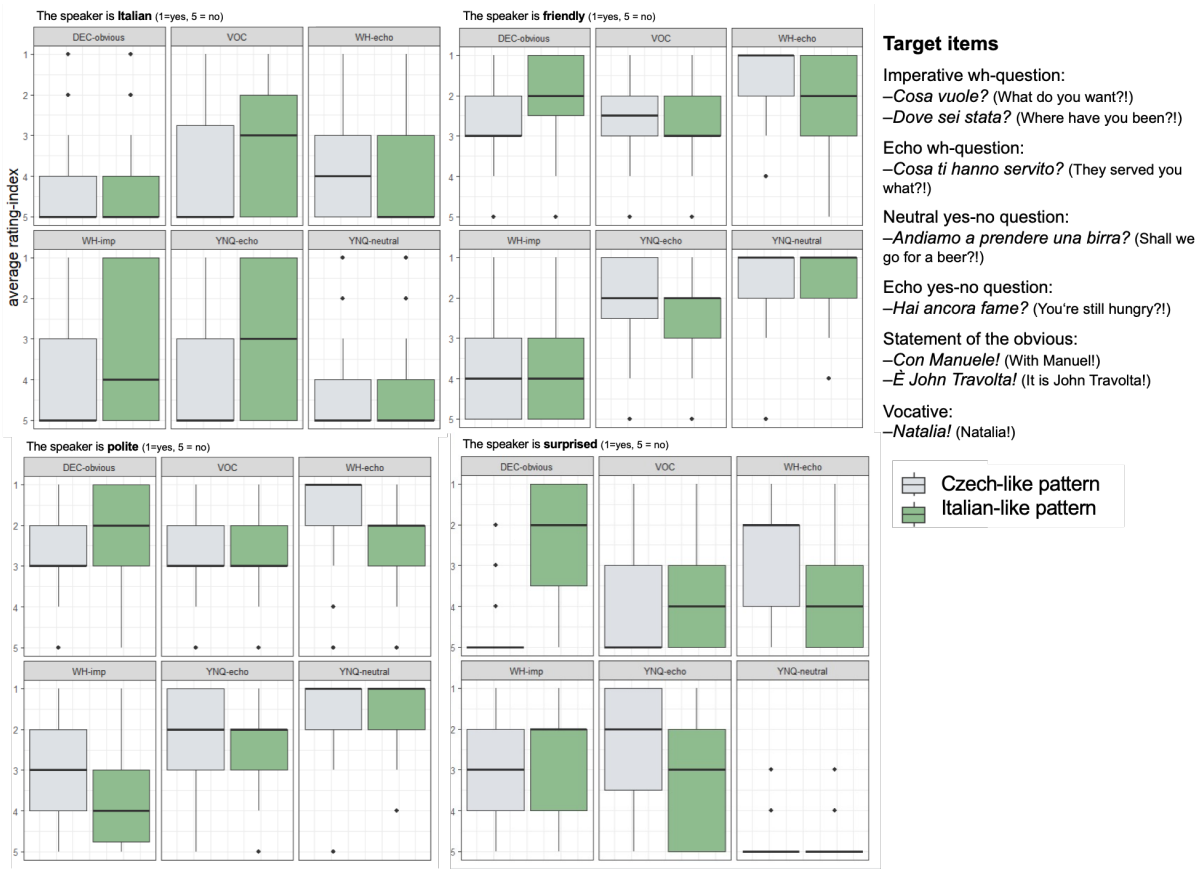


Figure 1. *Examples of the evaluation of the items according to the attributes "Italian"* (top left), *"Friendliness"* (top right), *"Politeness"* (bottom left), *"Surprise"* (bottom right).

References

[1] Chun, D. 2002. *Discourse Intonation in L2. From Theory and Research to Practice*. Amsterdam/Philadelphia: John Benjamins.

[2] Pešková, A. 2020. *L2 Spanish and Italian intonation: Accounting for the different patterns displayed by L1 Czech And German learners*. [Postdoc. Manuscript; Osnabrück]

[3] Colantoni, L.; Steele, J. & P. Escudero. 2015. *Second Language Speech. Theory and Practice*. Cambridge: Cambridge University Press.

[4] 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*. Heidelberg: Springer, 171–188.

[5] Leiner, D. J. (2019). *SoSci Survey* (Version 3.1.06) [Computer software]. Available at https://www.soscisurvey.de

[6] Frota, S. & P. Prieto (Eds.). 2015. *Intonation in Romance*. Oxford: Oxford University Press.

[7] Ladd, R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[8] Gili Fivela, B.; Avesani, C.; Barone, M.; Bocci, G.; Crocco, C.; D'Imperio, M., Giordano, R.; Marotta, G.; Savino, M. & P. Sorianello. 2015. *Intonational Phonology of the Regional Varieties of Italian*. In S. Frota & P. Prieto (Eds.): *Intonation in Romance*. Oxford: Oxford University Press, 140–197.

[9] Pešková, A. 2017. *Czech ToBI*. Presented at Phonetics and Phonology in Europe 2017 in Cologne (Germany).

The Voiced Palatal Lateral in L1 Catalan: Maintenance or Substitution Process?

Agnès Rius-Escudé[1] and Dolors Font-Rotchés[1]
[1]*Universitat de Barcelona*

In the standard variety of Catalan, the contrast between a palatal lateral consonant or an approximant, [ʎ] and [j], respectively, is still present [1, 2, 3, 4, 5, 6]. However, some young Catalan speakers are now substituting the [ʎ] sound with [j], a phenomenon called *ieisme* — and, in some cases, with other sounds.

In view of this, we designed the study with the following objectives: a) analyse which phonetic categories are produced by young LI Catalan speakers when they have to produce the voiced palatal lateral sound [ʎ]; b) verify if, after doing perception and production activities on the voiced palatal lateral, there are changes in the categories and an improvement in the pronunciation of the sound; c) determine whether the position of the sound is a factor that influences its production and learning.

For this study, we created a corpus with 17 female L1 Catalan informants, aged between 18 and 22. These participants, in the pre-training phase, recorded (in .wav format) the names of the objects they saw in twenty images that contain the sound [ʎ] in different positions. We have, however, only used the words with the sound in the initial (#CVC, e.g. *ll*una 'moon') and intervocalic (VCV, e.g. ampo*ll*es 'bottles') positions. Once the results obtained had been analysed, tutored activities were carried out over three weeks with a professor in the classroom: a) three audio perception exercises using the *Kahoot* application to distinguish between [ʎ] and [j]; b) three reading exercises using texts containing 10 [ʎ] sounds in different positions (10 words, 10 phrases and a paragraph); and c) a 1 min 30 s speech containing at least 8 words with [ʎ]. In the post-training phase, participants repeated the activity carried out in pre-training.

We obtained and analysed a total of 284 samples, 138 from the pre-training phase and 146 from the post-training phase. In Praat [7], all the audio files were manually labelled based on the transcription of SAMPA symbols [8] and the category of the types of sounds based on the structure of the formants. F_1 and F_2 values from the mid-point of each isolated consonant production were measured in Praat. A GLMM was run in R (v. 4.1.1) [9] to analyse the production of [ʎ] versus other non-target categories ([j], [ʒ], [dʒ], [ɟ]) in initial and intervocalic positions in the pre-training and post-training phases.

In the pre-training phase, 82 % of palatal laterals [ʎ] were pronounced — with a certain tendency to better pronounce it in the initial position — and the rest have been sub-divided into four types in the onset position (#CVC and CVC): [j], [ʒ], [dʒ] and the voiced palatal plosive [ɟ], of which the majority is [j] with 11.5 %. Of note is that they only produced the voiced fricative and affricate in the initial position (#CVC). After completing the perception and production activities, in the post-training phase, a significant improvement in the pronunciation of [ʎ] was attested (p=.0017): the only case with a [ɟ] disappeared, there were fewer cases of [j] and [ʒ] and the number of voiced palatal laterals, [ʎ], increased to 91 % (Fig. 1), in both the initial (p=.008) and intervocalic (p=.014) positions. In contrast, the affricate sound [dʒ] has not evolved at all to the palatal lateral, which we believe is due to the idiosyncrasy of the speaker but would need to be verified.

For L1 Catalan speakers, even though the [ʎ] sound clearly makes up the majority, a tendency has commenced, which is still weak, to substitute it for the voiced approximant [j] and, to a lesser degree, for the fricative [ʒ]. The activities significantly favour the learning of the voiced palatal lateral among L1 Catalan speakers and we have detected no differences in learning conditioned by the position, either initial or intervocalic.

References

[1] Recasens, D. 1991. *Fonètica descriptiva del català (Assaig de caracterització de la pronúncia del vocalisme i consonantisme del català al segle XX)*. IEC.
[2] Veny, J. 1978. *Els parlars catalans*. Moll. 1987 (7a edició).
[3] Prieto, P. 2004. *Fonètica i fonologia: els sons del català*. UOC.
[4] Rost, A. 2016. La percepción de [ʎ] y /j/ en catalán i en español. Implicaciones en la explicación del yeísmo. *Estudios de Fonètica Experimental* XXV, 39-80.
[5] Rost, A. 2020. Bilingualism and sound change: perception in the /ʎ/-/j/ merger process in Majorcan Spanish. Zeitschrift für romanische Philologie (ZrP) 136(1), 106-133.
[6] Recasens, D. 2014. *Fonètica i fonologia experimentals del català*. IEC.
[7] Boersma, P. & Weenink, D. 2021. *Praat: Doing phonetics by computer* (v. 6.1.40). University of Amsterdam. http://www.praat.org
[8] Llisterri, J. 1995. *A proposal for Catalan SAMPA*. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona. Retrieved from http://liceu.uab.cat/~joaquim/language_resources/SAMPA_Catalan.html
[9] RStudio 4.1.1. 2021. *RStudio: Integrated Development Environment* for R. Boston, MA: RStudio, Inc. [Computer program]. Retrieved from http://www.rstudio.com

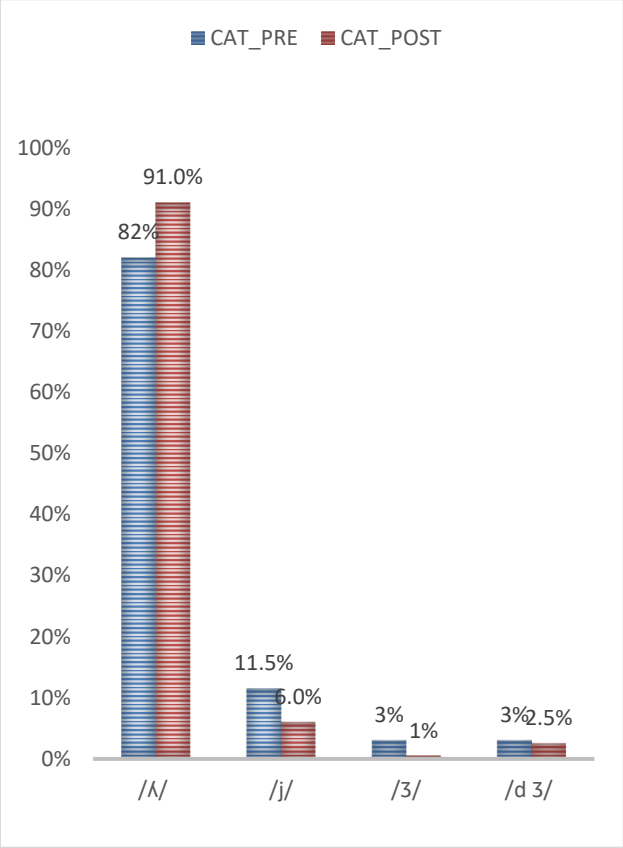


Figure 1. Phonetic categories of pre-training and post-training productions of [ʎ] in the onset position of Catalan.

Prominence augmentation and maximization of contrast via glide epenthesis in Brazilian Portuguese

Lucas Pereira Eberle[1]
[1]*Universidade Estadual de Campinas*

This study analyzes hiatus resolution through the epenthesis of a homorganic glide in Brazilian Portuguese, as in 'fr[ea]r' – 'fr[eja]r' ('to break') and 'v[oɐ]' – 'v[owɐ]' ('to fly' 3p.sg.), and also the monophthongization via glide deletion, as in 'cad[ej]ra' – 'cad[e]ra' ('chair') and 'r[ow]pa' – 'r[o]pa' ('cloth') through a new perspective which considers positional prominence and vocoids contrast as relevant factors for the triggering or not of these phenomena.

According to Beckman (1998), Smith (2005), and Becker et al. (2018), root, stressed, or initial syllables (also monosyllables) are prominent positions, either psycholinguistically and/or phonetically, and, therefore, demand prominent segmental material. Thus, the hypothesis is that through epenthesis the prominence is increased in strong syllables and that glide deletion will be less frequent in these syllables because it would decrease prominence. While in non-prominent positions, V1 raising or monophthongization is favored.

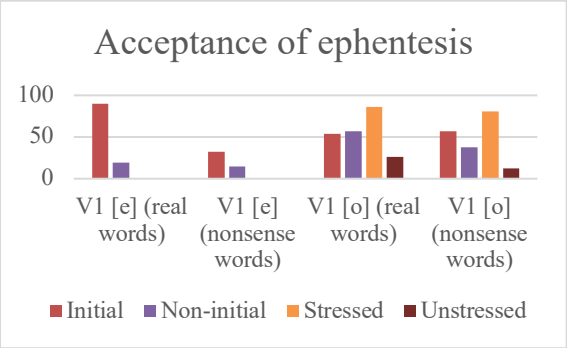
Regarding vocoids contrast, according to Flemming (2004), Nevins (2012), Becker et al. (2018), Borroff (2003), Casali (2011) and others, vowel sequences with low or indistinct contrast (both in sonority and in acoustic dimensions F1 and F2) are less preferred in languages because they are less perceptually distinctive and, consequently, more avoided. Thus, the hypothesis is that the epenthesis also increases the sonority contrast between the vowels of a hiatus, likewise, vowel rising, though decreases sonority of the syllable, increases sonority contrast between V1 and V2, while less distinct diphthongs will tend to monophthongization because they are less perceptible.

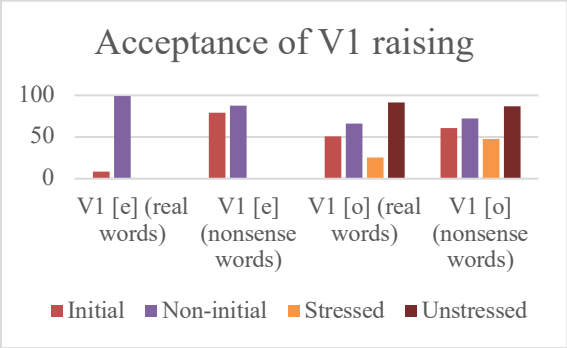
To test the hypotheses that positional prominence and vocoid contrast influence those phenomena, experiments were conducted to investigate the acceptability of glide epenthesis and deletion in real words and nonsense words. The experiments consisted of a Yes/No test, in which participants were presented with different pronunciations of the same word ('fr[ea]r', 'fr[eja]r', 'fr[ia]r' – 'to break' and 'cad[ej]ra', 'cad[e]ra' – 'chair'), and instructed to evaluate them as 'natural' or 'unnatural'. The conditions manipulated were stressed vs unstressed positions, ['vo.ɐ] – [vo'aɾ] ('to fly'), initial vs non-initial, ['how.pɐ] ('cloth') – ['aw.kow] ('alcohol'), and monossylables vs polissylables [gow] ('goal') – [dow'toɾ] (doctor').

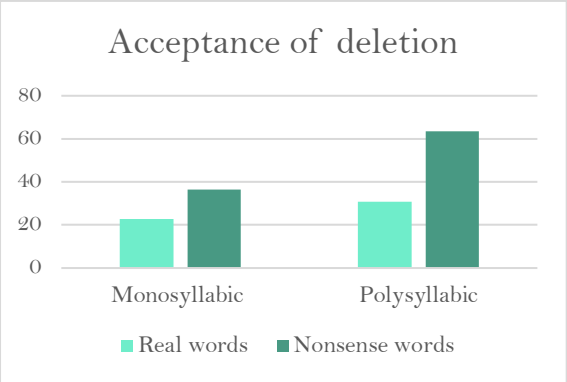
A mixed-effects logistic regression model was fitted using lme4 (Bates et al. 2015) in R (R Development Core Team 2016) and the obtained results corroborated the hypotheses, although asymmetrically. The glide deletion was less preferred (judged 'unnatural') in strong positions, mainly in monosyllables, and the diphthongs in which the glide deletion was considered more acceptable (judged as 'natural') were those with low or indistinct contrast in F2 [ow ej ɔw uw]. Epenthesis was more judged as 'natural' in stressed-V1 positions and roots (strong positions) while raising of V1 was more accepted in unstressed-V1 positions, as presented in (1).

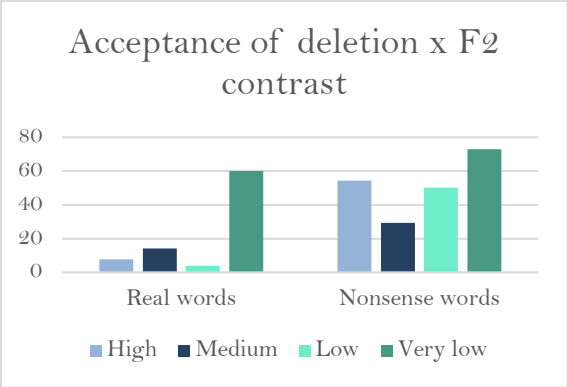
The Maximum Entropy Grammar (Goldwater & Johnson, 2003) was the reference to model the grammar that explains these phenomena. Unlike Optimality Theory (Prince & Smolensky, 1993), it allowed more elegant modeling of grammar with variation, i.e., with the possibility of more than one output. By assigning weights to the constraints and using the frequencies of acceptance obtained in the experiments, it was possible to calculate, using the MaxEnt Grammar Tool (Hayes & Wilson, 2008), which output had the highest prediction in the object language. Although the results showed that the effects of positional prominence and acoustic dispersion were not uniform and symmetrical in these phenomena, the results demonstrate how acoustics and suprasegmental factors influence the perception of vowel sequences in Brazilian Portuguese.

(1) Results summary









References

[1] BATES et. al. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67. 1–48. R package version 0.999999-4.

[2] BECKER, M., NEVINS, A., SANDALO, F., & RIZZATO, É. The Acquisition Path of [w]- final Plurals in Brazilian Portuguese. Journal of Portuguese Linguistics, 17(1), 4. 2018. DOI: http://doi.org/10.5334/jpl.189

[3] BECKMAN, J. N. *Positional faithfulness*. 1998. 270fls. Thesis – University of Massachusetts Amherst, Massachusetts, 1998.

[4] BORROFF, M. L. Against an Onset approach to hiatus resolution. Paper presented at the 77th Annual Meeting of the Linguistic Society of America, Atlanta (ROA-586). 2003.

[5] CASALI, R. F. Hiatus resolution. The Blackwell companion to phonology, p. 1-27, 2011.

[6] FLEMMING, E. Contrast and perceptual distinctiveness. Phonetically based phonology, p. 232-276, 2004.

[7] GOLDWATER, S. e JOHNSON, M. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm workshop on variation within Optimality Theory*. p. 111-120. 2003.

[8] HAYES, B.; WILSON, C. A maximum entropy model of phonotactics and phonotactic learning. Linguistic inquiry, v. 39, n. 3, p. 379-440, 2008.

[9] NEVINS, A. Enfraquecimento e fortalecimento de vogal em português brasileiro. Letras de Hoje, v. 47, n. 3, p. 228-233, 2012.

[10] PRINCE, A. e SMOLENSKY, P. Optimality Theory: Constraint interaction in generative grammar. 1993. Não publicado.

[11] R Development Core Team (2016). R: A language and environment for statistical computing. Vienna, Austria. (http://www.r-project.org).

[12] SMITH, J. L. *Phonological augmentation in prominent positions*. Routledge. 2005.

Conflicting effects of orthography on the de-regionalization of adolescent speech

Šárka Šimáčková[1], Václav Jonáš Podlipský[1], Natalia Nudga[2,3], Nikola Paillereau[2,3] and Kateřina Chládková[2,3]

[1]*Faculty of Arts, Palacký University Olomouc,* [2]*Faculty of Arts, Charles University,* [3]*Institute of Psychology, Czech Academy of Sciences*

Orthography affects the production and perception of sounds in an L2 [1, 2]. Children's L1 phonology also interacts with their knowledge of orthography [3]. Our study asks whether the degree to which adolescents' regional accent approximates non-local speech depends on orthography.

The phonology of Silesian Czech (SC), spoken in the northeast of the Czech Republic, differs from Common Czech (CC), the most widely spoken prestige interlanguage in the country [4], in the presence of the /i/-/ɨ/ contrast (CC having only 1 short high front V, /ɪ/) and the absence of contrasting V length [5, 6] (CC having long Vs, including /i:/). While these two features still distinguish contemporary SC from speech outside the region, they are conditioned by the speaker's socio-economic status, urban or rural background, age, and education.

This study tests these two characteristics in the speech of teenage grammar-school students from Opava, the cultural and economic center of western Silesia. We expect their accent to be affected by schooling, i.e. intensive experience with orthography, predicting opposite outcomes for the two contrasts: maintaining the local /i/-/ɨ/ contrast is supported by the use of two different letters (*"i"* and *"y"* respectively) in standard spelling, while the absence of the long-short V contrast in regional speech is undermined by standard spelling, which consistently marks V length with a diacritic (e.g. *"y"* versus *"ý"*). Productions of the SC speakers are compared to CC speakers from Central Bohemia.

Currently, data from 5 SC and 4 CC speakers (all female, age range 17–20) are analyzed (results of the full dataset of 9+9 speakers will be presented at the conference). Each speaker produced a total of 34 trisyllabic isolated words in a picture-naming task which was interspersed within a longer phrase-production task collecting a larger corpus. The orthographic representations of the intended words comprise 275 instances of the letter *"i"* (32 word-final, 243 non-final), 400 of *"y"* (266 fin., 134 n-fin.), 162 of *"í"* (56 fin., 106 n-fin.) and 132 of *"ý"* (all n-fin.).

Log-transformed V durations of the target Vs, and F1 and F2 (in ERB) of the short *"i"* or *"y"*, were modelled with 3 linear mixed-effects models in R. See Tables 1, 2 and 3 for the model formulas and fixed-effect coefficient estimates. The significant effect of V length without an interaction with Region indicates that the durational distinction between Vs represented in orthography as long and short did not differ reliably between SC and CC speakers. The significant interaction of V length with Position reflects final lengthening, evident especially in SC speakers (Fig. 1). Results for F1 or F2 did not show a significant effect of orthography or an interaction between Orthography and Region, indicating that there is only one short high front vowel category for both the CC and SC speakers irrespective of the *"i"*-*"y"* distinction in spelling. A significant effect of Position on F2 shows that word-final vowels were more retracted (Fig. 3). The significant interaction between Orthography and Position for F1 indicates that vowels represented as *"y"* were more open when word-final (Fig. 2).

These preliminary results suggest that teenage grammar-school-educated Silesian female speakers make a duration difference between vowels represented by standard orthography as long and short (at least in isolated words). Such distinction is *not* part of the phonology of the SC dialect. However, the importance of orthography in shaping these youngsters' sound system is somewhat undermined by their *merging* of the contrasting Silesian vowels /i/ and /ɨ/ – a contrast which in orthography is represented by separate letters. Possible reasons for such conflicting effects of orthography will be discussed.

Fig. 1: Modelled duration of high front Vs

Fig. 2: Modelled F1 of high front short Vs

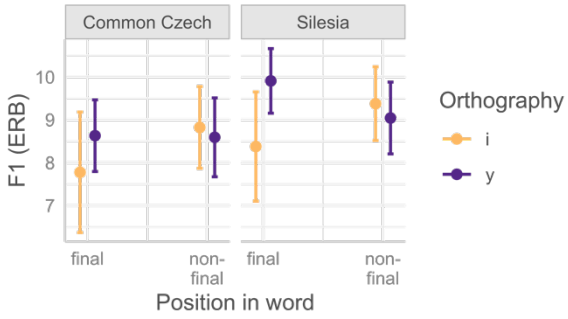
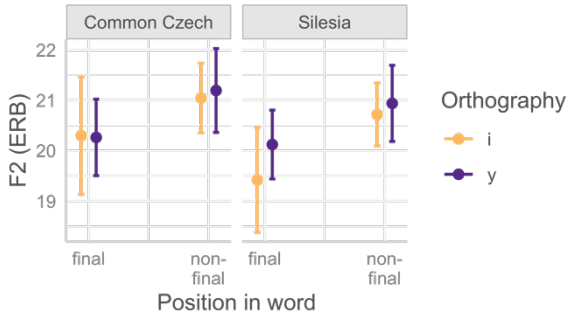


Fig. 3: Modelled F2 of high front short Vs



Model: log(Vdur) ~ Region * V length * Position + (1 + V length | Speaker) + (1 | Word)					
Table 1	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	4.5751	0.0671	10.9601	68.1516	**<0.0001**
Region (-1 CCz, +1 SC)	0.0602	0.0600	7.1123	1.0028	0.3489
V length (-1 long, +1 short)	-0.1313	0.0288	26.6361	-4.5596	**0.0001**
Position (-1 wd-final, +1 non-final)	-0.3080	0.0166	518.9106	-18.5970	**<0.0001**
Region : V length	0.0064	0.0211	8.0416	0.3020	0.7703
Region : Position	-0.0500	0.0098	923.4034	-5.1279	**<0.0001**
V length : Position	-0.0850	0.0172	538.9757	-4.9356	**<0.0001**
Region : V length : Position	-0.0167	0.0098	924.6301	-1.7067	0.0882

Model: F1 ~ Region * Orthography * Position + (1 + Orthography | Speaker) + (1 | Word)					
Table 2	Estimate	Std.	df	t value	Pr(>|t|)
(Intercept)	8.8148	0.3102	10.0546	28.4199	**<0.0001**
Region (-1 CCz, +1 SC)	0.3608	0.2878	7.6593	1.2536	0.2469
Orthography (-1 “i”, +1 “y”)	0.2276	0.1414	20.2914	1.6098	0.1229
Position (-1 wd-final, +1 non-final)	0.1427	0.1392	23.1576	1.0250	0.3160
Region : Orthography	0.0717	0.0905	19.7814	0.7924	0.4375
Region : Position	-0.1094	0.0871	605.1757	-1.2568	0.2093
Orthography : Position	-0.3689	0.1447	19.5110	-2.5506	**0.0193**
Region : Orthography : Position	-0.0974	0.0870	606.6661	-1.1201	0.2631

Model: F2 ~ Region * Orthography * Position + (1 + Orthography | Speaker) + (1 | Word)					
Table 3	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	20.5012	0.2325	10.9717	88.1958	**<0.0001**
Region (-1 CCz, +1 SC)	-0.2005	0.2125	8.1026	-0.9436	0.3726
Orthography (-1 “i”, +1 “y”)	0.1299	0.1525	17.0266	0.8519	0.4061
Position (-1 wd-final, +1 non-final)	0.4742	0.1225	20.9225	3.8707	**0.0009**
Region : Orthography	0.1014	0.1238	11.0649	0.8195	0.4298
Region : Position	0.0547	0.0842	603.8365	0.6503	0.5157
Orthography : Position	-0.0369	0.1263	17.7761	-0.2924	0.7734
Region : Orthography : Postion	-0.0833	0.0842	604.3006	-0.9896	0.3228

References

[1] Bassetti, B., Sokolović-Perović, M., Mairano, P., & Cerni, T. (2018). Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Eviden-ce from minimal pairs. *Lang. Speech*, *61*(4), 577-597.

[2] Stoehr, A., & Martin, C. D. (2022). The impact of orthographic forms on speech production and perception: An artificial vowel-learning study. *J. Phonetics*, *94*, 101180.

[3] Stage, S. A., & Wagner, R. K. (1992). Development of young children's phonological and orthographic knowledge as revealed by their spellings. *Developmental Psychology*, 28(2), 287.

[4] Krčmová, M. (2017): Obecná čeština [Common Czech]. In: Karlík, P. et al. (eds.), CzechEncy: Nový encyklopedický slovník češtiny. https://www.czechency.org/slovnik/OBECNÁ%20ČEŠTINA. Accessed 15/1/2023.

[5] Bělič, J. 1972. *Nástin české dialektologie [A sketch of Czech Dialectology]*. SPN.

[6] Blažková, J. 2008. *Výslovnost dvojího "i" ve východolašském dialektu [Pronunciation of two i’s in the Eastern Lachian dialect]*. MA dis., Charles University, Prague.

Higher pitch, lower range, and slower gestures but not slower speech: An audiovisual comparison of child-directed with adult-directed broadcasting

Yan Gu, University of Essex
Yanran Zhang, Communication University of Zhejiang

Researchers who study language production debate whether audience design is driven by speakers' needs or by their audiences' needs (e.g., Arnold et al., 2012; Aylett & Turk, 2004). TV broadcasting has a distinct recipient design where speakers seriously care about their imagined audiences. In broadcast discourse, intonation is typically carried by a higher pitch and a larger pitch variation than in ordinary conversations (Price, 2008). In addition, a study showed that head beats and eyebrow movements were associated with the phonological prosodic structure in Swedish newsreaders (Ambrazaitis & House, 2017). Furthermore, Swerts and Krahmer (2010) made the first attempt to investigate multiple cues in adult- and child-directed programmes and found that newsreaders are more expressive when addressing children than adults. However, their broadcasting research had a small sample size (N=2) and used a between-subject comparison. Moreover, across languages (Cox et al., 2022), child-directed speech is characterised by exaggerated intonation, a wider pitch range and a slower speaking rate (e.g., Cristia, 2013; Han et al., 2022), and speakers increase their iconic gesture rate for a child listener compared to an adult listener (Campisi & Özyürek, 2013). It is still unclear how the same broadcasters adjust their multimodal communication when producing the same content for adults and children.

We investigated how broadcasters organize their audiovisual language production on both a regular adult-directed and a child-directed programme. Thirty-six future broadcasters produced live programmes in which they explained four pictures to an imagined adult and child audience, respectively (within-subject, counterbalancing sequences). To avoid gender differences (e.g., Kiepura et al., 2021; Sim, 2021), participants were all female. We analysed speech prosody (e.g., F0, intensity, speaking rate) of 3888 utterances through Praat (Boersma & Weenink, 2019) and visual cues of 8486 gestures (e.g., gesture types, gesture rate, duration, and saliency, Chu et al., 2014) through Elan (Wittenburg et al., 2006) as a function of broadcasting programmes.

The findings revealed that (Table 1 and Table 2), compared to adult-directed programme, child-directed broadcasting had 1) a higher mean F0, F0_max and F0_min (all *ps* < .001), but a *smaller* F0 range ($p < .001$). The smaller F0 range is due to a ceiling effect that the increase in F0_max (0.47 semitone, $p<.001$) was less than that of the F0_min (0.89 semitone, $p<.001$); 2) more salient gestures (larger size) ($p=.006$); 3) a higher representational ($p < .001$) and pointing gesture rate ($p < .001$) but a lower pragmatic/interactive gesture rate ($p < .001$); 4) was not slowed down in speaking rate ($p=.65$) but slowed down in each representational gesture (gesture phases lasted longer) ($p=.002$).

Our study made the first attempt to quantitively document multimodal child-directed and adult-directed communication in a broadcasting context. It shows that certain signal channels are not audience-oriented, whereas other channels may compensate for it (gesture rate vs. speaking rate). The distinction between speaker-and audience- orientation is not binary, but the mechanism should be understood as the incorporation of both processes simultaneously via different information channels according to the context.

Table 1. Means (and SD) and results (beta and *p*) of prosodic cues for each condition.

Dep Variables	ProgrammeADB	ProgrammeCDB	*p* values
Mean_F0 (ST)	27.15(1.22)	27.99(1.29)	0.83***
SDF0 (ST)	58.47(6.39)	60.73(6.98)	2.31***
Max_F0 (ST)	36.07(1.52)	36.53(1.37)	0.47***
Min_F0 (ST)	16.14(2.07)	17.04(1.91)	0.89***
F0 range (ST)	15.08(2.59)	14.28(2.30)	-0.78***
Speaking rate	0.71(0.03)	0.71(0.04)	-0.002
Avg_Intensity (dB)	58.13(3.33)	58.30(3.57)	0.18
Intensity range (dB)	34.91(2.47)	34.88(2.55)	-0.034
Pausing rate	0.33(0.07)	0.33(0.07)	0.36
Avg. pause dur (sec)	0.40(0.09)	0.40(0.10)	0.54
Numbers of pauses	24.72(8.60)	23.83(8.64)	-0.94

Table 2. Means (and SD) and results (beta and *p*) of gestures for each condition.

Dep Variables	ProgrammeADB	ProgrammeCDB	*p* values
reprentationalgesture_rate (per sec)	0.07(0.06)	0.10(0.08)	2.51***
pointing_rate (per sec)	0.06(0.05)	0.08(0.07)	0.02***
pragmatic_rate (per sec)	0.19(0.11)	0.12(0.10)	-0.04***
beat_rate (per sec)	0.11(0.06)	0.10(0.07)	-0.01
average_beat_times	2.26(1.52)	1.94(0.89)	-0.32**
mean_duration_represent_gesture (sec)	1.17(0.59)	1.35(0.54)	0.19**
gesture saliency	6.32(0.97)	6.55(1.43)	0.23**

References

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.

Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, *19*(3), 505-512.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14-27.

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour,* 1-20.

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, *7*(3), 157-170.

Han, M., De Jong, N. H., & Kager, R. (2022). Prosodic input and children's word learning in infant- and adult-directed speech. *Infant Behavior and Development*, *68*, 101728.

Price, J. (2008). New news old news: A sociophonetic study of spoken Australian English in news broadcast speech. *AAA: Arbeiten Aus Anglistik Und Amerikanistik*, *33*(2), 285-310.

Swerts, M., & Krahmer, E. (2010). Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics*, *38*(2), 197-206.

Poster Session 2

Saturday, 15:00 - 16:20

The relevance and weighting of prosodic cues in question interpretation

Marieke Einfeldt[1], Ekaterina Kazak[2], Angela James[1], Rita Sevastjanova[1], Daniela Wochner[1], Katharina Zahner-Ritter[3], Nicole Dehé[1] and Bettina Braun[1]
[1]*University of Konstanz,* [2]*University of Manchester,* [3]*University of Trier*

We investigate the well-formedness of rhetorical questions (RQs) and information-seeking questions (ISQs) in German by means of a multiple-cue perception study. Previous research has shown that RQs are produced with longer duration and more frequent breathy voice quality than ISQs. Rhetorical *wh*-questions showed a steep low-rising accent (LH)* and an almost obligatory fall (L-%); *wh*-ISQs had a rising nuclear accent L+H* and both low and rising edge tones. Polar RQs mostly had a steep low-rising nuclear accent (LH)* and a final mid-high plateau (H-%), while polar ISQs most often had a low or low-rising nuclear accent (L*/L+H*) and a high-rising edge tone (H-^H%; [1]). Data from a semi-spontaneous production task showed that speakers preferred *wh*-syntax for RQs over polar question syntax [2].

To test the relative importance of different prosodic cues and cue combinations in the prenuclear and nuclear region, two perception studies were conducted using an Active Learning System (AL [3], cf. [4]). We created 12 polar (1) and 12 *wh*-questions (2). Overall, we tested 9216 items, which prosodically were combinations of (i) one of three prenuclear accents or no accent (later recoded as the binary variable present/absent), (ii) one of six nuclear accents, (iii) one of four edge tones, (iv) modal or breathy voice in the prenuclear region, and (v) overall long or short duration. Participants judged whether a given question was an ISQ or not (Ex. I) or an RQ or not (Ex. II). The AL allowed for keeping the human labelling effort feasible; each participant labelled 88 items in each experiment. The AL learned the cue weights iteratively. Participants were split into two groups by the AL according to whether they were sensitive to the nuclear accent and edge tone (Group I) or not (Group II), based on their responses to the first 24 items. For each experiment, we report results of the larger Group I (Ex. I: 77% of participants; Ex. II: 84%). We used a boosting algorithm [5] on the labelled data to determine the cues that improved each model's accuracy. Separate linear regression models tested the cues' contribution to ISQ/RQ interpretation (N=384 conditions). We used the cues with an accuracy gain <1% as reference level (intercept) in the regression models.

In **Ex. I (ISQs)**, the responses of 62 participants led to the following cues as intercept: *wh*; breathy; long; prenuclear accent present; the sets H*/L+H* and L-%/H-%. Nine cues or combinations thereof increased the probability of ISQ interpretation (light grey bars in the left-hand panel of Figure 1), while five decreased it (dark grey bars). In **Ex. II (RQs)**, the responses of 67 participants led to the following cues as intercept: polar; modal; short; no prenuclear accent; the sets H*/L+H*/H+!H* and L-%. Ten cues or combinations thereof increased the probability of RQ interpretation (light grey bars in the right-hand panel of Figure 1). Only the presence of a prenuclear accent decreased the probability of RQ interpretation (dark grey bar).

Ex. I and II corroborate findings from previous production studies ([1], [2]): The cues long duration, breathy voice quality and *wh*-syntax as well as combinations thereof increased the RQ interpretation in Ex. II (see Figure 1), while the respective counterparts increased the ISQ interpretation in Ex. I. With respect to intonation, (LH)* and H-% increased the RQ interpretation in Ex. II and decreased an ISQ interpretation in Ex. I. Also, L* L-% increased the RQ interpretation and decreased an ISQ interpretation, indicating that flat contours are more often interpreted as rhetorical. Two accent types (H*, L+H*) and the edge tone L-% were compatible with both ISQs and RQs. Overall, our data corroborate earlier studies on the cues relevant to RQs and ISQs, yet with a much bigger dataset and a larger range of cues. Results will be discussed with respect to the modelling of a prosodic grammar for RQs vs. ISQs.

(1) Mag denn jemand Karneval? (polar question)
Does anyone like carnival?
(2) Wer mag denn Karneval? (wh-question)
Who likes carnival?

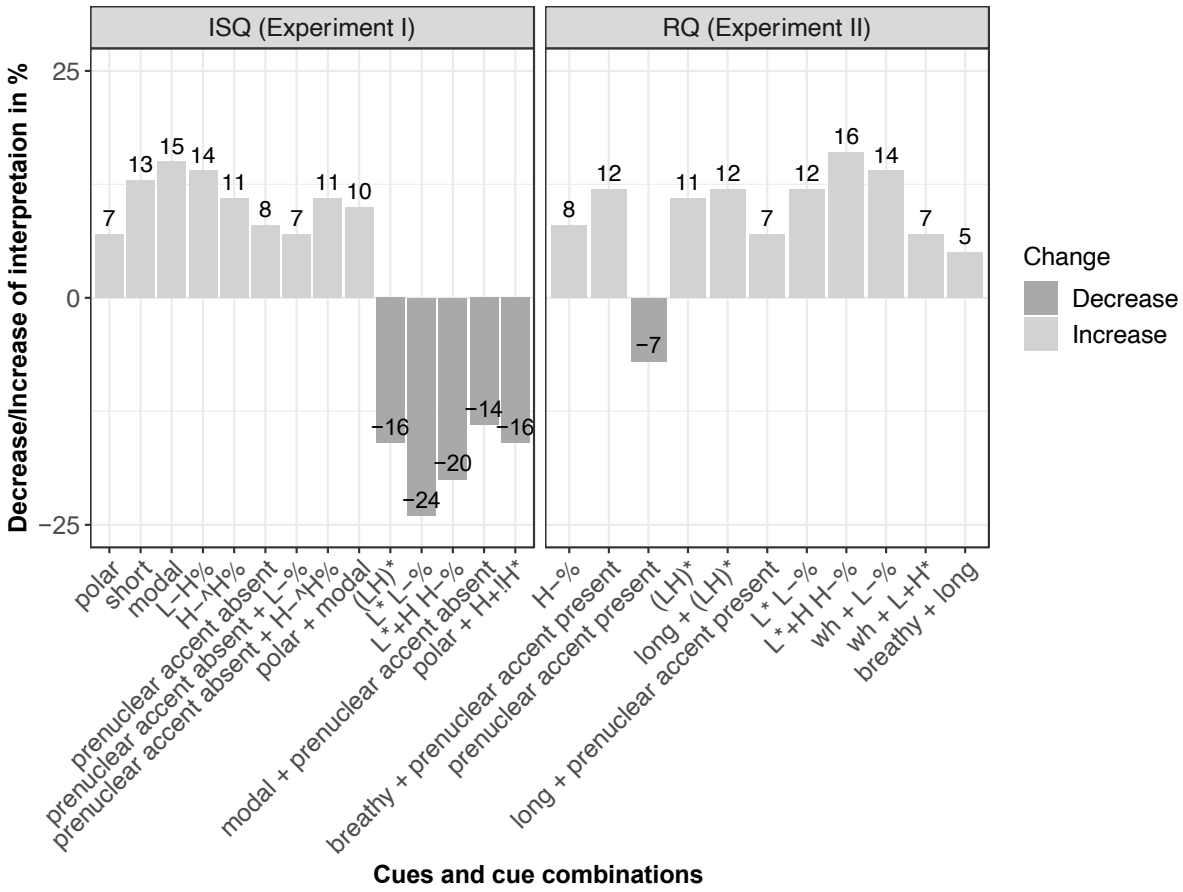


Figure 1. *Cues increasing (light grey) and decreasing (dark grey) the ISQ (left panel) or RQ interpretation (right panel).*

References

[1] Braun, B., Dehé, N., Neitsch, J., Wochner, D. & Zahner, K. 2019. The prosody of rhetorical and information-seeking questions in German. *Language and Speech* 62(4), 779–807.

[2] Dehé, N., Wochner, D. & Einfeldt, M. 2022. The interaction of discourse markers and prosody in rhetorical questions in German. *Journal of Linguistics*, 1–25.

[3] Settles, B. 2009. *Active Learning literature survey*. Computer Sciences Technical Report at the University of Wisconsin-Madison.

[4] Einfeldt, M., Sevastjanova, R., Zahner-Ritter, K., Kazak, E. & Braun, B. 2021. Reliable estimates of interpretable cue effects with Active Learning in psycholinguistic research. *Proceedings of Interspeech 2021* (Brno, Czechia (virtual)).

[5] Chen, T. & Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco California USA), 785–794.

A comparison of Spanish and Gipuzkoa Basque vowels produced by Basque-Spanish bilinguals

Peng Li[1], Clara Martin[2], and Natalia Kartushina[1]
[1]*Center for Multilingualism in Society across the Lifespan, University of Oslo, Norway*
[2]*Basque Center on Cognition, Brain and Language, Spain*

Bilinguals partition the phonological systems of their languages as a function of language proficiency and use. For instance, Catalan-dominant, but not Spanish-dominant, Catalan-Spanish bilinguals keep the mid-vowel distinction in Catalan speech, a contrast that does not exist in Spanish [1], [2]. Simultaneous Basque-Spanish bilinguals delaterize the palatal lateral /ʎ/ to fricative /ʝ/ in Spanish but keep the /ʎ-ʝ/ distinction in Basque [3]. Interestingly, simultaneous bilinguals may divide the vowel space to fit two L1 vowel systems despite redundancies [4]. However, previous research mostly looked at the phonological differences between bilinguals' two L1s where one language lacks a contrast present in the other (e.g., Catalan /e-ɛ/ vs. Spanish /e/). Are there differences systematically? Do bilinguals distinguish the phonemes in production when their L1s have identical phonemes, for instance, vowels?

This study took Basque and Spanish as our target languages as both languages have five vowel phonemes /a, e, i, o, u/. Based on previous research on early bilinguals, we hypothesized that bilinguals would show distinctive phonetic realizations of the two sets of vowels in Basque and Spanish. In addition, although the acoustic descriptions of several Basque variants have been documented, like Mixean [5] and Goizueta Basque [6], some variants are still underinvestigated, such as Gipuzkoa Basque, which is the most spoken in the Basque Country.

Fourty-seven Basque-Spanish early bilinguals from Gipuzkoa, Basque Country, Spain, participated in a word reading task. They read five Basque words and five Spanish words five times. The first syllable of each word contained one of the five Basque and Spanish vowels (e.g., Basque *bOtu* 'vote'; Spanish *bOta* 'boot'). This design elicited 2,349 tokens (47 participants × 5 words × 2 languages × 5 repetitions - 1 missing token). We manually annotated the steady part of each target vowel and extracted the mid-point of the first three formants (F1, F2, and F3). Then, the formant values were transformed into Bark values (Z1, Z2, and Z3). The vowel height was thus represented by Z3-Z1 (henceforth Height), and the vowel backness was represented by Z3-Z2 (henceforth Backness). Figure 1 visually plots the data with Bark normalized values.

We built two Linear Mixed Models with vowel Height and Backness as the dependent variables. For both models, the fixed effects included vowel (/a, e, i, o, u/), language (Basque vs. Spanish), and their interaction; the random effects included a random intercept of participant with two random slopes, one for vowel and the other for language. We found a significant two-way Vowel by Language interaction for Height, $\chi^2 = 13.21$, $p = .010$, and Backness, $\chi^2 = 19.87$, $p = .001$. Post-hoc pairwise comparisons adjusted with false discovery rate method confirmed that the Basque /i/ was lower than the Spanish /i/, $t = 2.35$, $p = .019$; the Basque /o/ and /u/ were more fronted than the Spanish /o/, $t = 3.89$, $p < .001$, and /u/, $t = 4.29$, $p < .001$, respectively. No significant between-language differences were found for /a/ and /e/. Hence, bilinguals demonstrated differentiation of the two vocal systems of their native languages, although the two languages had the same vowel inventories at the phonological level. Note that there was considerable between-speaker variability in production.

To conclude, this study provided new data on vowel production in a relatively less reported language, the Gipuzkoa Basque, that could be used as a reference for Basque (monolingual Basque speakers are quasi-inexistent in the Basque Country). Our bilingual data partially supported the assumption that early bilingual speakers try to tease the two phonological systems apart in speech production. Since there was considerable variability between participants, further analyses will tackle the role of language use and switching habits in the degree of overlap between bilinguals' two phonological systems.

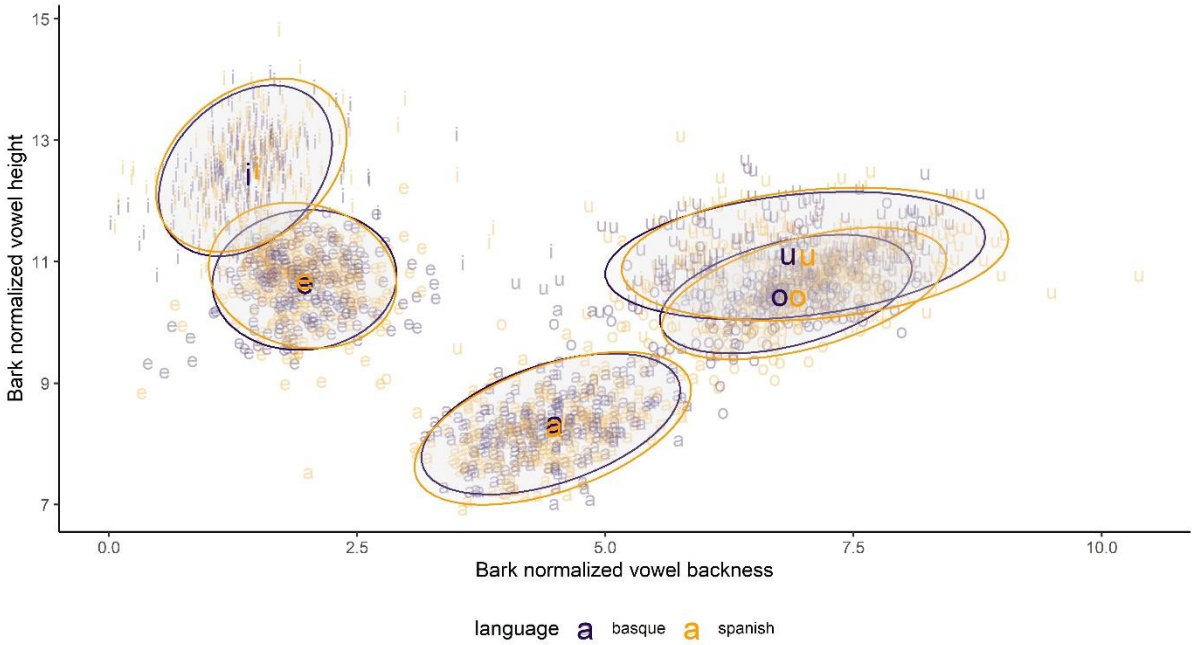


Figure 1. *Bark-normalized vowel height and vowel backness for Gipuzkoa Basque and Spanish vowels. Boldface letters represent the mean value; ellipses cover 67% of the data in each vowel category; and the light-coloured letters represent individual tokens.*

References

[1] Amengual, M. (2016). The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. *International Journal of Bilingualism*, *20*(2), 133–152. https://doi.org/10.1177/1367006914544988

[2] Mora, J. C., Keidel, J. L., & Flege, J. E. (2015). Effects of Spanish use on the production of Catalan vowels by early Spanish-Catalan bilinguals. In J. Romero & M. Riera (Eds.), *The Phonetics–Phonology Interface: Representations and methodologies* (pp. 33–54). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.335.02mor

[3] Beristain, A. (2021). Type of early bilingualism effect on the delateralization of /ʎ/ in Basque and Spanish. *Linguistic Approaches to Bilingualism*, *11*(5), 700–738. https://doi.org/10.1075/lab.19082.ber

[4] Guion, S. G. (2003). The Vowel Systems of Quichua-Spanish Bilinguals. *Phonetica*, *60*(2), 98–128. https://doi.org/10.1159/000071449

[5] Egurtzegi, A., & Carignan, C. (2020). An acoustic description of Mixean Basque. *The Journal of the Acoustical Society of America*, 147(4), 2791–2802. https://doi.org/10.1121/10.0000996

[6] Hualde, J. I., Lujanbio, O., & Zubiri, J. J. (2010). Goizueta Basque. *Journal of the International Phonetic Association*, *40*(1), 113–127. https://doi.org/10.1017/S0025100309990260

Dialogical speech and quality of life in Italian speakers affected by hypokinetic dysarthria

Barbara Gili Fivela, Anna Chiara Pagliaro, Sonia d'Apolito
University of Salento & CRIL-DReAM, Lecce

Hypokinetic dysarthria is a speech motor disorder that involves a decrease in muscle strength and tone, resulting in reduced range and/or accuracy in the coordination of movements [1]. As far as speech is concerned, it causes changes in phonation, in the articulation of segments and in the production of prosody, with a marked variability in speech rate related to sudden accelerations and 'inappropriate' pauses [2,3,4,5]. The resulting speech characteristics play an important role in modifying communicative abilities, as lowering speech accuracy affects the intelligibility of the message, but also in modifying the speaker's perception of his or her psycho-social identity - which is built in linguistic interactions, changes over time and is not homogeneous [6,7] - with consequences for the speaker's perception of the quality of his or her life. Various self-assessment questionnaires are used to quantify what impact a certain pathology, a certain stage of its evolution, or a therapeutic intervention has on a patient's life. The main goals of the present investigation is to verify if the phonetic analysis of the oral productions of Italian dysarthric speakers supports the self-assessment of their quality of life (QoL-Dys [8]), with a focus on interaction-related aspects. The hypotheses are that phonetic features are affected by pathology and correlate with subject's assessments.

Twelve Italian speakers particpated in the study: eight subjects suffering from Parkinson's disease and mild hypokinetic dysarthria (PDs), together with four control speakers (CTRs). They were not cognitively impaired (MOCA >= 24), they were from and lived in the Apulia region and were age matched as much as possible (mean age: PDs 63 y.o., CTRs 59 y.o.). Subjects were asked 1) to participate in a Map-Task dialogue [9], and 2) to describe, also through interaction with the experimenter, images corresponding to some of the icons also found in the Map-Task. Acoustic data were acquired in a quite room with an external sound card (Edirol capture UA-5) and Shure directional microphones (SM86), using two separate channels. Speech productions were transcribed orthographically [10] and manually segmented into units of different hierarchical and qualitative levels by implementing different tiers in PRAAT. Analyses regard (a) descriptive statistics on control and dysarthric speech, concerning a1) percentage and type (e.g., Giver/Follower) of dialogic turns, a2) number and position of pauses, broken down into infra- and intra-speech turn pauses; b) acoustic measures, such as b1) dialogue duration; (b2) pause duration; (b3) articulation and speech rate; (b4) the disfluency index [11]. A Z test of proportions is used to statistically compare descriptive data (a) concerning PD and CTR speech, while Linear Mixed Effect Models are used to investigate acoustic measures in (b). Finally, correlations between QoL-Dys scores and acoustic measurements are investigated by means of the Kendall's coefficient tau-b.

Preliminary results on 4 out of 8 PDs and 3 out of 4 CTRs show that some of the investigated measures are affected by pathology and also seem to be correlated with subjects' self-evaluation. Specifically, average duration of dialogues and turns tend to be shorter in pathological subjects, the duration of Pause_infra is highly variable (e.g., remarkably long in PD2_F – see Fig.1), Pause_intra seems less variable and slightly shorter in PDs in MapTask dialogues only (Fig.1, left), and the Disfluency Index tend to be higher in PDs, especially in MapTasks and mostly in Giver turns (see Fig.2). Further, speakers QoL-Dys evaluation correlates with some acoustic measures, as in the case of duration of infra_turn Pauses and the perception of being slowly in speaking and finding difficulties in talk during emergency, or the high disfluency index correlating with the difficulty in asking long and articulated questions. Of course, more material and analyses are needed to robustly generalize results, though they suggest that self-evaluations may also offer keys to interpret the variability that is often observed in speech characteristics of dysarthric subjects.

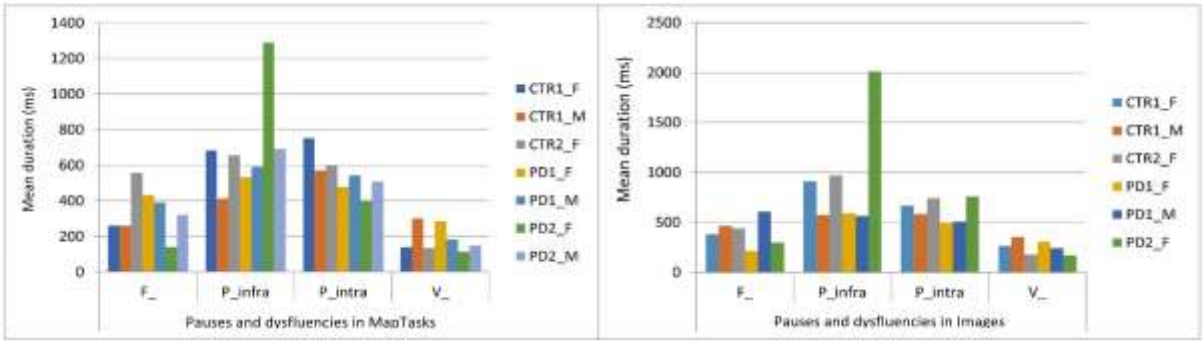


Figure 1. *Average duration of pauses and disfluencies in Map-task dialogues (left) and Image interactive descriptions (right) by 4 PDs and 3 CTRs.*

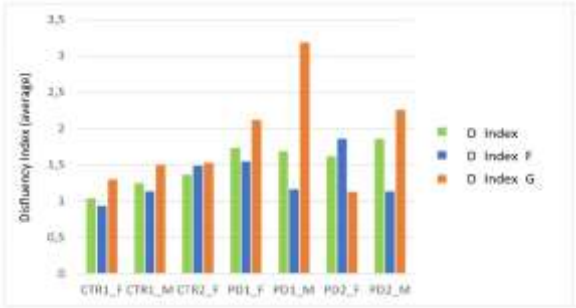


Figure 2. *Dysfluency index in Map-Tasks by 4 PDs and 3 CTRs (F= follower, G= giver).*

References

[1] Darley, F.L., Aronson, A.E., Brown, J.R. 1975. *Motor speech disorders*. Philadelphia: WB Saunders Company.

[2] Canter, G.J. 1963. Speech characteristics of patients with Parkinson's disease: Intensity, pitch and duration. *Journal of Speech Hearing Disorders* 28, 217-224.

[3] Monrad-Krohn, G.H. 1947. Dysprosody or altered melody of language. *Brain* 70, 405-415.

[4] Teston, B., & Viallet, F. 2001. L'évaluation objective de la prosodie in Les dysarthries. In P. Auzou, C. Ozsancak, V. Brun (Eds), *Problèmes en médecine*, Paris: Masson, 109-121.

[5] Duez, D. 2006. Syllable structure, syllable duration and final lengthening in parkinsonian French speech. *Journal of Multilingual Communication Disorders*, 4/1, pp.45-57.

[6] Norton, B. 2000. *Identity and language learning: gender, ethnicity and educational change*, London: Longman.

[7] Zimmerman, D.H. 1998. Identity, context and interaction. In C. Antaki, S. Widdicombe (Eds), *Identities in talk,* London: Sage, 87-106.

[8] Piacentini, V., Zuin, A., Cattaneo, D., & Schindler, A. 2011. Reliability and validity of an instrument to measure quality of life in the dysarthric speaker. *Folia Phoniatrica Logopedica* 63/6, 289-295.

[9] Anderson, A.H., Bader, M., Gurman Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S., & Weinert, R. 1991. The HCRC Map Task Corpus. *Language and Speech* 34/4, 351 -366.

[10] Savy, R. 2006. Specifiche per la trascrizione ortografica annotata dei testi raccolti. *Progetto CLIPS*, http://www.clips.unina.it/it/.

[11] Pettorino, M. & M., Giannini, A. 2005. Analisi delle disfluenze e del ritmo del dialogo romano. In F. Albano Leoni, R. Giordano (Ed), *Italiano parlato. Analisi di un dialogo*, Napoli: Liguori editore, 89-104.

The Influence of Contextual and Talker F0 Information on Fricative Perception

Orhun Uluşahin[1], Hans Rutger Bosker[1,2], James M. McQueen[1,2], Antje S. Meyer[1]
[1] Max Planck Institute for Psycholinguistics, [2] Donders Institute for Brain, Cognition and Behaviour

Speech perception is extensively influenced by contrastive acoustic context effects [1], such as the contrastive effect of fundamental frequency (F0) on the perception of voiceless fricatives' spectral center of gravity (CoG). That is, lower F0 contexts elicit a higher CoG perception and higher F0 contexts elicit a lower CoG perception [2]. However, it remains unknown whether knowledge of a talker's typical F0 profile (i.e., as opposed to context) can have similar effects. This study therefore investigated whether talker-bound F0 information can cause perceptual biases in the same contrastive direction.

In Experiment 1, female native Dutch listeners (N=10) categorized target words as the Dutch words *sok* "sock" (/sɔk/) or *sjok* "(I) trudge" (/ʃɔk/). The target words were created by replacing the original fricatives in a female native Dutch speaker's natural utterances of the words *sok* and *sjok* with tokens from an 8-step fricative continuum between /s/ and /ʃ/ (modelled on the same speaker). The target words were preceded by the carrier sentence *Nu komt het word…* "Now comes the word…" The carrier and the vowel /ɔ/ in the target words were pitch-shifted ±4 semitones to create High-F0 and Low-F0 conditions respectively, accompanied by an unshifted Mid-F0 control condition. Across 240 randomized trials containing all F0 conditions and fricative steps, participants categorized ambiguous fricatives from the synthesized fricative continuum as being more /s/-like in the Low-F0 trials compared to the High-F0 trials.

In Experiment 2, another group of participants (N = 32) listened to 20 minutes of speech from the same talker whose speech had been pitch-shifted ±4 semitones to create High-F0 and Low-F0 talker groups, respectively. After the exposure phase, participants performed a 2AFC task in which they categorized words containing fricatives from a 5-step subset (on account of ceiling effects observed in Exp. 1) of the original 8-step continuum (i.e., original steps 3-7) as *sok* or *sjok*. Crucially, in the test phase, the carrier sentence and the F0 context manipulation were removed, given the tendency of proximal context to take over talker information (e.g., [3]). Thus, participants encountered the same Mid-F0 acoustic context on each trial. Despite the lack of variability in the immediate context, participants in the Low-F0 talker group perceived the synthesized fricative continuum as being more /s/-like compared to the High-F0 talker group. This pattern persisted over a large number of trials (i.e., 160) but only became statistically robust after the first 40 trials, suggesting that participants may have needed to train themselves on the continuum. During this training period, participants from both groups overwhelmingly categorized targets as containing /ʃ/, and a global /ʃ/ bias was observed throughout the experiment, despite the subsequent divergence in response proportions.

Two further experiments (N = 32 in each) were run with methodological adjustments in an attempt to minimize the interference of the biases observed in Experiment 2. Experiment 3 was run online, and Experiment 4 was run in person. In both of these experiments, the 5-step subset of the original continuum was shifted to the original continuum steps 2-6 to sound more /s/-like, four practice trials with feedback (i.e., correct/incorrect) were introduced to the 2AFC task with original steps 1 and 8 as stronger endpoints, and breaks were removed from the 2AFC task. While these methodological changes eliminated the early /ʃ/ bias, they failed to eliminate the global /ʃ/ bias. Furthermore, neither experiment replicated the results of Experiment 2 as the observed effects were assimilatory rather than contrastive (i.e., the High-F0 talker groups perceived the continuum as being more /s/-like than the Low-F0 groups), and this effect was significant in Experiment 4.

Overall, the effect of the immediate acoustic context in Experiment 1 aligns with previous work while the effect of talker F0 remains unclear. Further research is required to establish the reliability of the talker effects, and whether they are contrastive or assimilatory in nature.

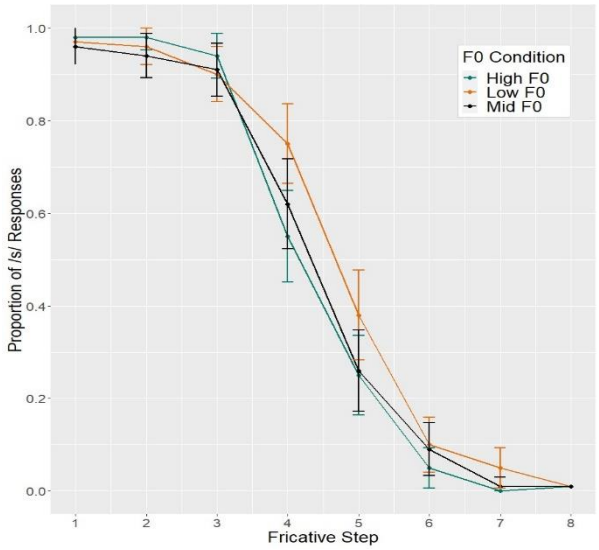


Figure 1. *Response proportions across F0 conditions and fricative steps in Experiment 1.*

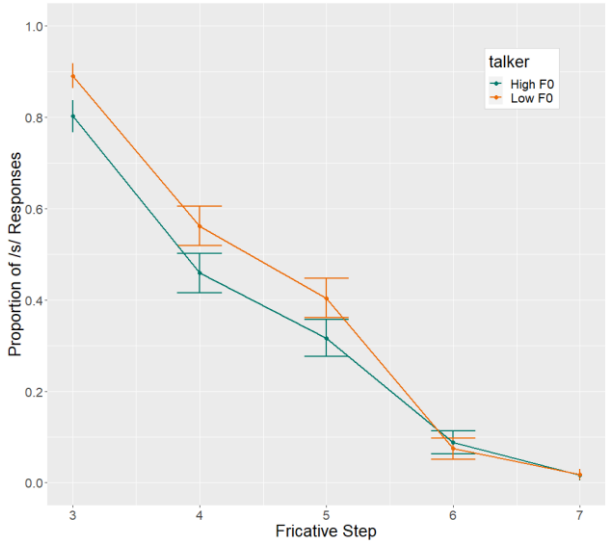


Figure 2. *Response proportions across talker groups and fricative steps in Experiment 2.*

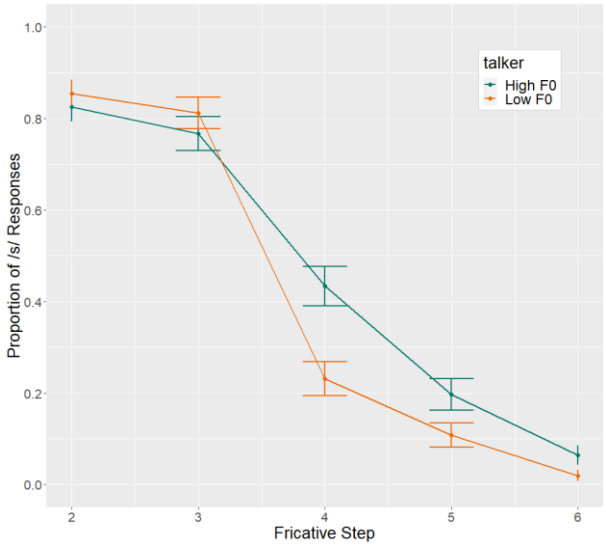


Figure 3. *Response proportions across talker groups and fricative steps in Experiment 3.*

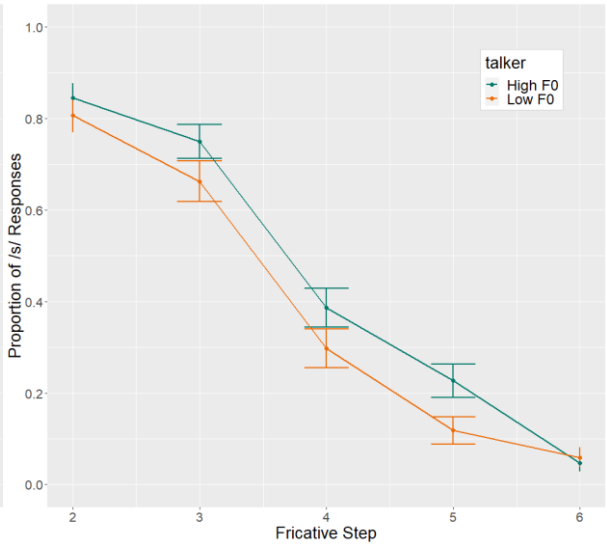


Figure 4. *Response proportions across talker groups and fricative steps in Experiment 4.*

References

[1] C. Stilp, 'Acoustic context effects in speech perception', *WIREs Cogn. Sci.*, vol. 11, no. 1, p. e1517, 2020, doi: 10.1002/wcs.1517.

[2] O. Niebuhr, 'On the perception of "segmental intonation": F0 context effects on sibilant identification in German', *EURASIP J. Audio Speech Music Process.*, vol. 2017, no. 1, p. 19, Aug. 2017, doi: 10.1186/s13636-017-0115-3.

[3] E. Reinisch, 'Speaker-specific processing and local context information: The case of speaking rate', *Appl. Psycholinguist.*, vol. 37, no. 6, pp. 1397–1415, Nov. 2016, doi: 10.1017/S0142716415000612.

Plapper – a smartphone app for self recording: Exploration of vowel spaces

Stefanie Jannedy[1] and Melanie Weirich[2]
[1]*Leibniz ZAS Berlin,* [2]*Friedrich-Schiller-Univiersity, Jena*

Criado-Perez's 2019 book *Invisible Women: Exposing Data Bias in a World Designed for Men* [1], was a painful reminder that phonetic studies, which are traditionally conducted in our laboratories, samples, at best, data from a fairly homogenous groups of volunteers rather than a cross-section of the population. Therefore, phonetics is generally unable to account for or even approximate "speech variation in the wild". With the Covid-pandemic coming into full swing in 2020 and lockdowns preventing any data-driven phonetics work, the need for a method to collect speech data remotely became inevitable. Aside from this, phoneticians have long observed their own data-gap as we often sample student populations rather than people adhering to regular work hours and then rush home to take care of their families, that work two shifts or that simple will not ever be convinced to come to our laboratory.

We now have developed a smart-device app to sample the German-speaking population in Germany. It's name is *Plapper* and it runs on Android and iOS and allows for people recording themselves in their own natural habitat without having to come to the lab thereby addressing the *observer's paradox*. The smart-device app we are using is modelled after earlier versions of a similar app deployed in Switzerland [2], the UK [3, 4] and Belgium [5].

Plapper is a tool that can be used for recordings of sentences balanced for specific target sounds (like vowels or fricatives) but also for monologues or responses to open questions. Recordings are in wav-format and have a sampling frequency of 44.1 kHz. What sets Plapper apart from other crowd-sourcing efforts is that we specifically inquire participants' social attributes like age, education, sex, gender and sexual orientation. Through this meta data, we gain insights into social variation aside from regional variation.

Due to full control over the backend, we can add different sentences for reading tasks or open questions for free recordings. While Plapper is already available in app stores, we've engaged in an extended testing phase now, gathering data from family, friends, students, colleagues and acquaintances and all of their social networks in the north and east of Germany for a proof-of-concept study. Previous work has shown that the spectral characteristics of fricative recordings done via smart device versus in the laboratory did not differ [6] (see Fig. 1: from [6]) and that the audio quality is suitable for analyses of spectral differences [5].

So far, more than 75 people from in or around Hamburg, Berlin, and Dresden have submitted their data for analyses. We are consistently downloading and semi-automatically labelling the audio recordings with *WebMaus* [7] and hand-correcting the demarcations for specific segments. Currently, we focus on vowel differences and overall vowel spaces but also particular vowel characteristics (e.g. a more fronted /u/ in eastern parts of Germany).

References

[1] Criado-Perez, C. 2020. *Unsichtbare_ Frauen._Wie_eine_von Daten beherrschte Welt die Hälfte der Bevölkerung ignoriert.* BTB Verlag.
[2] Leemann, A. (2016) *Deutschklang*. Smartphone App.
[3] Leemann, A. (2017) *English Dialects*. Smartphone App.
[4] Leemann, A., Kolly, M.-J., Purves, R., Britain, D. und E. Glaser. 2016. Crowdsourcing Language Change with smartphone applications. *PLoS ONE* 11(1): e0143060. https://doi.org/10.1371/journal.pone.0143060
[5] Gilles, P. (2019) Using crowd-sourced data to analyse the ongoing merger of [ɛ] and [ʃ] in Luxembourgish. Proceedings of the 19th International Congress of Phonetic Sciences Melbourne, Australia.

[6] Jannedy, S., Weirich, M. & Leemann, A. (2018) The ecological validity of crowd-sourced data. Talk presented at Phonetik & Phonologie in deutschsprachigen Ländern 14, Univ. of Vienna. 6.-7.09.2018.
[7] Kisler, T. and Reichel U. D. and Schiel, F. (2017) Multilingual processing of speech via web services, *Computer Speech & Language*, Vol. 45, pp. 326–347.

(1)

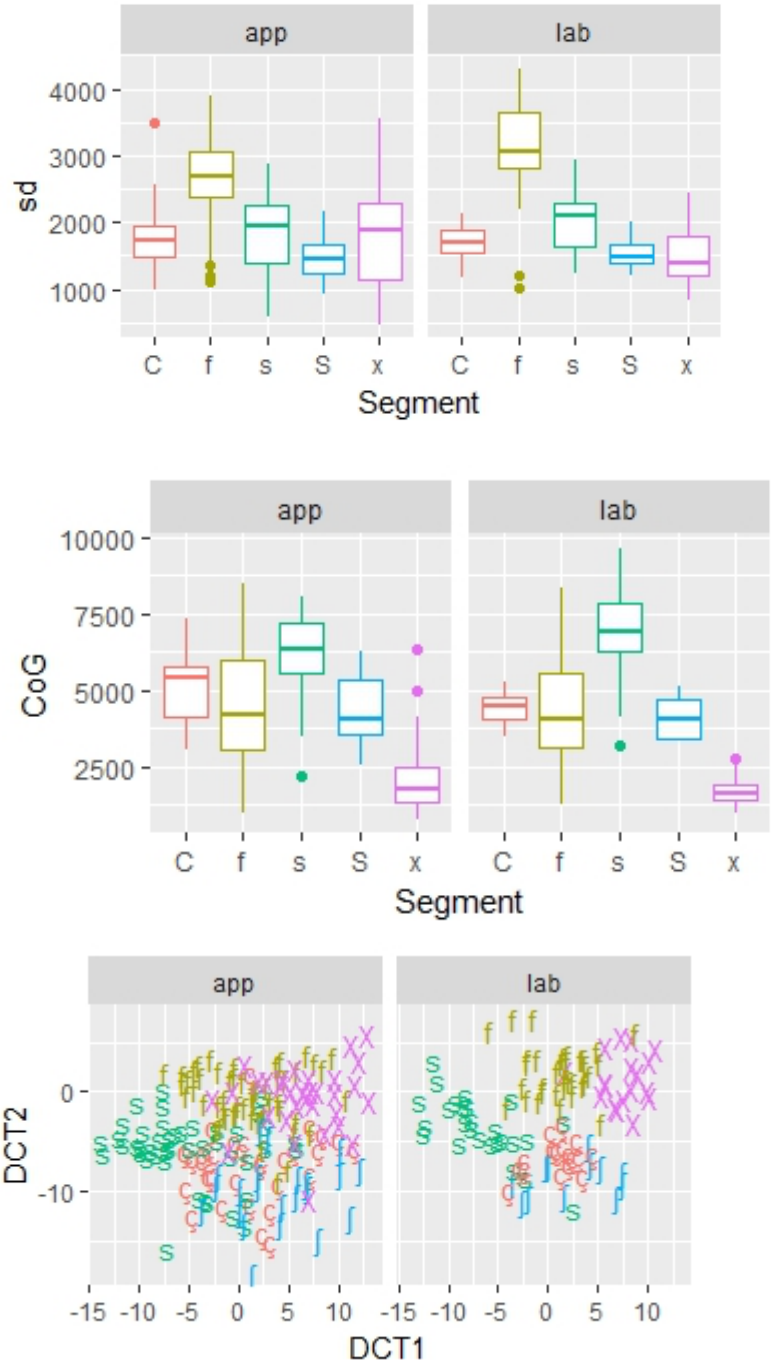


Figure 1. *Comparison of spectral characteristics (Spectral moments: Standard deviation (sd), Center of Gravity (CoG) and Discrete Cosine Transformation Coefficients (DCT1, DCT2)), of the five German fricatives /ç f s ʃ χ/ of data recorded via smart devices (left) versus in the laboratory (right). Data from [2], analyses by [6].*

Beat gestures can drive recalibration of lexical stress perception

Ronny Bujok[1], David Peeters[1,2], Antje Meyer[1], and Hans Rutger Bosker[1,3]
[1]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands,*
[2]*TiCC Tilburg University, Tilburg, The Netherlands,*
[3]*Donders Institute, Radboud University, Nijmegen, The Netherlands*

The acoustic realization of speech can vary dramatically between speakers and situations, potentially making it difficult to interpret it unimodally. Listeners may therefore use visual cues to help them disambiguate the auditory signal. For instance, when an ambiguous sound /a?a/ midway between /aba/ and /ada/ is presented to a listener together with a video of a talker producing either a visual /aba/ or /ada/, listeners – over time – learn to perceive the ambiguous sound as the sound indicated by the visual articulatory cues [1]. That is, participants who heard ambiguous /a?a/ paired with visual /aba/ categorized a subsequently presented audio-only /aba – ada/ continuum as more /aba/-like. Conversely, when /a?a/ was paired with visual /ada/, listeners were more likely to perceive /ada/. Listeners are thus capable of adjusting their perceptual boundaries based on visual information, which leads to long lasting changes in their auditory perception. This effect is called visually-guided *recalibration*.

In this study we tested if manual beat gestures can also recalibrate listeners' perceptual categories of suprasegmental aspects of speech, specifically lexical stress. Beat gestures, simple up-and-down movements of the hand, are usually aligned to stressed syllables. As such their alignment can influence online stress perception [2]. That is, if an ambiguously stressed word is paired with a video of a talker producing a beat gesture, listeners are more likely to perceive stress on the syllable indicated by the beat gesture. However, it is unclear if this beat gesture effect has long-lasting consequences for perception. In other words, can beat gestures recalibrate perception of lexical stress?

We tested 80 participants using a recalibration paradigm, including audiovisual exposure and an audio-only test phase. In exposure, participants were repeatedly presented with an ambiguously stressed /ka.nɔn/ midway between Dutch *CAnon* [strong-weak (SW); "canon"] and *kaNON* [weak-strong (WS); "cannon"]. Ambiguously stressed items were created by F0 interpolation, while duration and intensity were kept constant and average values. Critically, ambiguous /ka.nɔn/ was disambiguated by a beat gesture on either the first (SW-bias group) or second syllable (WS-bias group). Participants were expected to learn during passive viewing/listening in the exposure phase that the ambiguous stress cues on /ka.nɔn/ indicated either initial (SW-bias group) or final stress (WS-bias group). Participants were then presented with an audio-only test phase with two blocks (order counterbalanced between participants): segmental overlap and generalization. In the segmental overlap block, participants were asked to categorize five ambiguous tokens, each presented 15 times, from the *CAnon – kaNON* continuum (based on F0 manipulation) as either SW or WS using a 2AFC task. In the generalization block, participants categorized a continuum of a novel word pair: *VOORnaam* [SW; "first name"] – *voorNAAM* [WS; "respectable"]. Results from GLMMs indicated that participants from the SW-bias group gave more SW responses across the entire continuum in both blocks compared to the WS-bias group (Figure 1).

This study suggests that visual beat gestures can have a lasting effect on subsequent audio-only perception, which we interpret as recalibration. Moreover, this effect generalizes to novel items, which is considered evidence for abstraction of phonological units [3]. We conclude that listeners can use the visual modality and the timing of seemingly meaningless hand gestures to adapt to suprasegmental variability in speech.

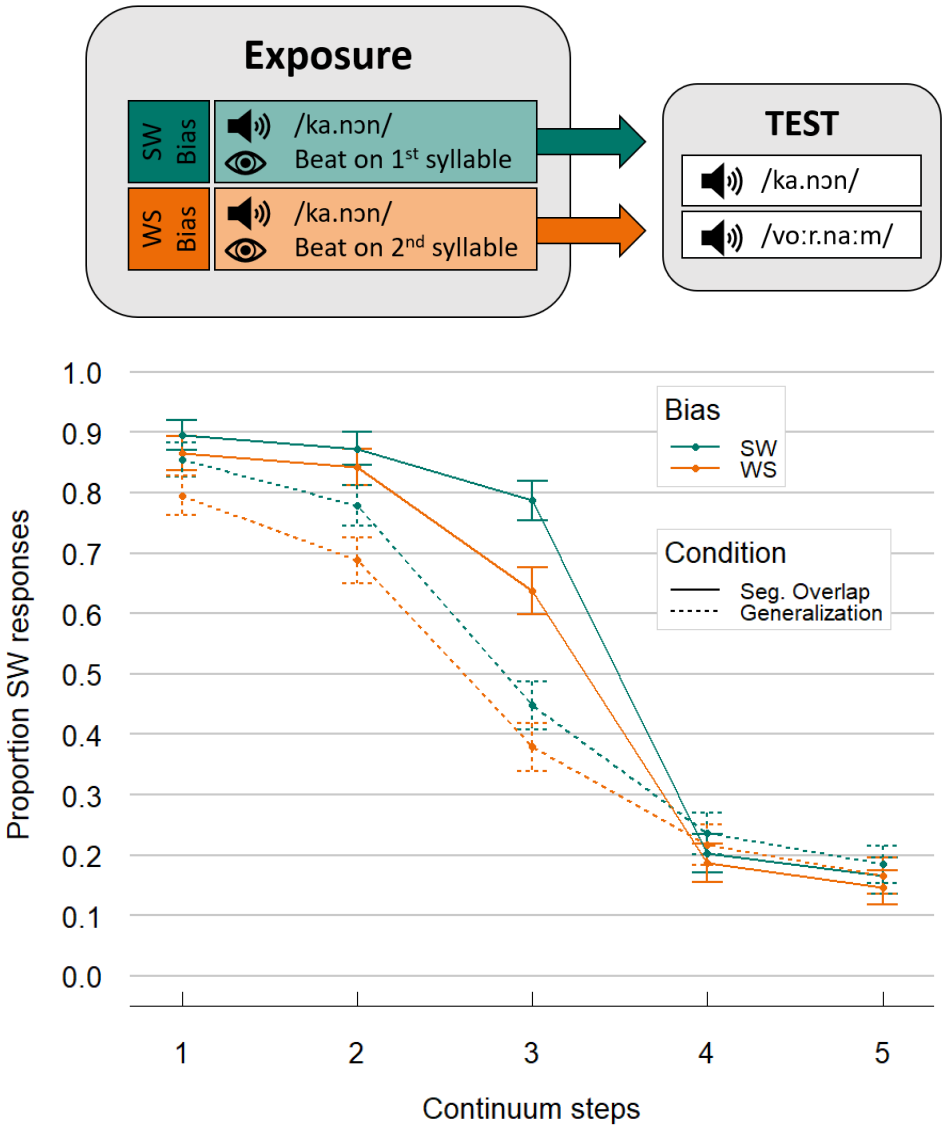


Figure 1. *Design and results*. In the exposure phase, the SW-bias group (beat on 1st syllable) learned that the ambiguous /ka.nɔn/ indicated initial stress. The WS-bias group (beat on 2nd syllable) learned the same auditory token indicated final stress. In the subsequent audio-only test phase participants from the SW-bias group perceived the CAnon – kaNON (seg. overlap) and VOORnaam – voorNAAM (generalization) continua as more SW-like, while the WS-bias group perceived the same continua as more WS-like.

References

[1] P. Bertelson, J. Vroomen, and B. de Gelder, "Visual Recalibration of Auditory Speech Identification: A McGurk Aftereffect," *Psychol Sci*, vol. 14, no. 6, pp. 592–597, Nov. 2003, doi: 10.1046/j.0956-7976.2003.psci_1470.x.

[2] R. Bujok, A. Meyer, and H. R. Bosker, "Audiovisual Perception of Lexical Stress: Beat Gestures are stronger Visual Cues for Lexical Stress than visible Articulatory Cues on the Face," PsyArXiv, preprint, May 2022. doi: 10.31234/osf.io/y9jck.

[3] J. M. McQueen, A. Cutler, and D. Norris, "Phonological Abstraction in the Mental Lexicon," *Cognitive Science*, vol. 30, no. 6, pp. 1113–1126, 2006, doi: https://doi.org/10.1207/s15516709cog0000_79.

Characteristics of boundary marking for parenthesis - Relationship between prosodic boundary marking and breathing patterns

Valéria Krepsz
[1]*Humboldt-Universität zu Berlin,* [2]*Hungarian Research Centre for Linguistics*

According to a group of previous studies on the prosodic characteristics of parenthesis (cf. [1], [2]), linguistic elements embedded into matrix sentences generate new intonational phrases (IP), namely, they split the original sentence into two IPs. However, other studies suggest that different prosodic strategies can also be observed in the case of embedded clauses. In some examples, the inserted IP may be built in as a unit of the tone group of the first clause [3]. Peters [4] analyzed the realization of different syntactic parentheticals in spontaneous speech. The results corroborated that the embedding speech unit does not necessarily create a separate IP, moreover, although the parenthesis can form its own IP, it does not split the original phrase into two separate IPs in all cases. In addition, respiration pattern is closely related to the length and syntactic structure of the intended speech production (cf. [5]). Previous research has shown that inhalation mostly occurs at the syntactic boundary [6], but the synchronicity of speech and breath planning and production is still quite unclear. The previous research mostly focused on English and German, but not for Hungarian, however its prosody is typologically different from the prosody of the other languages, e.g. information structure in Hungarian is primarily expressed by word order, and the position of focus is defined syntactically. The question arises as to how the prosodic structure and the breathing patterns of the parenthesis interrelate, and what kind of correlation can be found with the length and syntactic structure of the speech units. The aim of the study is therefore to investigate how prosodic boundary markers and breathing circles are realized in shorter and longer embedded sentences of different structures.

The material consisted of 16 sentences (3 times repetitions) read out by 10 Hungarian speakers. The original sentences consisted of 2 clauses, to which an embedding one was added. The sentences are built up of clauses with different lengths in different combinations: long (L, 21±1 syll) - short (S; 7±1 syll); S – L; S – L – L, etc. The recordings were annotated at sentence and IP level, and the last syllable before the boundaries was marked using Praat [7]. To examine prosodic boundary marking, the occurrence and duration of silent pauses and creaky voice, the variation in f0 and intensity values, and the appearance of phrase-final lengthening at the boundary of clauses and sentences were analyzed. Respiration was measured using the RespTrack device that consisted of two belts attached to the abdomen and chest to record the timing pattern and amplitude of inhalation and exhalation in synchrony with speech. These features were focusing on the possible boundaries within the sentences. The statistical analysis was conducted using R program [8].

The results showed considerable heterogeneity depending on the structure of the sentence and the individual strategies of the speakers. The results corroborated the effect of the length and syntactic structure of the clauses: the occurrence of the prosodic boundary markers are less frequent and showed less divergence from realization in non-boundary position after the shorter clauses than after the longer ones, i.e. shorter sentences formed a discrete IP less often than longer sentences. The most systematic prosodic boundary marker was the occurrence of a silent pause, followed by phrase final lengthening and changes in f0, while the decrease in intensity appeared mainly in the sentence-end position. The prosodic marking of the boundaries accompanied by breathing was also more prominent. In addition to the effect of the syntactic structure and length of the embedding clauses, there were also significant individual differences in the realization of the sentences both in terms of prosody and breathing patterns.

The results provide more information on the interdependence between the syntactic and prosodic structure of communication and the prosodic and respiratory planning of speech production, and a basis for comparison with other types of languages.

References

[1] Selkirk, E.O. 1980. *On prosodic structure and its relation to syntactic structure*. Bloomington.

[2] Nespor, M. & Vogel, I. 1986. *Prosodic phonology*. Foris.

[3] Armstrong, L.E. & Ward, I. C. 1926. *Handbook of English intonation*. Heffer & Sons.

[4] Peters, J. 2006. Syntactic and prosodic parenthesis. *Proceedings of International Conference on Speech Prosody 2006* (Dresden, Germany).

[5] Rochet-Capellan, A. & Fuchs, S. 2013. The interplay of linguistic structure and breathing in German spontaneous speech. *Proceedings of Interspeech 2013* (Lyon, France).

[6] Wang Y.T., Green J.R., Nip I.S., Kent R.D. & Kent J.F. 2010. Breath group analysis for reading and spontaneous speech in healthy adults. *Folia Phoniatr. Logopaedica* 62, 297-302.

[7] Fuchs, S., Petrone, C., Krivokapic, J. & Hoole, P. 2013. Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41, 29-47.

[8] Boersma, P. & Weenink, D. 2021. *Praat: doing phonetics by computer*. Computer program.

Response tokens in the lab and in the wild: Evidence from task-based and spontaneous conversations

Alicia Janz, Simon Wehrle & Martine Grice
IfL Phonetik – University of Cologne

Interactive feedback is one of the main mechanisms in the coordination of speaker turns in human dialogue. Verbal response tokens produced by the listener, indicating passive recipiency (backchannels) or incipient speakership, are known to play an important role in the coordination of who speaks when [1]. They have been investigated in several studies, especially in task-oriented settings [1, 2]. However, the conversational demands of structured, task-oriented communication may differ from those of naturally occurring free conversation, in which response tokens may be required to serve a wider range of functions [3].

Previous studies on German task-oriented dialogues have reported a predominance of the response tokens *ja, okay, mmhm* (mostly with rising intonation) and *genau* (mostly with falling intonation) [4, 5]. In this study we investigate the lexical choice and prosodic realisation (pitch movements in semitones) of passive recipiency (PR) and incipient speakership (IS) tokens in two conversational settings: (1) spontaneous face-to-face conversations and (2) task-oriented (Maptask [6]) conversations without visual contact. We recorded 10 German speakers in pairs in each of the two settings.

In both settings speakers used similar proportions of PR and IS tokens (task-oriented: 83.8% (PR), 16.2% (IS), spontaneous: 76.7% (PR), 23.3% (IS)). For indicating passive recipiency, speakers used predominantly the standard response tokens *ja* and *mmhm*. However, in the spontaneous conversations, the proportion of *other* (non-standard) response tokens was much higher. For indicating incipient speakership in the Maptask, speakers showed a preference for different lexical tokens such as *ja, okay,* and *genau*, whilst in the spontaneous conversations the vast majority of response tokens were of the *ja* type (see Figure 1). In terms of prosodic realisation, most PR tokens in task-oriented conversations were produced with a rising intonation contour, while most PR tokens in spontaneous conversations carried a level or falling intonation contour. To indicate incipient speakership, speakers used mostly falling and level contours in both conversational settings (see Figure 2). In sum, we found similar proportions of passive recipiency and incipient speakership tokens in task-oriented and spontaneous conversations. However, while in task-oriented conversations the tokens' functions are distinguished on the lexical and the prosodic level, speakers seemed to make less of a clear distinction between the two types of response tokens in the spontaneous conversational setting.

Since social communication is an inherently multimodal process [7], in which non-verbal signals, such as the speaker's eye-gaze have been shown to substantially influence the coordination of speaker turns [8], we speculate that while our speakers made extensive use of the speech channel in the task-oriented dialogues (where the visual channel is unavailable), they use this channel to a lesser extent in naturalistic dialogue where cues on the visual channel are also available. Our findings emphasise the importance of considering different conversational settings when investigating conversation. Additionally, they point us in new directions, such as the multimodal investigation of feedback in conversation, taking into account non-verbal channels such as eye gaze and gesture as well [9].

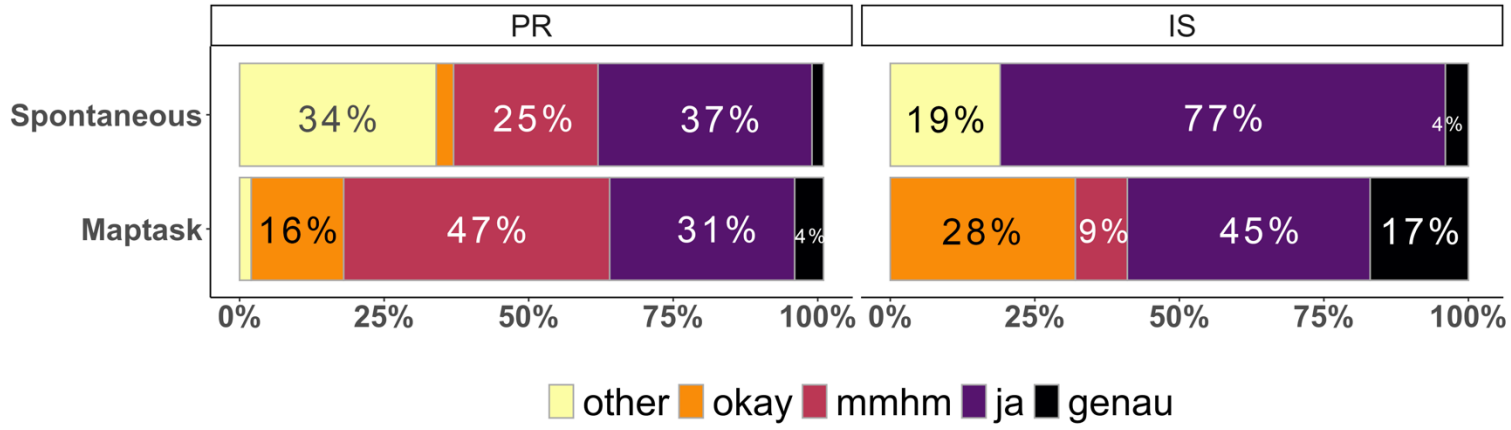


Figure 1. *Proportions of response token types and functions (PR: passive recipiency and IS: incipient speakership) in spontaneous and task-oriented conversational settings.*

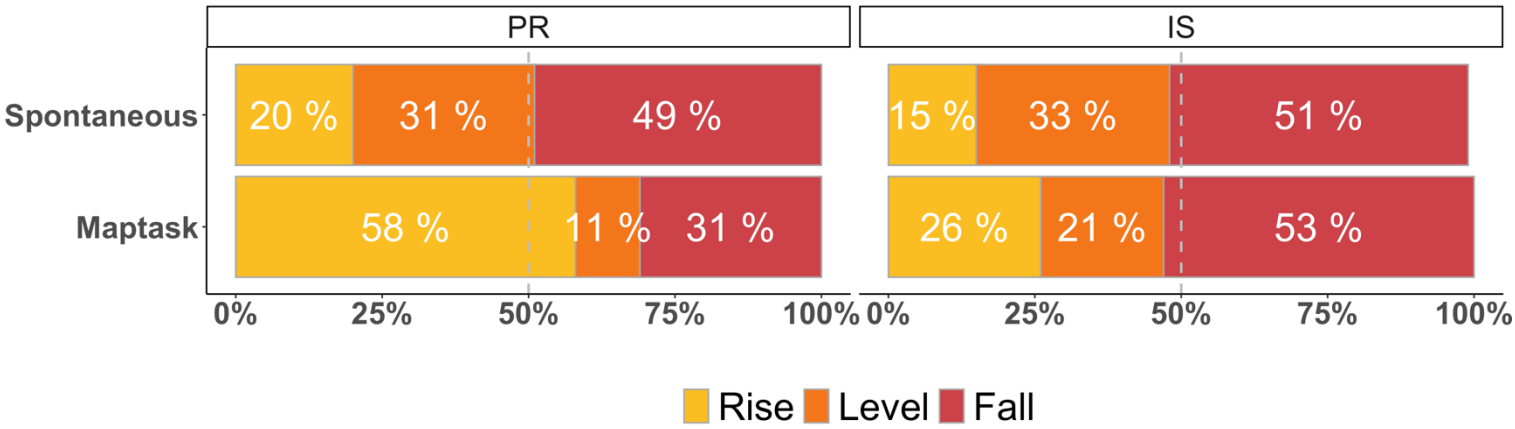


Figure 2. *Proportions of intonation contours indicating passive recipiency and incipient speakership in spontaneous and task-oriented conversational settings.*

References

[1] Bangerter, A. & Clark H. H. (2003). Navigating joint projects with dialogue. *Cog. Sci.*, 27(2), 195-225

[2] Savino, M. (2011). The intonation of backchannel tokens in Italian collaborative dialogues. *Language and Technology Conference*, 28-39

[3] Fusaroli, R. et al. (2017). Measures and mechanisms of common ground: Backchannels, conversational repair, and interactive alignment in free and task-oriented social-interactions. *Proc. CogSci 2017*, 2055-2060

[4] Wehrle, S. (2022). *A Multi-Dimensional Analysis of Conversation and Intonation in Autism Spectrum Disorder.* PhD Dissertation

[5] 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,* 734-738

[6] Anderson, A., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., ... Miller, J. (1991). The HCRC Map Task corpus. Language and Speech, 34(4), 351–366.

[7] Hadley L. V., Naylor G., Hamilton A. F. D. C (2022). A review of theories and methods in the science of face-to-face social interaction. *Nature Reviews Psychology,* 1(1), 42-54

[8] Degutyte Z., Astell A. (2021). The Role of Eye Gaze in Regulating Turn Taking in Conversations: A Systematized Review of Methods and Findings. *Front. Psychol.* 12, 616471

[9] Anonymous

Phonetic Variation and Syllabic Structures in Italian Connected Speech

Loredana Schettino and Francesco Cutugno[1]
[1]*University of Naples Federico II*

In spontaneous conditions, speakers' utterances consist of continuous chains of words. In the economy of production, segments tend to be under-specified in the speech signal and can undergo variation ranging from subtle weakening, vowel centralization or consonant lenition up to to deletion of segments or even of multiple syllables [1]. This kind of acoustic ``reduction'' has been described as a common feature in spoken languages [2] which may result from speakers' production accuracy in specific communicative situations [3] and linguistic factors, like prosodic features, e.g., lexical stress [4], lexical category, the discursive function [1]. In particular, [4] provided evidence that, in Italian, vowel centralization in unstressed syllables represents a structural feature independent of diaphasic and diatopic variation. This study aims at building upon this by investigating the phonetic variation patterns that may be observed in the speech chain with relation to specific linguistic structures, i.e., syllabic structures and stress.

Given the evidence on the role of syllables as basic units of speech production and perception [2, 5], we investigate reduction phenomena in connected speech by comparing the syllables sequence expected at the phonological level to that observed at phonetic level. The data consists in Italian spoken narrative texts, "frog stories", produced by 11 university students (Nocando corpus [6]). The phonological and phonetic annotations were obtained using the WebMAUS Basic services [7], manually edited in Praat [8] and syllabified according to the principles of sonority sequencing and onset maximization [9]. The alignment between the sequences of phonological and phonetic was evaluated using SCLITE, a tool included in the Speech Recognition Scoring Toolkit (SCTK) provided by the NIST. The evaluation output reports as "Deletion" (D) cases of phonological syllable without correspondent at the phonetic level; "Substitutions" (S) phonetic syllables that differ from the corresponding phonological one; "Correct" (C), case of equivalence between phonetic and phonologic syllables. The phonological syllables were also annotated for their structure (V, C) and contextual saliency (lexical stress). For Substitutions, the syllabic position subjected to variation is also considered (Onset, Nucleus, Coda). To evaluate the role of lexical stress and syllabic structure on phonetic variation and control for individual variability Generalized Linear Mixed Models were fitted. First, with the evaluation levels as binomial responses, Syllabic Structure and Lexical Stress as fixed effects, and Speaker as a random effect; then, with the Substitutions subset, the syllabic position subjected to variation as binomial responses, Lexical Stress as fixed effect.

The analysis is based on 940 syllables at the phonological level. 67% are phonetically realized as expected (C), 24% are subjected to variation (S), 9% have been deleted. As expected [4], reduction processes mostly, and significantly, concern unstressed syllables, fig1.a. However, we found that while unstressed syllables are more likely to be deleted rather than stressed ones, both can be almost equally subjected to substitution. In fact, systematic relations emerged between the stress condition and the position of the change in the syllable. Namely, lexical stress prevents variation of the syllables' nuclei, i.e., deletion or change like centralization, which concerns unstressed syllables instead, but allows for onset changes (such as lenition or assimilation phenomena). As for the syllabic structure, fig. 1.b. CV represents the most frequent structure [9] and the most stable and resistant to variation. Instead, V and VC structures are most prone to deletion and to restructuring processes in the speech chain whereas complex structures (CCV, CVC) tend to simplification. This finding seems in line with Greenberg's observation that syllable onsets are generally preserved while coda or nuclear constituents are more frequently subjected to underspecification [5]. This study allowed for the description of reduction processes related to linguistic structures which are, to a certain extent, independent from sociolinguistic factors, which contributes to improve our insight on speech production and comprehension mechanisms in spontaneous communication.

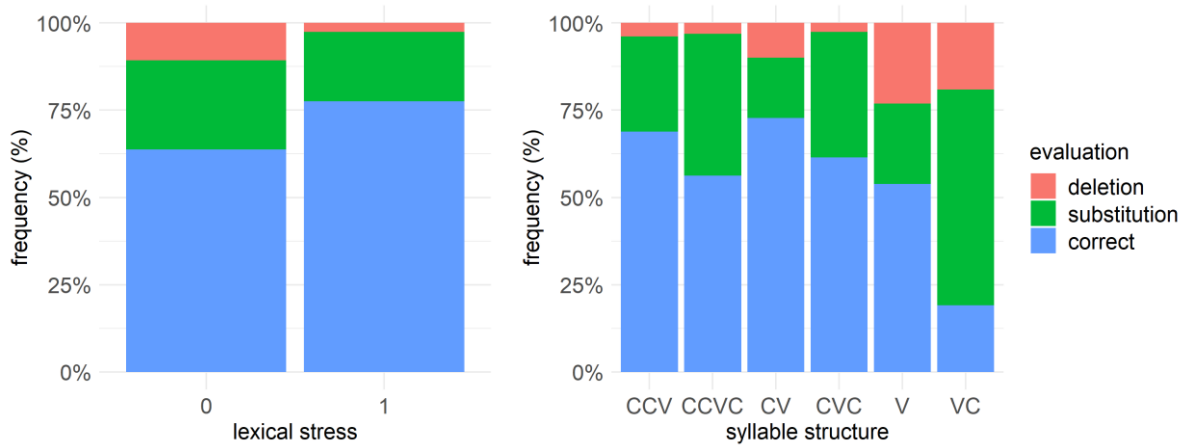


Figure 1. *Frequency (%) of the evaluation output cases per lexical stress condition (1 = stressed, 0 = unstressed, 1.a, left) and per syllabic structure (1.b, right).*

Dataset	Response	Fixed Effect	Estimate	SE	Pr(>|z|)
All syll	Deletion	stress	-1.56	0.40	0.40
"	Substitution	stress	-0.33	0.18	0.055
"	Substitution	CV	-0.63	0.28	0.023
"	Deletion	V	2.43	0.92	0.008
"	Deletion	VC	1.68	0.81	0.039
Substituted syll	Nucleus_substitution	stress	-1.13,	0.40	0.005
Substituted syll	Onset_substitution	stress	0.81,	0.32	0.012

Table 1. *Relevant results from the statistical analysis*.

References

[1] Ernestus, M. & Warner, N. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39, 253-260.

[2] Cangemi, F. & Niebuhr, O. 2018. Rethinking reduction and canonical forms. In Cangemi, F, Clayards, M., Niebuhr, O., Schuppler, B. & Zellers, M. (Eds.), *Rethinking reduction*, Berlin: De Gruyter Mouton, 277-302.

[3] Ernestus, M., Hanique, I. & Verboom, E. 2015. The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics* 48, 60-75.

[4] Savy, R. & Cutugno, F. 1998. Hypospeech, vowel reduction, centralization: how do they interact in diaphasic variations. *Proceedings of the XVIth International Congress of Linguists* (Paris, France).

[5] Greenberg, S. 1999. Speaking in shorthand – a syllable-centric perspective for understanding pronunciation variation. *Speech Communication* 29(2-4), 159-176.

[6] Brunetti, L., Bott, S., Costa, J. & Vallduví, E. 2011. A multilingual annotated corpus for the study of information structure. Proceedings of the Grammatik und Korpora 2009. Dritte internationale Konferenz (Mannheim, Germany).

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

[8] Boersma, P., & Weenink, D. 1999-2022. *Praat: doing phonetics by computer*. Computer program.

[9] Nespor, M. 1993. *Fonologia*. Bologna: Il Mulino, 1993.

Individual differences in child-directed prosody and gestures in broadcasting: The role of empathy

Yanran Zhang, Communication University of Zhejiang
Yan Gu, University of Essex

Child-directed language (CDL) is a special multimodal communicative behaviour that differs from adult-directed language (ADL) in many aspects (e.g., prosody; gestures) (Campisi & Ozyurek, 2013; Han et al., 2022; Hoff & Naigles, 2002; Perniss et al., 2018). Despite a general tendency to find infant-directed speech across cultures (Cox et al., 2022), not everyone makes adaptations when talking to children. The degree of prosodic adjustments between ADL and CDL may be affected by individuals' sensitivity and empathy levels (e.g., Kempe, 2009; Leerkes et al., 2009; Spinelli & Mesman, 2018). For example, Kempe (2009) found that empathy level has a positive correlation with adult-directed pitch, but that there is no such correlation when addressing infants. However, the experiment was only based on six given sentences, questioning the generalizability of the results. In addition, empathy levels predict gesture frequency and saliency (Chu et al., 2014), but no research has investigated the influence of empathy on individuals' gesture adaptation for CDL. Furthermore, the setting of CDL is not confined to parent-child dyads - we know little about individual differences in prosodic and gesture adaptation in TV broadcasting, a distinctive non-caregiver context where broadcasters seriously care about their imagined audience. If empathy levels relate to language production, broadcasters who care more about their audiences should adjust their language more than broadcasters who care less. However, it is entirely unknown how empathy level will affect the prosodic and gestural adjustments in CDL.

The present study aimed to understand individual differences in adaptation between adult-directed and child-directed broadcasting. Forty-six future broadcasters were required to do a live broadcast explaining pictures both in a regular adults' programme and a children's programme (counterbalancing sequences). We examined whether participants' empathy affected their degree of adjustment between the two programmes. To avoid gender differences, participants were all female.

The multimodal behaviours were annotated and analyzed in prosodic features (e.g., pitch, intensity, speaking rate) of 4916 utterances, and manual features (e.g., types, saliency, frequency) of 10843 gestures. The empathy levels of participants were coded through the Empathy Quotient questionnaire (Baron-Cohen & Wheelwright, 2004) with 60 questions and a total score of 80. The mean score of empathy level was 43.39 (SD=9.68, range 22-66). Results showed that there were interactions between participants' empathy and broadcasting programmes: Compared to participants with lower empathy, those with a higher level of empathy adjusted significantly more in mean pitch, speaking rate, intensity, overall gesture saliency, frequency of representational and beat gestures between programmes. Specifically, higher-empathetic participants talked faster and louder with more variation in pitch, gestured with a larger size, and used more representational but fewer beat gestures in the child-directed programme, whereas lower-empathetic participants had a reversed pattern or displayed no differences. Additionally, regardless of programmes, participants with a higher empathy level used more representational but fewer pragmatic gestures. In sum, child-directed broadcasting has more audiovisual adjustment, but individuals' empathy levels play a critical role in shaping the degree of adjustment between child-directed prosody and gestures.

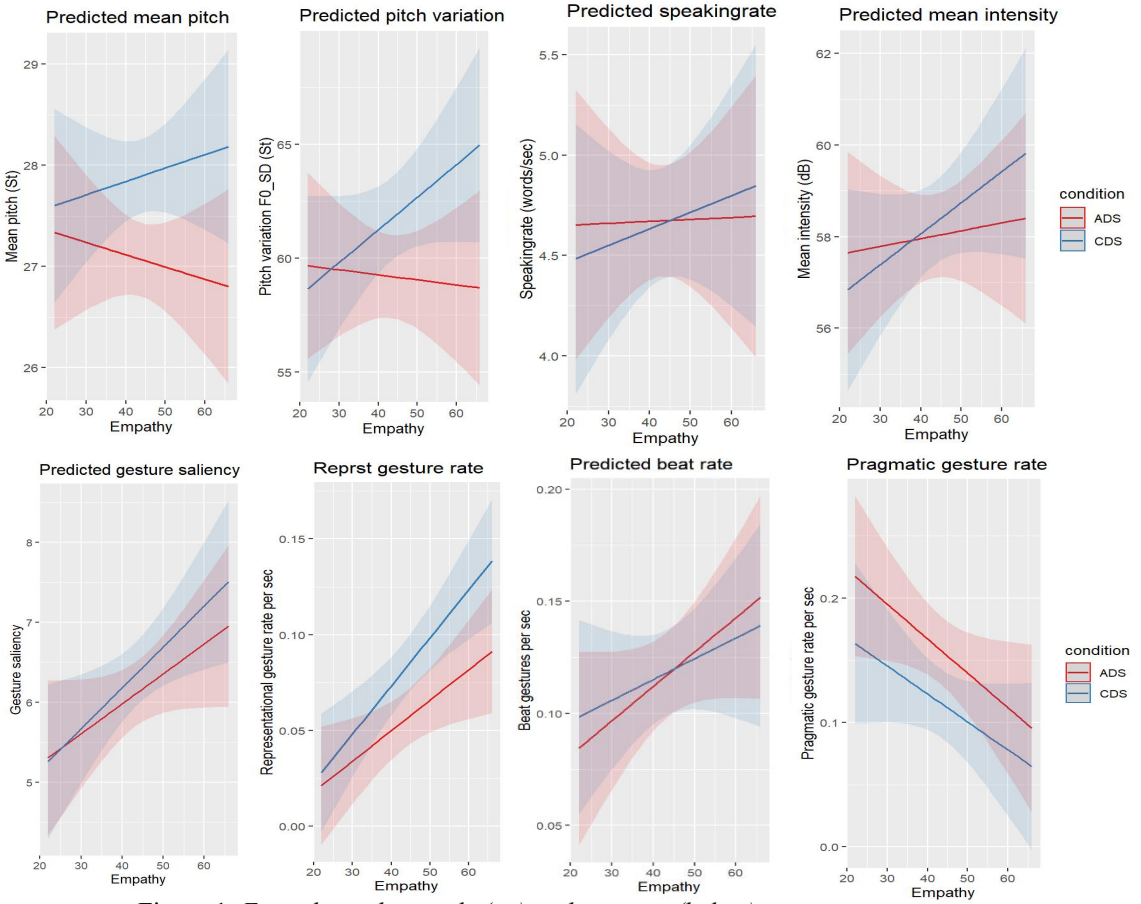


Figure 1. *Empathy and prosody (up) and gestures (below) in CD and AD- broadcasting*

References

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*(2), 163-175.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14-27.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*(2), 694-709.

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour,* 1-20.

Han, M., De Jong, N. H., & Kager, R. (2022). Prosodic input and children's word learning in infant- and adult-directed speech. *Infant Behavior and Development*, *68*, 101728.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Dev.*, *73*, 418–433.

Kempe, V. (2009). Child-directed speech prosody in adolescents: Relationship to 2D:4D, empathy, and attitudes towards children. *Personality and Individual Differences*, *47*(6), 610-615.

Leerkes, E. M., Blankson, A. N., & O'Brien, M. (2009). Differential effects of maternal sensitivity to infant distress and nondistress on social-emotional functioning. *Child Development*, *80*, 762-775.

Perniss, P., Lu, J. C., Morgan, G., & Vigliocco, G. (2017). Mapping language to the world: the role of iconicity in the sign language input. *Developmental Science*, *21*(2), e12551.

Spinelli, M., & Mesman, J. (2018). The regulation of infant negative emotions: The role of maternal sensitivity and infant-directed speech prosody. *Infancy*, *23*(4), 502-518.

Sound symbolic interactions of cuteness and size in German long vowels

Dominic Schmitz
Heinrich Heine University Düsseldorf, Germany

When certain sounds combined with further sensory information become meaningful, one speaks of sound symbolism. One of the most well-researched types of sound symbolism is "size sound symbolism": Some speech sounds, e.g. /i/, are associated with smallness, while other speech sounds, e.g. /a/, are associated with bigness [1, 2]. While there is a rather large body of research concerned with size sound symbolism itself, there is lack of research connecting size to other dimensions of the visual domain. The present study aims to deliver first results to fill this research gap.

While the dimension of size was investigated in a multitude of studies on sound symbolism during the last decades [3, 4], another visual dimension was rarely considered: cuteness. Cuteness can be understood as a more complex form of simple geometric shape, as was investigated in previous research [5, 6]. Cuteness, especially from its biological perspective as comprised in the so-called "baby schema" [7], is a fundamental feature of human perception and correlates, among other things, with size [8]. Research on Japanese has shown that cuteness is also found as sensory information to be combined with speech sounds [9].

Taking into account both size and cuteness, the present study aimed at establishing a relation from "small" to "big" and from "not cute" to "cute" for long vowels of Standard German (i.e. /aː, ɛː, eː, iː, oː, øː, uː, yː/), providing further insight into the multimodal nature of sound symbolism.

Two online forced-choice tasks (pilot study with 21 participants; main study with 80 participants) were conducted using OpenSesame [10]. Disyllabic pseudowords were used as auditory stimuli, controlling for potentially confounding lexical [11] and contextual [12, 13] effects. In either syllable, stimuli's nuclei consisted of one of the vowels under investigation. The simplex onsets of the open syllables consisted of one consonant, i.e. /d, f, j, k/ or /r/. In total, 96 pseudowords were used, i.e. 12 per vowel. Images of phantasy creatures [14] were used as visual stimuli. In each trial, participants were shown five differently sized versions of a randomly chosen creature. The participants' task was to decide which image version matched the audio stimulus of a trial best. As cuteness judgements likely differ by participants, afterwards participants were again shown all creature images to judge them for their cuteness on a five point scale.

The size response then entered a generalised additive mixed model regression analysis as dependent variable. Cuteness judgments, vowel, onset consonant types and phonological neighbourhood density were introduced as independent variables, while participant ID and age were included as random effects. Overall, /aː/ was found to be bigger than all other vowels, while /iː, yː/ were found to be smallest. Cuteness judgements did not show a significant effect on their own. However, having vowel quality and cuteness judgements interact, it was found that the size of the open vowel /aː/ increased with cuteness, while the size of the close vowels /iː, yː/ further decreased. Results were consistent across both the pilot and the main study.

The present findings demonstrate that cuteness modifies the effect of size sound symbolism. With increasing cuteness, the vowel considered to be biggest is judged to be even bigger, while the vowels considered to be smallest are judged to be even smaller. It appears that sound symbolistic effects manifest in an intricate interaction when multiple visual dimensions are considered. The present findings contribute to the growing body of evidence for and the nature of sound symbolism and call for the incorporation of multiple dimensions into analyses.

References

[1] Tarte, R. D. (1982). The relationship between monosyllables and pure tones: An investigation of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior*. Academic Press 21(3). 352–360. https://doi.org/10.1016/S0022-5371.

[2] Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*. Springer US 7(1). 5562. https://doi.org/10.1038/s41598-017-05965-y.

[3] Berlin, B. (1995). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. *Sound Symbolism*. Cambridge University Press 76–93. https://doi.org/10.1017/CBO9780511751806.006.

[4] Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P.F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences 113(39). 10818–10823. https://doi.org/10.1073/PNAS.1605782113.

[5] Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language*. Academic Press 99. 122–150. https://doi.org/10.1016/J.JML.2017.09.006.

[6] Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition*. Elsevier 126(2). 165–172. https://doi.org/10.1016/J.COGNITION.2012.09.007.

[7] Lehmann, V., Huis in't Veld, E. M. J., & Vingerhoets, A. J.J.M.. (2013). The human and animal baby schema effect: Correlates of individual differences. *Behavioural Processes*. Elsevier 94. 99–108. https://doi.org/10.1016/j.beproc.2013.01.001.

[8] Kringelbach, M. L., Stark, E. A., Catherine, A., Bornstein, M. H., & Stein, A. (2016). On Cuteness: Unlocking the Parental Brain and Beyond. *Trends in Cognitive Sciences* 20(7). 545–558. https://doi.org/10.1016/j.tics.2016.05.003.

[9] Kumagai, G. (2019). A sound-symbolic alternation to express cuteness and the orthographic Lyman's Law in Japanese. *Journal of Japanese Linguistics*. Walter de Gruyter GmbH 35(1). 39–74. https://doi.org/10.1515/JJL-2019-2004.

[10] Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*. Springer 44(2). 314–324. https://doi.org/10.3758/s13428-011-0168-7.

[11] Caselli, N. K., Caselli, M. K. & Cohen-Goldberg, A. M. (2016). Inflected words in production: Evidence for a morphologically rich lexicon. *Quarterly Journal of Experimental Psychology* 69(3). 432–454. https://doi.org/10.1080/17470218.2015.1054847.

[12] Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59(5). 1208. https://doi.org/10.1121/1.380986.

[13] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the cicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717. https://doi.org/10.1121/1.402450.

[14] van de Vijver, R., & Baer-Henney, D. (2014). Developing biases. *Frontiers in Psychology* 5. https://doi.org/10.3389/fpsyg.2014.00634.

Beyond WEIRDY: Learning of nonnative speech sounds by seniors

Ocke-Schwen Bohn, Birgitte Poulsen, and Sidsel Rasmussen
Department of English, Aarhus University, Denmark

Just about everything that is known about speech sound learning involves participants who are WEIRDY: Western, Educated, from Industrialized, Rich and Democratic countries, and Young of age. The near-exclusive focus on young learners is unfortunate for both practical and theoretical reasons. Practical, because it is not known how successful the ever-increasing number of mentally well-functioning seniors can be at learning additional languages, and theoretical because current speech learning models [1, 2] base their claims regarding life-long speech learning abilities on evidence from young adults.

This presentation reports on the first of a series of studies which examine speech learning abilities of people above the age of sixty. Our main question of this first study was whether age differences exist with respect to the efficacy of training syllable-initial /s-z/. To examine whether training, if effective, was restricted to one phonetic context, we were also interested in whether training of syllable-initial /s-z/ would have an effect on the perception of /s-z/ in the untrained final position. We also examined whether we could confirm previous findings which had shown that perceptual training had an effect on production (which was not trained).

Fifty-one native Danish listeners with minimal English language experience participated in the first study: 26 participants aged 18-35 years (10 of which were controls), and 25 participants aged 60-75, of which 8 were controls. The controls and the trainees took part in pre- and post-tests which examined the identification accuracy for English /z/ and /s/ in the two phonetic contexts (initial and final), with naturally produced tokens of CV and VC, with V = /i ɑ, u/. Both controls and trainees were also recorded for productions of syllable-initial and syllable-final words with /s/ and /z/ in a delayed repetition task. Only the trainees took 10 training sessions, equally spaced over three weeks, at their home using online links provided by the experimenters.

Results show that training was effective for both age groups: For the trained syllable-initial /s-z/, the young participants' percent correct rate changed from 75.8% at pre-test to 92.5% at post-test, and the old participants' percent correct rate changed from 78.5% at pre-test to 89.2% at post-test. Paired t-test revealed that in both cases, these increases were highly significant ($p < .001$). For the untrained final /s-z/, the young participants' percent correct rate changed from 65.3% at pre-test to 75.9% at post-test, and the old participants' percent correct rate changed from 79.1% at pre-test to 81.3% at post-test. Paired t-test revealed that the changes were highly significant for the young participants ($p < .001$) and marginally significant for the older participants ($p = .0585$).

The results suggest that phonetic learning in a perceptual training paradigm is not compromised by an advanced age of the trainees, thus providing support for the claim that the ability for reorganization of phonetic systems remains intact over the entire life span. The presentation will report on potential differences in learning trajectories between the age groups, and it will report on whether and how perceptual training affected speech production.

References

[1] Best, C. T., & Tyler, M. 2007. Nonnative and second-language speech perception. In Bohn, O.-S. & Munro, M. J (Eds), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, J. Benjamins, 13–34.

[2] Flege, J. E., & Bohn, O.-S. 2021. The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning.* Cambridge University Press, 3-83.

Jaw movements in speech during physical activity

Heather Weston[1], Malin Svensson Lundmark[2,3], Donna Erickson[4] and Oliver Niebuhr[3]
[1]*Humboldt-Universität zu Berlin*, [2]*Lund University,* [3]*University of Southern Denmark,*
[4]*Yale University*

Diverse situations influence speech production in daily life. One common case is physical activity, which has been found to affect such varied aspects of speech as f0 [e.g., 1], pause placement [2], and voice quality [3]. Few articulatory investigations have been published, however, although modifications could be strongly expected to arise. For one, physical activity increases overall muscular tension, which could affect articulatory trajectories. For another, respiratory drive increases, which could lead to a larger mouth aperture to facilitate airflow, resulting in greater jaw displacement.

Jaw articulation has many functions in speech. A basic function relates to the fact that languages organize speech units into 'syllables' (or 'mora'). To produce syllable-/mora-sized units, the jaw/mouth opens and shuts. The degree of jaw opening (jaw displacement) is affected by intrinsic vowel effects: the jaw is more open in open vowels and less open in closed vowels [4, 5]. Moreover, the degree of syllable prominence correlates strongly with the amount of jaw displacement once the intrinsic vowel height effects have been factored out [6].

How physical activity affects jaw articulation remains largely unclear – to our knowledge, no direct investigations have been published. However, studies of loud speech, which shares some production characteristics with speech during exercise, have shown that speakers increased oral aperture compared with typical speech [7, 8]. Additionally, our own acoustic work in progress has suggested larger jaw displacements with increasing physical workload. The first two formants of the German point vowels [i, a, u] were assessed in stressed position in sentences read by 48 participants exercising at low and moderate intensity levels. While changes in F2 were vowel- and speaker-dependent, F1 values generally increased across vowels between all workload comparisons (rest vs. low, low vs. moderate, rest vs. moderate), though the greatest increases were seen for low vowel [a]. Increases in vocal intensity were also greatest for [a]. Taken together, these results suggest greater jaw displacement in speech during exercise compared with speech at rest.

One barrier to investigating jaw articulation during exercise is the difficulty of data collection with traditional methods, such as electromagnetic articulography (EMA), which preclude participant movement. To test the predictions derived from our acoustic data, we are therefore using a new wireless device for recording jaw articulations: the MARRYS helmet [9]. The MARRYS helmet is currently being developed at the University of Southern Denmark. It mechanically measures jaw displacement using two bending sensors on either side of the face (Fig. 1). In addition, it simultaneously records acoustic data using an attached microphone that is time-aligned with the jaw kinematics to keep the mouth–microphone distance constant. Moreover, as the MARRYS helmet is portable and worn by the participant, it can be used in natural settings outside of the lab and during various activities.

Following the design of our previous acoustic study, the present study assesses jaw displacement during production of the German vowels [i, a, e] during low and moderate exercise. Based on the formant changes we observed previously, we predict that jaw displacement will increase for all vowels as physical-activity level increases, with greatest displacements for the low vowel [a]. We further predict that the increase in jaw displacement will give rise to local increases in prosodic prominence that will be conditioned by vowel, with [a] being most prominent. This line of inquiry gives insight into how and why real-life situations, from walking to physically demanding jobs, can affect different aspects of speech, while simultaneously demonstrating the utility of a new piece of equipment, the MARRYS helmet, to obtain articulatory data in the wild.

Figure 1. *The MARRYS helmet*.

References

[1] Primov-Fever, A., Lidor, R., Meckel, Y., & Amir, O. 2013. The effect of physical effort on voice characteristics. *Folia Phoniatrica et Logopaedica*, *65*(6), 288-293.

[2] Baker, S.E., Hipp, J., & Alessio, H. 2008. Ventilation and speech characteristics during submaximal aerobic exercise. *Journal of Speech, Language, and Hearing Research*, *51*(5), 1203-1214.

[3] Godin, K.W., & Hansen, J.H.L. 2015. Physical task stress and speaker variability in voice quality. *EURASIP Journal on Audio, Speech, and Music Processing*, 2015(1).

[4] Erickson, D. 2002. Articulation of extreme formant patterns for emphasized vowels. *Phonetica* 59, 134-149.

[5] Williams, J.C., Erickson, D., Ozaki, Y., Suemitsu, A., Minematsu, N., & Fujimura, O. 2013. Neutralizing differences in jaw displacement for English vowels. *Proceedings of International Congress of Acoustics, POMA* 19, 060268.

[6] Erickson, D., & Kawahara, S. 2016. Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistic Vanguard* 2, 102-110. De Gruyter Mouton. DOI 10.1515/lingvan-2015-0025.

[7] Šimko, J., Benus, S., & Vainio, M. 2016. Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue. *The Journal of the Acoustical Society of America, 139 1*, 151-162.

[8] Huber, J. E., & Chandrasekaran, B. 2006. Effects of increasing sound pressure level on lip and jaw movement parameters and consistency in young adults. *Journal of Speech, Language, and Hearing Research, 49*(6), 1368-1379.

[9] Erickson, D., Niebuhr, O., Gu, W., Huang, T., & Geng, P. 2020. The MARRYS cap: A new method for analyzing and teaching the importance of jaw movements in speech production. *Proceedings of International Seminar of Speech Production*, 48-51.

Hyperarticulated speech sounds fast: Effects of speech style on perceived tempo

Leendert Plug,[1] Rachel Smith,[2] and Yue Zheng[3]
[1]University of Leeds, [2]University of Glasgow, [3]University of Nottingham

Many acoustic cues for perceived speech tempo variation have been identified [1 2 3 4], which together support the proposal that speech which conveys more complex spectral information is heard as taking more time than speech conveying less complex information. That is, if duration is controlled, spectrally more complex speech sounds faster than spectrally less complex speech. We therefore hypothesize that hyper-articulated speech sounds faster than normal speech when articulation rates are controlled.

We report two listening experiments which address this hypothesis using clear and normal sentence productions. Experiment 1 assessed listeners' ability to separate tempo and speaking mode judgements. We used sentences from the LUCID corpus [5] produced in two modes, *normal*, where speakers were instructed to speak 'casually, as if talking to a friend'; and *clear*, with speakers talking 'clearly as if talking to someone who is hearing impaired'. 10 sentences, each produced by 6 speakers, were used to create the stimuli.. By compressing the duration of *clear* sentences to that of *normal* ones, we generated four pairs per sentence: SPEED pairs which differed in tempo (*clear* uncompressed + *clear* compressed), PRECISION pairs which differed in speaking mode (*normal* uncompressed + *clear* compressed), BOTH pairs which differed on both dimensions (*normal* uncompressed + *clear* uncompressed), and NEITHER pairs which were identical (*normal* uncompressed + *normal* uncompressed; *clear* uncompressed + *clear* uncompressed). 82 native British English speakers listened to 120 sentence pairs, counterbalanced for order. They judged whether the members of each pair differed in speed, speaking mode, both, or neither. Results (Figure 1) showed that PRECISION pairs attracted significantly more 'speed' and 'both' responses than SPEED pairs attracted 'precision' and 'both' responses (37% vs. 22%; $\chi 2=131.2$, df=1, $p<0.001$), supporting our hypothesis that when duration is controlled, differences in speaking mode can trigger the percept of differences in speed.

Experiment 2 used the same stimuli to probe the direction of the effect of speaking mode variation on perceived tempo. 26 listeners judged which pair member was faster ('first', 'second', 'neither'). We created a variable to reflect our predictions as to which pair member should be heard as faster: the clear production for PRECISION pairs and the higher-rate production for SPEED and BOTH pairs. We fitted a conditional inference regression tree model to the responses with this variable (*Predicted*) as well as the pair type (*Type*: 'PRECISION', 'SPEED' or 'BOTH') as predictors. The tree algorithm iteratively implements binary data splits according to the strongest predictor for the relevant data subset (Figure 2). The first split is at the top of the tree (node 1): Figure 2 shows that a clear majority of responses involved, as predicted, the clear member of PRECISION pairs and the fast member of SPEED pairs being heard as faster. Lower nodes of the tree reveal that compared to SPEED and BOTH pairs, PRECISION pairs were more often associated with 'neither' responses, and had a slightly larger number of responses in the non-predicted direction. Still, the modelling confirms that responses to PRECISION pairs show a significant listener preference for hearing compressed clear sentence productions as relatively fast.

In summary, the results of both experiments, taken together, confirm that compressed clear sentence productions sound faster than normal productions with the same duration. Experiment 3, in progress, manipulates tempo range (fast, mid, slow) to test the generality of the effect. The precise mechanism underlying this effect remains to be established, as does its relation to other potentially competing influences on tempo perception, such as the fact that listeners tend to perceive speech as faster if it is hard to understand [6], which clear speech usually is not [7].

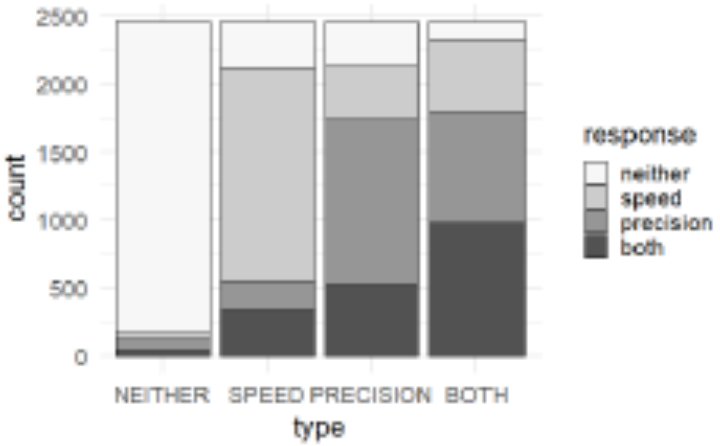


Figure 1. *Cumulative bar chart of the Experiment 1 response proportions by stimulus type.*

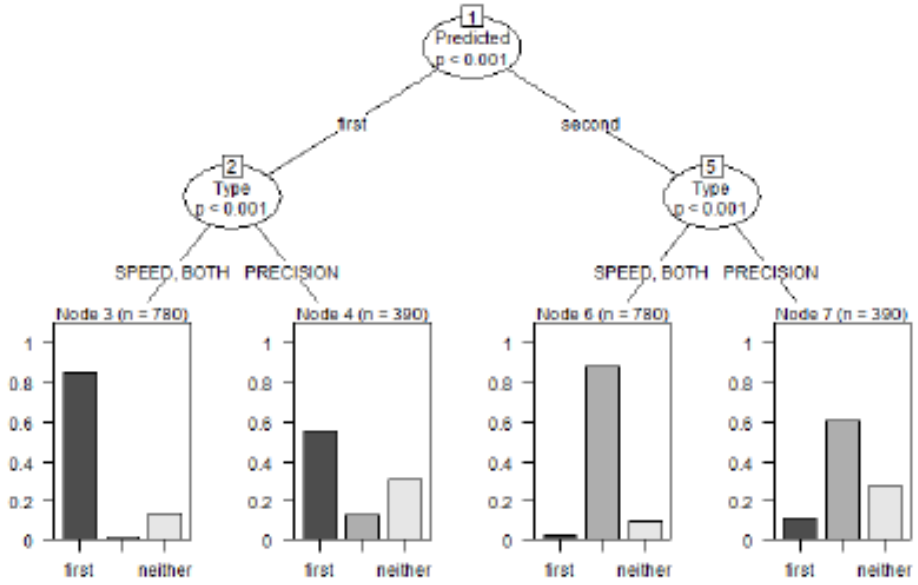


Figure 2. *Regression tree for responses to Expt 2 PRECISION, SPEED, and BOTH pairs.*

References

[1] Kohler, K. J. 1986. Parameters of speech rate perception in German words and sentences: Duration, f0 movement, and f0 level. *Language and Speech* 29, 115–139.

[2] Rietveld, A. C. M., Gussenhoven, C. 1987. Perceived speech rate and intonation. *Journal of Phonetics* 15, 273–285.

[3] Feldstein, S., Bond, R. N. 1981. Perception of speech rate as a function of vocal intensity and frequency. *Language and Speech* 24, 387–394.

[4] Weirich, M., Simpson, A. P. 2014. Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics* 43, 1–10.

[5] Hazan, V., Baker, R. 2011. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America* 130, 2139–2152.

[6] Bosker, H. R., Reinisch, E. 2017. Foreign languages sound fast: Evidence from implicit rate normalization. *Frontiers in Psychology* 8.

[7] Smiljanić, R., Bradlow, A. R. 1999. Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass* 3, 236–264.

A prosodically-annotated corpus of spontaneous narrations in mono- and heritage Russian

Yulia Zuban[1], Martin Klotz[2] & Sabine Zerbian[1]
[1]*University of Stuttgart,* [2]*Humboldt University Berlin*

Studies on prosodic variation and change in marginalized dialects have recently received some more attention (see [1]). Following this line of research, we present the prosodically-annotated corpus of heritage Russian (RuPro; spoken in the U.S.) which allows to also investigate global intonational features and compare them with monolingual speakers. Prosody in heritage languages has so far mostly been investigated for specific prosodic aspects, (e.g. [4], [15]).

RuPro is a subcorpus of the RUEG corpus [13], collected and annotated by the collaborative research unit «Emerging Grammars». Naturalistic repertoire data were elicited by means of presenting participants with a fictional incident. Participants were instructed to describe the accident, acting out different communicative situations which covered different modes and settings (spoken/written x formal/informal) (for details on methodology, see [12]). The overall corpus contains data from bilingual speakers of heritage Russian, heritage Greek and heritage Turkish and dominant English and dominant German respectively. Monolingual data of German, English, Russian, Greek and Turkish were elicited to enable comparisons.

The RuPro-subcorpus contains spoken Russian data of about 25k word tokens by monolingual and bilingual Russian heritage speakers in the United States (see [11] for a corpus on monolingual Russian). The corpus contains reports of 53 heritage speakers of Russian in the US and 40 monolingual Russian speakers. The speakers fall into two age groups, i.e. adolescents (14-18 years; 20 mono, 22 bilingual) and adults (22-35 years; 20 mono, 31 bilingual). The corpus data of each participant is enriched with extensive metadata on their language background, socio-economic status, and personality traits.

All data were transcribed, normalized, and annotated for communicative unit (CU, [6]) and part-of-speech. All annotations in the corpus have been created manually, part of speech tags and lemmas are manual corrections of automatic annotations created with UDPipe [9]. The annotated corpus data is stored in two formats, EXMARaLDA [10] for morphological and lexical annotations, PRAAT [3] for prosodic annotations. For the final corpus both data sources are merged using Pepper [14], aligning the annotated tokens from the EXMARaLDA and the PRAAT data. This way, both morphological and prosodic annotations can be used in a query in ANNIS [7] for search and visualization of our phenomena of interest.

The data of the spoken modes have been annotated with three prosodic layers, namely pitch accent placement on word-level, pitch accent type, and IP-boundaries. Pitch accent placement and pitch accent type (PA) were determined auditorily and phonetically, i.e., based on the F0 examination (low and high turning points). IP-boundaries were annotated following the annotation guidelines stated in [5]. Six types of pitch accents (L*, H*, L*+H, L+H*, H+L* and H*+L) and three phrase-final boundary tones (BT; L%, H% and LH%) emerged as relevant in these data. All H tones could be additionally upstepped (pitch range expansion compared to a preceding high tone) or downstepped (pitch range compression). PRAAT was used for phonetic annotation of the audio files [2]. Data were annotated by linguistically trained annotators, and interrater agreement for a subsample of the heritage Russian data was calculated (κ= 0.65).

As a starting point, we looked at the mean length of IPs and the number of PAs per IP in mono- and heritage speakers of Russian, averaged across formal and informal communicative situation. We found that heritage and monolingual speakers produce a similar number of PAs per IP (1,71 vs. 1,69), but the IPs of heritage speakers contained significantly less words than the IPs of monolingual speakers (2,6 words/IP vs. 2,9). Shorter IPs are likely to influence the perception of fluency and/or rhythm and could thus contribute to an overall perceived accent in heritage speakers (see [8]).

References

[1] Armstrong, M., Breen, M., Gooden, S., Levon, E., & Yu, K. M. (2022). Sociolectal and Dialectal Variation in Prosody. *Language and Speech*, 65(4), 783–790. https://doi.org/10.1177/00238309221122105

[2] Boersma, P. & D. Weenink. 2007 *Praat: doing phonetics by computer* (version 5.3.51).

[3] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

[4] Dehé, N. 2018. The Intonation of Polar Questions in North American ("Heritage") Icelandic," *Journal of Germanic Linguistics*, vol. 30, pp. 213-259.

[5] Himmelmann, Nikolaus P. et al. 2018. On the universality of intonational phrases – a crosslinguistic interrater study. *Phonology* 35.2, 207-245.

[6] Hughes, D., L. McGillivray & M. Schmidek. 1997. *Guide to narrative language: procedures for assessment*. Eau Claire, Wisconsin: Thinking Publications.

[7] Krause, T. 2019. *ANNIS: A graph-based query system for deeply annotated text corpora* [PhD Thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät]. https://doi.org/10.18452/19659

[8] Kupisch, T., Barton, D., Hailer, K., Klaschik, E., Stangen, I., Lein, T., & Weijer, J. v. d. 2014. Foreign Accent in Adult Simultaneous Bilinguals, *Heritage Language Journal*, 11(2), 123-150. doi: https://doi.org/10.46538/hlj.11.2.2

[9] Straka, M., Hajic, J., & Strakova, J. 2016. *UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing*. 8.

[10] Schmidt, T., K. Wörner. 2014. "EXMARaLDA". In: Handbook on Corpus Phonology. Ed. by Ulrike Gut Jacques Durand and Gjert Kristoffersen. Oxford University Press, pp. 402–419. URL: http://ukcatalogue.oup.com/product/9780199571932.do

[11] Volskaya, N., Kachkovskaia. 2016. Prosodic annotation in the new corpus of Russian spontaneous speech CoRuSS. *Proceedings of Speech Prosody* 2016 Boston, USA.

[12] Wiese, H. 2020. Language Situations: A method for capturing variation within speakers' repertoires. In: Yoshiyuki Asahi (Ed.), *Methods in Dialectology* XVI. Frankfurt a. M.: Peter Lang [Bamberg Studies in English Linguistics]. Pp. 105-117.

[13] Wiese, H., Alexiadou, A., Allen, S., Bunk, O., Gagarina, N., Iefremenko, K., Jahns, E., Klotz, M., Krause, T., Labrenz, A., Lüdeling, A., Martynova, M., Neuhaus, K., Pashkova, T., Rizou, V., Tracy, R., Schroeder, C., Szucsich, L., Tsehaye, W., Zerbian, S., Zuban, Y. 2019. RUEG Corpus (Version 0.2.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3236069

[14] Zipser, F., & Romary, L. 2010. A model oriented approach to the mapping of annotation formats using standards. *Workshop on Language Resource and Language Technology Standards, LREC 2010*. https://hal.inria.fr/inria-00527799

[15] Zuban, Y., Rathcke, T. & S. Zerbian. 2020. Intonation of yes-no questions by heritage speakers of Russian. In *Proceedings of 10th Speech Prosody*, Tokio, Japan.

Final shortening: Pre-boundary syllable duration and f_0 excursion decrease as boundary strength levels increase

Gerrit Kentner[1,2], Isabelle Franz[2,3], Christine A. Knoop[2], Winfried Menninghaus[2]
[1]*Goethe University,* [2]*MPI for Empirical Aesthetics,* [3]*HS Gesundheit Bochum*

Prosodic boundaries are marked by a variety of phonetic cues, most importantly by pauses, boundary tones (pronounced excursions of the pitch contour preceding a boundary), and by a slow-down in the speech rate approaching the boundary (pre-boundary lengthening) (see Krivokapić [1] for a review). Several studies suggest that pre-boundary lengthening, like pausing, is additive and increases with the strength of the prosodic boundary (e.g., Cambier-Langeveld [2]; Wightman et al., [3]). However, most previous studies consider short and predictable experimental sentences and disregard potential effects of sentence length and intonation, which are known to independently affect syllable duration: Boundary tones, which can have rising or falling contours, take time to execute. Rising contours typically indicate incompleteness (e.g., Bolinger, [4]; Tyler, [5]) and are often produced at boundaries within a sentence. Falling contours are mostly found at the end of declarative sentences. Rising contours take more time than falling ones (Sundberg [6]). Moreover, the pitch range that speakers can exploit for producing tonal events decreases as the sentence or utterance progresses (the so-called *declination* effect; see Cohen et al., [7]; Ladd, [8]). Correspondingly, prosodic boundaries that are produced early on in a given sentence may be marked by greater pitch excursions (which take longer) than sentence-final boundaries. Intonational effects on pre-boundary syllable duration may therefore be stronger in early regions of a sentence and decrease as the sentence progresses. Correspondingly, intonational effects might work against a positive correlation between pre-boundary syllable durations and boundary strength.

In order to examine the relationship between boundary strength, pausing, pre-boundary syllable duration and f_0 excursion on the pre-boundary syllable, we analyzed a corpus of prose texts by four German authors (Kleist, Goethe, Fontane, Kafka), each read aloud by eight professional speakers (>90k syllables, ~6h of speech). The sentences in these texts vary greatly in length (range: 1 to 170 syllables; mean: 30.4, median: 22.5 syllables). We compared the phonetic realization of prosodic boundaries with various degrees of boundary strength (level 0: no break; 1: simple phrase break; 2: short comma phrase; 3: long comma phrase; 4: sentence boundary; 5: direct speech boundary) and specifically studied the relationship between pre-boundary lengthening and pause durations. In addition, we examined the tonal realization of pre-boundary syllables. We find that pausing and pre-boundary lengthening are not correlated in a simple monotonic fashion. Whereas pause duration monotonically increases with predicted boundary strength (Figure 1), pre-boundary syllable duration and the pitch excursion on the pre-boundary syllable are largest for level-2 breaks and decrease significantly through levels 3 to 5 (Figure 2). Our analysis suggests that pre-boundary syllable duration is partly contingent on the tonal realization, which is subject to f_0 declination as the utterance progresses. We also surmise that the monotonic increase of pause duration along the predicted scale and the nonmonotonic pattern of pre-boundary syllable duration indicate that these phonetic signals reflect different speech production processes: The former likely reflect the closure of units of increasing size and planning time for upcoming clauses; the latter likely reflect current planning complexity: Compared to short comma phrases (level 2), finalized clauses and sentences (levels 3 to 5) require less or no time for planning the current phrase, so that the syllable duration decreases. Overall, this study shows that a simple monotonic correlation between boundary strength, pausing, and pre-boundary syllable duration is not valid. We emphasize the importance of examining complex spoken prose (apart from single self-contained experimental sentences) when studying the phonetic realization of the various degrees of prosodic boundary strength.

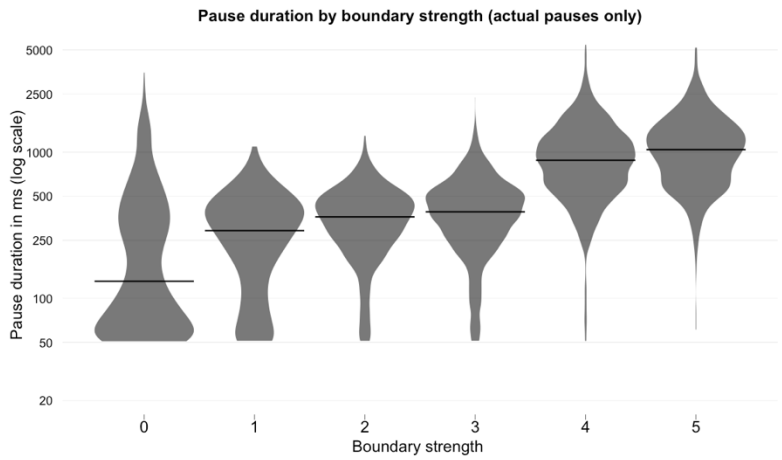


Figure 1. *Violin plot with pause durations in ms (log scale), broken down by the boundary strength index. The average line represents the median duration.*

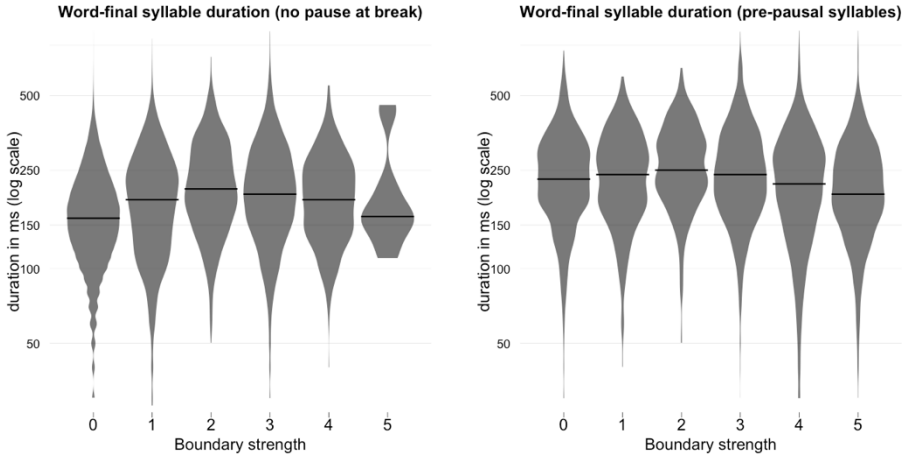


Figure 2. *Violin plots with pre-boundary syllable durations in ms (log scale), broken down by the boundary strength index and presence (right panel)/absence (left panel) of a pause.*

References

[1] Krivokapić, J. (2022). Prosody in articulatory phonology. In J. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic theory and practice* (pp. 213–236). MIT Press.

[2] Cambier-Langeveld, G. M. (2000). *Temporal marking of accents and boundaries* [Doctoral dissertation, University of Amsterdam].

[3] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717.

[4] Bolinger, Dwight. 1989. *Intonation and its uses: Melody in grammar and discourse*. Stanford University Press.

[5] Tyler, J. (2014). Rising pitch, continuation, and the hierarchical structure of discourse. *University of Pennsylvania Working Papers in Linguistics*, 20(1), 36.

[6] Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics*, *7*(2), 71–79.

[7] Cohen, A., Collier, R. & ‘t Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch? *Phonetica*, *39*(4–5), 254–273.

[8] Ladd, D. R. (1984). Declination: A review and some hypotheses. *Phonology*, 1, 53–74.

Pitch perturbations at vowel onset in different linguistic contexts in Thai

Alif Silpachai
Radboud University

A consonant may perturb the following pitch at vowel onset. For example, fundamental frequency (F_0) might be lower following a voiced stop than following a voiceless stop [1]. This perturbation (henceforth CF_0, following [2]) can develop into a lexical tone in some languages (see [3]). CF_0 can also change in different linguistic contexts. For example, CF_0 effects might be clearer in a high-pitch environment than in a low-pitch environment [4].

Research on how linguistic context modulates CF_0 effects is needed because different linguistic contexts, including tone context, sentential context, and place of articulation, can change the way in which CF_0 effects are observed and, consequently, the way in which conclusions about CF_0 effects are drawn. Previous research has provided conflicting findings regarding these modulations (e.g., compare the assumptions and/or findings of [5] with those of [6] , [7], or [8]). It is not clear whether the height of pitch following CF_0 effects in a tautosyllabic or tautomorphemic vowel modulates the CF_0 effects in tone languages. It is also unclear how CF_0 effects are produced in words compared to in phrases and in different phrase types. Lastly, previous studies have not provided consistent findings regarding whether CF_0 effects change in different places of articulation.

Twelve native speakers of Central Thai (6 females and 6 males, mean age = 45.75, *SD* 13.97, years produced monosyllabic words in Thai starting with /b/, /p/, /pʰ/, /m/, /d/, /t/, /tʰ/, /n/, /l/, /k/, /kʰ/, or /ŋ/, having /a/ or /a:/ as the vowel, bearing the falling (/51/), mid (/32/), or low (/21/) tone, and placed in a word either in isolation, a declarative statement, or an alternative question. The declarative statement, adapted from the one used in the study by [5], could be translated as "(I) will say the word _____ for (you) to hear," and the alternative question, based on the one used in the study by [7], had the following structure: "Did {you, s/he} {say, read, write} _____ or Y again?" where Y was a word that differed from the target word either in its vowel quality or onset consonant. The final analysis included 5,184 tokens. Three F_0 points, which were 10 ms, 20 ms, and 30 ms, following the onset of voicing, were analyzed and transformed to semitones relative to each speaker's mean F_0.

The results of mixed effects models suggested that CF_0 effects were modulated by tone context, sentential context, and place of articulation. CF_0 effects occurred most in the low tone, in a word in isolation, or following an alveolar consonant. Different patterns of CF_0 effects also occurred in different linguistic contexts. For example, F_0 was lower following a prevoiced stop than following a voiceless aspirated stop in the falling tone or a word in isolation, but F_0 was higher following a prevoiced stop than following a voiceless aspirated stop in the low tone or a declarative statement. F_0 was lower following a voiceless aspirated stop than following a voiceless unaspirated stop only in the low tone. F_0 was not different following a prevoiced stop compared to following voiced sonorant after bilabial and velar consonants or in the falling and mid tones.

The results are discussed in terms of physiological (e.g., vocal fold tension and aerodynamics) and/or nonphysiological (a controlled phonetics, [9]) factors that likely caused CF_0 effects in different linguistic contexts. The findings highlight the importance of the roles of different linguistic contexts in modulating CF_0 effects in a tone language. The implications of the findings are the following: (a) differences in CF_0 patterns may reflect individual differences or real differences between languages, (b) some languages may have more than one pattern of CF_0 effects, and (c) F_0 following a voiced sonorant in the low tone or /n/ or /l/ in any tone should not be used as a baseline to determine the direction in which F_0 following a stop is perturbed in Thai.

References

[1] House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105–113. https://doi.org/10.1121/1.1906982

[2] di Cristo, A., & Hirst, D. J. (1986). Modelling French micromelody: Analysis and synthesis. *Phonetica*, *43*(1–3), 11–30. https://doi.org/10.1159/000261758

[3] Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, *55*(1), 37–58. https://doi.org/10.2307/412518

[4] Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, *125*(1), 425–441. https://doi.org/10.1121/1.3021306

[5] Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *The Journal of the Acoustical Society of America*, *120*(5), 2884–2895. https://doi.org/10.1121/1.2346131

[6] Lai, Y., Huff, C., Sereno, J., & Jongman, A. (2009). The raising effect of aspirated prevocalic consonants on F_0 in Taiwanese. In J. Brooke, G. Coppola, E. Görgülü, M. Mameni, E. Mileva, S. Morton, & A. Rimrott (Eds.), *Proceedings of the 2nd International Conference on East Asian Linguistics*.

[7] Kirby, J. P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics*, *71*, 326–354. https://doi.org/10.1016/j.wocn.2018.09.009

[8] Shimizu, K. (1990). *Cross-language study of voicing contrasts of stop consonants in Asian languages* [Unpublished doctoral dissertation]. University of Edinburgh.

[9] Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.

Synchronic and diachronic variation in Munster Irish intensity prominence: systematic modelling of naturalistic storytelling data

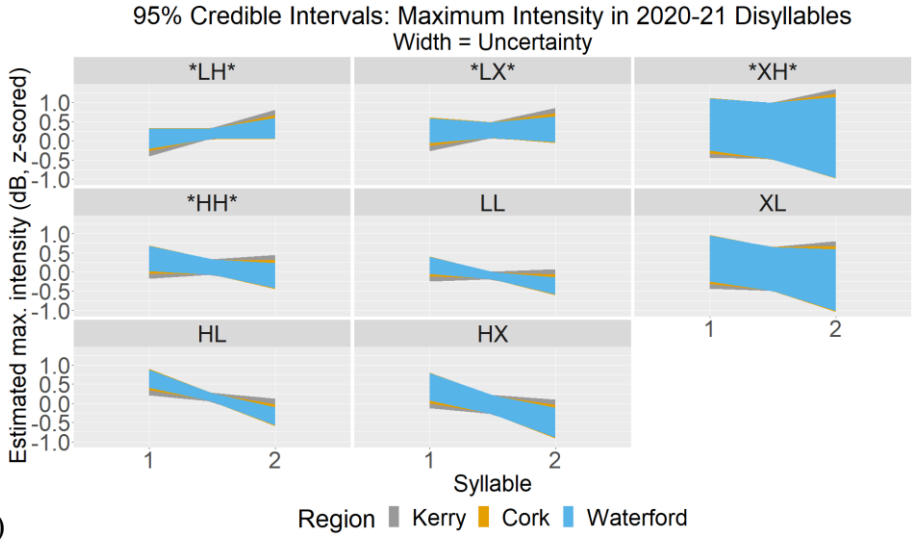
Connor McCabe (cmccab1@tcd.ie)[1]
[1]Trinity College, Dublin

Munster Irish (MI) is one of the three traditional macrovarieties of Modern Irish (Gaelic), spoken as a community language in small pockets of Counties Kerry, Cork, and Waterford. The variety is perhaps most noted for its lexical stress, described as weight sensitive in contrast to the initial-syllable stress of other Irish varieties. Limited to the first three syllables of a word, stress is said to occur on non-initial heavy syllables (i.e. with long vowels or diphthongs), with the sequence /ax/ sometimes suggested to be 'medium' heavy. Philological and dialectological treatments of this, e.g. [1,2], rely on impressionistic descriptions of stress location using ambiguous terminology. Phonological studies, e.g. [3-5], have frequently used analyses based on these 'data' as evidence regarding theories of metrical structure and stress assignment, often while disregarding regional/intraspeaker variation noted in traditional descriptions.

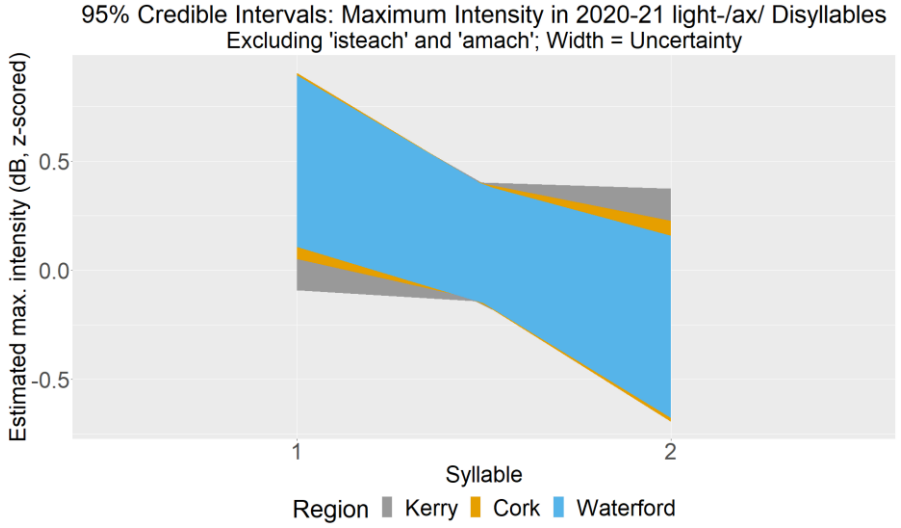
As part of a larger doctoral study of lexical and phrasal prominence in MI, the distribution of possible acoustic exponents of lexical stress across syllable positions in di- and trisyllables of different weight structures was examined statistically using L1 MI storytelling data from 1928 [6] and 2021-22. There were 20 speakers in the 1928 sample, supplying 3941 disyllables and 342 trisyllables, and 14 in the Zoom-collected 2020-21 sample, supplying 3758 disyllables and 425 trisyllables. This was the first instrumental phonetic study of the 1928 data, allowing for comparison of MI speakers a century apart. Instead of assuming stress location, acoustic measures were set as dependent variables modelled as a function of syllable position with item weight structure as a random slope (along with speaker). For brevity, this paper presents only modelling of maximum syllable intensity, as for both naturalistic and contextually controlled nonword data, this measure showed the most correspondence to existing descriptions of weight-sensitive lexical stress. Other measures modelled were maximum F0, F0 range, and vowel duration, modelling for which did not indicate weight-sensitive variation in prominence distribution even under experimental controls for phrasal context and segmental composition.

Maximum syllable intensity was z-scored by participant to facilitate pooling of data. Four linear mixed-effect models were constructed for di- and trisyllables in each dataset using Bayesian methods with the `brms` package in RStudio [7]. Each model was supplied with weakly informative priors, assuming a null effect of syllable position on intensity as most likely (mean 0*z*, standard deviation 1*z*). 95% credible intervals were plotted by weight-structure category to evaluate conformity of the measure's distribution in different weight structures to expected stress location (e.g. expected medial prominence in a light-heavy-light trisyllable).

As shown in Figure 1 with reference to 2020-21 disyllables, intensity prominence sometimes aligned with predicted non-initial stress, but divergence between previous descriptions and model results was frequent. Results suggesting non-initial intensity prominence for disyllable light-/ax/ in both eras and for all-light trisyllables in the modern data showed evidence of being influenced by the high-frequency adverbs *isteach* 'inside', *amach* 'outside', and *abhaile* '(to/at) home' when the data were remodelled with these items excluded (Figure 1b). This is taken to suggest possible lexical biases in the original description of non-initial stress in MI, notably because these particular items have exceptional non-initial stress in *all* dialects of Irish; the productivity of /ax/'s special status is therefore also questionable. Non-initial intensity prominence was only predicted for 2020-21 light-heavy disyllables (final), 1928 light-heavy-/ax/ trisyllables (medial), and 2020-21 light-heavy-light trisyllables (medial). Variability in model results across regions and between eras highlights the danger of (i) taking small samples from multiple regions as representative of MI as a whole, and (ii) treating modern data as implicitly compatible with impressionistic 19th and 20th-century descriptions.



(a)



(b)

Figure 1. *95% credible intervals for models of maximum intensity across syllable positions in 2020-21 disyllables. Abbreviated weight structures (H = 'heavy', L = 'light', X = '/ax/' marked with asterisks are those expected to exhibit non-initial stress.*

References

[1] O'Rahilly, T.F. 1932. *Irish Dialects Past and Present*. Dublin: Browne & Nolan.

[2] Ó Sé, D. 1989. "Contributions to the Study of Word Stress in Irish". *Ériu* 40, 147-178.

[3] Green, A.D. 1997. *The prosodic structure of Irish, Scots Gaelic, and Manx*. PhD Thesis. Cornell University.

[4] Iosad, P. 2013. "Head-dependent asymmetries in Munster Irish prosody". *Nordlyd* 40(1), 66-107.

[5] Windsor, J.W., S. Coward & D. Flynn. 2018. "Disentangling Stress and Pitch Accent in Munster Irish". *Proceedings of the 35th West Coast Conference on Formal Linguistics*, 430-437.

[6] Royal Irish Academy. 2009. *The Doegen Records Web Project*. Accessed 2020-2022. https://www.doegen.ie

[7] Bürkner, P.-C. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1), 1-28.

Perceptual illusions of ungrammaticality in foreign-accented speech

Sarah Wesolek[1,2], Piotr Gulgowski[1,3], Marzena Zygis[1,2]
[1]Leibniz-Centre General Linguistics, [2]Humboldt University Berlin, [3]University of Wrocław

Foreign accent can evoke biases affecting the way in which the speaker is perceived [1], or even influencing basic perceptual processes [2, 3]. Additionally, in past research, accented speech has been linked to processing challenges such as a decreased intelligibility [4], processing slow-down [5] and increased listening effort [6]. The reduced sensitivity to violations for foreign-accented speech in ERP studies [7, 8] suggests that listeners generally expect low level of grammatical and semantic proficiency from L2 speakers.

We hypothesize that the same expectation of ungrammaticality will also prime listeners to perceive grammatical errors in fully correct foreign-accented utterances. We predict that listeners exposed to foreign-accented speech will perceive illusory grammatical errors, a phenomenon previously dubbed 'grammatical tinnitus' in the press [9]. We also hypothesize that this effect will increase when foreign-accented sentences characterized by foreign suprasegmental features will contain additional phonemic changes.

To test our predictions, we conducted two mirror experiments for German and Polish, contrasting the perception of speech recorded by two foreign and two native-accented speakers. Sentences in the foreign condition were marked by a subtle foreign accent characterized by suprasegmental variation. Furthermore, stimuli of both accent types contained (i) well-formed sentences (no further manipulation), (ii) sentences with a phonological vowel substitution, and (iii) sentences with a grammatical error (in the critical region of the sentence). All stimuli were of comparable lengths and followed the same syntactic structure, see Example (1). For audio examples visit our OSF repository. Ungrammatical sentences were introduced to provide cases of grammatical violations. Here we will focus on well-formed and phonologically anomalous cases. In our study, 33 native speakers of German and 30 native speakers of Polish were asked to listen to sentences and answer the question 'Is this sentence grammatically correct?' by pressing the button corresponding to 'Yes' or 'No' as fast and accurately as possible.

To statistically model our results, for each experiment (Polish, German) a binomial logistic regression model (package 'Lme4', [10]) was fitted with JUDGMENT ACCURACY [correct, incorrect] as the dependent variable and ACCENT Type [native, foreign], ERROR TYPE [well-formed, phonological substitution, grammatical error], and their interaction as fixed factors. We also included PARTICIPANT and SENTENCE as random intercepts with SENTENCE ACCENT, SENTENCE TYPE, and their interaction as slopes. For the analysis of German data, 5917 data points were submitted. The dataset of the Polish experiment contained 5369 data points.

Our results show that in both languages sentences with no segmental or grammatical violations (well-formed) were more often judged as ungrammatical in foreign-accented than native-accented speech (GER: z=-3.3, p<.05; PL: z=3.25, p<.01). Contrary to our expectations, the 'tinnitus' was absent in the phonologically anomalous condition in the German experiment, see Figure 1). In the Polish experiment (Figure 2), this condition was associated with an increased number of incorrect judgments (z=2.33, p<.05), but against our second prediction this effect was not significantly different from the well-formed condition (no significant interaction).

The results from both languages suggest that well-formed utterances are more likely to be judged as ungrammatical when spoken with foreign accent as opposed to speech produced by natives. This effect is consistent with the hypothesis of 'grammatical tinnitus', a form of (un)grammaticality illusion. In regard to sentences containing phonological substitutions, the German data revealed no difference between foreign and nativeaccented speech. Possible explanations for between-language differences include differences in accent familiarity and differences in the strength of phonological violations used for manipulations.

(1) Example Sentence triple German experiment

Lena befragt die Lehrerin ...
Lena consults the Teacher

(i) well-formed
...zu ihrem Fehler in der Klausur.
about her[fem.] mistake[fem.] in the exam

(ii) phonological substitution
...zu ihrem F[ɛ]ler in der Klausur.
about her[fem.] mistake[fem.] in the exam

(iii) grammatical mistake
...zu ihrer Fehler in der Klausur.
about her[masc.] mistake[fem.] in the exam

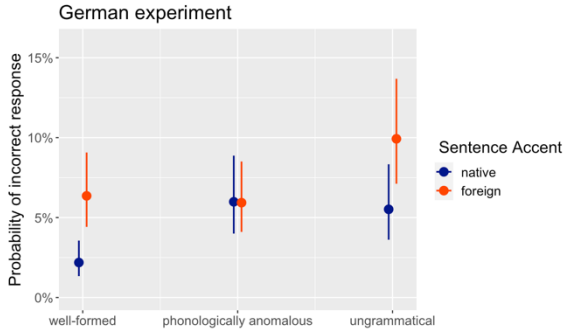


Figure 1: *Probability of incorrect response for Accent and Error Type, German.*

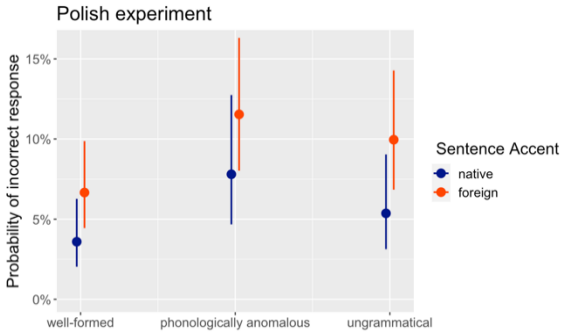


Figure 2: *Probability of incorrect response for Accent and Error Type, Polish.*

References

[1] Fuertes, J., Gottdiener, W., Martin, H., Gilbert, T., Giles, H. 2012. A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, 42, 120–133.

[2] Jannedy, S., Weirich, M. 2014. Sound change in an urban setting: Category instability of the palatal fricative in Berlin. *Laboratory Phonology*, 5, 91–122.

[3] Ingvalson, E. M., Lansford, K. L., Federova, V., Fernandez, G. 2017. Listeners' attitudes toward accented talkers uniquely predicts accented speech perception. *JASA,* 141 (3), EL234–EL238.

[4] Bent, T., Bradlow, A. 2003. The Interlanguage Speech Intelligibility Benefit. *JASA*, 114, 1600–1610.

[5] Floccia, C., Butler, J., Girard, F., Goslin, J. 2009. Categorization of regional and foreign accent in 5- to 7-year-old British children. *Inter. Journal of Behavioral Development*, 33 (4), 366–375.

[6] Van Engen, K. J., Peelle, J. E. 2014. Listening effort and accented speech. *Frontiers in Human Neuroscience*, vol. 8, 1–4.

[7] Hanulíková, A., van Alphen, P. M., van Goch, M. M., Weber, A. 2012. When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. *J Cogn Neurosci*, 24 (4), 878–887.

[8] Grey, S., van Hell, J. G. 2017. Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, 42, 93–108.

[9] Heiser, Sebastian. 2014. Berlins Bürgermeisterkandidat Saleh: Ein dubioses Hörproblem. TAZ: Gesellschaft/Medien. [Online]. Available: https://taz.de/Berlins-Buergermeisterkandidat-Saleh/!5034466/ [Accessed: January 6, 2023].

[10] Bates, D., Mächler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software*, 67 (1).

Poster Session 3

Sunday, 10:00 - 11:20

L1 writing system influence on consonant doubling in L2 speakers of English

Candice Frances[1], Eugenia Navarra-Barindelli[2], and Clara Martin[2,3]
[1]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands,* [2]*Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián, Spain,* [3]*Ikerbasque, Basque Foundation for Science, Bilbao, Spain*

Some languages, like English, do not have an exact one-to-one phoneme to grapheme correspondence in their spelling system. These languages are called opaque, as a single sound can be spelled in different ways (e.g., the phoneme /f/ spelled [ph] as in phone, [f] as in fill, or [ff] as in effect). Importantly, this spelling is not arbitrary and depends on different factors. These include graphotactic and phonotactic rules as well as the frequency of correspondence (between sound and spelling) and neighbor spellings. In other words, there are some norms that native speakers usually follow regarding how to spell a word they have never seen or heard before (Treiman & Wolter, 2018), even if they are not always aware that they do. When we think about L2 learners of English, they follow these same rules and norms to some extent, particularly when the writing systems do not overlap—as in e.g., Mandarin and Korean (Yin et al., 2020). This is further complicated by overlaps and contradictions with their L1 rules and norms in languages that share writing systems. In our current study, we asked whether L1 speakers of Spanish (n = 47)—a transparent language that shares the writing system with English—were able to follow the same rules (graphotactic and phonotactic) to inform their spelling of new items as L1 speakers of a language that does not share its writing system with English (namely, Mandarin, n = 46, and Korean, n = 39). In all groups, participants had started learning English in primary school. The Mandarin and Korean speakers were taking university level English and the Spanish speakers had at least a 70% on the English lextale (Lemhöfer & Broersma, 2012) and a 40 on the English portion of the BEST (de Bruin et al., 2017). In particular, we focused on consonant doublings (e.g., [f] versus [ff]) and compared results from L1 Mandarin and Korean speakers (Yin et al., 2020), L1 English speakers (Treiman & Wolter, 2018), and L1 Spanish speakers. Participants listened to each pseudoword twice and had to spell it out. We then counted the proportion of items that were spelled with a doubled medial consonant per condition.These data tested the following norms in spelling English pseudoword: (1) long vowels are usually followed by single consonant spellings and short vowels are often followed by double consonants and (2) if a vowel is spelled using two graphemes—regardless of the vowel—, the following consonant will be spelled using a single grapheme. We also compared the influence of a different L1 with the same orthographic system (Spanish) or a different orthographic system (Mandarin and Korean) using L1 English speakers as the baseline. We found that although overall patterns were the same between groups, the participants with a different L1 orthography adhered more closely to the L1 English speakers, particularly with respect to vowel length. Although it is possible that other differences between groups exist, we tentantively suggest that, although L2 speakers of English seem to understand the rules of their L2, they are influenced by the orthography of their L1 when spelling new words.

References

de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST Dataset of Language Proficiency. Frontiers in Psychology, 8. https://doi.org/10.3389/fpsyg.2017.00522

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. Behavior Research Methods, 44(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Treiman, R., & Wolter, S. (2018). Phonological and graphotactic influences on spellers' decisions about consonant doubling. Memory & Cognition, 46(4), 614–624. https://doi.org/10.3758/s13421-018-0793-9

Yin, L., Joshi, R. M., Li, D., & Kim, S.-K. (2020). Decisions about consonant doubling among non-native speakers of English: Graphotactic and phonological influences. Reading and Writing, 33(7), 1839–1858. https://doi.org/10.1007/s11145-020-10017-5

Plural alternations and consonant syllabification in Brazilian Veneto

Natália Brambatti Guzzo
Saint Mary's University

In Brazilian Veneto (BV; a heritage variety of Veneto spoken in several parts of Brazil), stress is mostly penultimate, except for when the word ends in a consonant, in which case stress is final (e.g., [biˈtʃeɾ] 'glass', [paˈɾoŋ] 'boss', [niˈsol] 'bedsheet'), suggesting that syllable weight is important for stress assignment in the language (Guzzo, 2022). However, different word-final consonants behave distinctly in pluralized nominals, which suggests that their syllabification is not identical. In this paper, I argue that an x-slot syllabification approach can account for the BV plural alternations without undermining the observations about stress assignment in the language.

In BV, the pluralization of masculine nominals involves the addition of suffix /-i/ (Stawinski, 1982; Luzzatto, 2000; see also Belloni, 2009). In nominals that end in an unstressed theme vowel (/-o/ or /-e/), this vowel is replaced by the plural suffix (1a). In nominals that end in a rhotic or a nasal, the plural suffix is added to the stem (1b; nasal quality is positionally conditioned in BV; Guzzo, 2022). However, in nominals that end in a lateral consonant, pluralization results in a VV string (produced as hiatus or diphthong), with no lateral on the surface (1c).

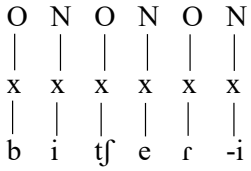
The question that arises is what motivates the alternation in plural forms observed in BV. One possibility is that BV avoids /li/ strings, which would result in the lateral being dropped. Nevertheless, this does not seem to be the case, as BV does exhibit /li/ word-internally ([gaˈlina] 'hen'), across words ([niˈsol impoɾˈta] 'imported bedsheet'), as well as in plurals where the lateral is the onset of a final CV syllable ([ˈbolo] → [ˈboli] 'cake.PL', [ˈvale] → [ˈvali] 'valley.PL').

Instead, I propose that the plural alternation observed in (1) results from word-final laterals being syllabified differently from word-final rhotics and nasals in BV. Specifically, while rhotics and nasals are onsets of syllables with empty nuclei, laterals are syllabified in the nucleus. Consequently, only rhotics and nasals may be resyllabified. In other words, when suffix /-i/ is added to stems ending in a rhotic or nasal, the consonant is resyllabified as the onset of the syllable whose nucleus is the suffix vowel (see (2); resyllabification is implied in the structure, which is simplified). On the other hand, in stems ending in a lateral, resyllabification of the lateral is prohibited. As a result, the lateral is dropped (see (3); the dropped lateral is in parentheses). (3) accounts for the VV string being variably produced as hiatus or diphthong, as the suffix vowel may be linked to the preceding N. The representations below include x-slots, following early approaches in Government Phonology (e.g., Kaye, 1990), to account for potential quantity effects (Passino et al., 2022), such as final stress in BV when the word ends in a lateral.

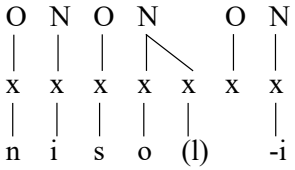
Two additional observations suggest that word-final laterals are syllabified differently from rhotics and nasals in BV. The first observation concerns the pluralization of monosyllables that end in consonant: while masculine monosyllabic nominals ending in a rhotic or nasal do not change their form in the plural (e.g., [kaŋ] 'dog.SG' → [kaŋ] 'dog.PL'), those ending in a lateral exhibit the same pluralization pattern as polysyllables (e.g. [kɔl] 'neck.SG' → [kɔi] 'neck.PL'). The second observation is about the pluralization of nominals that end in a stressed vowel, which also do not take plural suffix /-i/ (e.g., [puˈpa] 'dad.SG' → [puˈpa], *[puˈpai] 'dad.PL'; compare with [niˈsol] → [niˈsoi] 'bedsheet.PL'). In this case, I propose that adjacency between the suffix and the stressed vowel is prevented early in the inflection. This does not apply to forms such as [niˈsoi] since there is an intervening lateral at suffixation. The BV data thus further contribute to the observation that laterals in Romance often behave distinctly from other word-final consonants (for Portuguese, see Mateus & d'Andrade, 2000; for Old French, see Gess, 1998).

(1) a. [ˈsaso] → [ˈsasi] ‘stone.PL’, [ˈpaɾe] → [ˈpaɾi] ‘father.PL’
b. [biˈʧeɾ] → [biˈʧeɾi] ‘glass.PL’, [paˈɾoŋ] → [paˈɾoni] ‘boss.PL’
c. [niˈsol] → [niˈsoi] ‘bedsheet.PL’, [kaˈval] → [kaˈvai] ‘horse.PL’

(2)



(3)



References

Belloni, Silvano. 2009. *Grammatica veneta*. Esedra, Padova.

Gess, Randall. 1998. Old French NoCoda effects from constraint interaction. *Probus* 10: 207-218.

Guzzo, Natália Brambatti. 2022. Brazilian Veneto (Talian). *Journal of the International Phonetic Association*, Illustrations of the IPA.

Kaye, Jonathan. 1990. ‘Coda’ licensing. *Phonology* 7: 301-333.

Luzzatto, Darcy Loss. 2000. *Dissionàrio talian (véneto brasilian)–portoghese*. Porto Alegre: Sagra Luzzatto.

Mateus, Maria Helena and Ernesto d’Andrade. 2000. *The phonology of Portuguese*. Oxford: Oxford University Press.

Passino, Diana, Joaquim Brandão de Carvalho and Tobias Scheer. 2022. Syllable structure and (re)syllabification. In Gabriel, Christoph, Randall Gess and Trudel Meisenburg. *Manual of Romance: Phonetics and Phonology*. Berlin: De Gruyter.

Stawinski, Alberto Victor. 1982. *Gramática e vocabulário do dialeto italiano rio-grandense*. 3rd edn. Porto Alegre, Brazil: EST & Caxias do Sul, Brazil: EDUCS.

Spectral and durational properties of Judeo-Spanish and Bulgarian vowels

Mitko Sabev[1], Jonas Grünke[2], Christoph Gabriel[2], Bistra Andreeva[1]
[1]*Saarland University*, [2]*Johannes Gutenberg University Mainz*

Bulgarian Judeo-Spanish phonology is sorely understudied, with only a few empirical investigations conducted, e.g. [1,2]. We report the results of an examination of vowel F1, F2, F3 frequencies and duration in 3 minutes of spontaneous speech from each of 4 bilingual speakers of Judeo-Spanish (JSp) and Bulgarian (BL_BG), aged 80–88 (recorded 2012 in Sofia), in both of their languages, as well as from 4 monolingual Bulgarians (ML_BG), aged 79–86 (recorded 2016 in Sofia). 7285 vowel tokens were analysed in total. Although born in different parts of Bulgaria, all participants had lived in Sofia for over 60 years at the time of recording.

Standard Bulgarian has 6 contrastive stressed vowels, /ɛ a ɔ i ɤ u/; unstressed /a ɔ/ are raised and merge with /ɤ u/, respectively, while merger of unstressed /ɛ–i/ is nonstandard and stigmatised [3,4]. Judeo-Spanish has 5 corresponding phonemes, /ɛ a ɔ i u/, lacking /ɤ/.

Vowels were analysed in 3 prosodic conditions: stressed non-phrase-final, unstressed nonfinal and unstressed phrase-final. (There were insufficient stressed final tokens.) The following questions were addressed. (1) In stressed position, do vowel quality and duration differ across the 3 varieties? (2) What are the magnitude and main acoustic variables distinguishing stressed (nonfinal) and unstressed (nonfinal) vowels? (3) What are the magnitude and acoustic variables distinguishing unstressed nonfinal and unstressed final vowels? (4) What are the magnitude and acoustic variables distinguishing vowels in the pairs /ɛ–i/, /a–ɤ/ and /ɔ–u/ in stressed and unstressed position, and how much of the contrastiveness is lost in unstressed position?

MANOVAs and post hoc pairwise Games-Howell tests comparing the stressed vowels across the varieties showed that duration tends to be longer in JSp than in Bulgarian, which may be due to a slower speech rate. /a/ was significantly lower and /ɛ/ fronter in JSp than in BL_BG. However, the differences were small, pointing to shared vowel targets in the bilingual varieties. In ML_BG, /ɔ u/ were significantly fronter, while /u/ was also lower (Fig. 1).

Pillai's traces from MANOVA comparing stressed and unstressed (nonfinal) vowels revealed considerable change in unstressed position for the nonhigh vowels, /ɛ a ɔ/ (Fig. 2). Different reduction patterns, however, emerged from post hoc linear discriminant analysis (LDA). In BL_BG and ML_BG /a ɔ/, the primary change was F1 frequency reduction, while durational shortening played a secondary role. /ɛ/ was more reduced in BL_BG than in ML_BG (which is typical only of eastern dialects). JSp overall had less reduction, while F1 and duration played equal parts in distinguishing stressed and unstressed /ɛ a/. In high unstressed vowels, there was little spectral change, but notable durational reduction in JSp and BL_BG, but not in ML_BG.

Apart from Bulgarian /ɤ/, unstressed vowels were considerably longer in phrase-final than in nonfinal position in all varieties (Fig. 3). Final /ɤ/, however, was markedly underrepresented in the data (22 tokens), indicating that the results for this comparison may not be reliable. The final–nonfinal spectral differences, although low in relative weight compared to duration, result in smaller—centralised and raised—nonfinal vowel spaces.

The analysis of high vs nonhigh vowels (Fig. 4) showed dramatic contrast reduction in ML_BG and BL_BG unstressed /a–ɤ/ and /ɔ–u/, which can only be neutralising. Contrast loss in JSp /ɔ–u/ was smaller, but also resulted in a level of contrast that is likely to be perceptually undetectable. In unstressed /ɛ–i/, on the other hand, most acoustic overlap was found in JSp, and least in ML_BG. High and nonhigh vowels were predictably differentiated primarily by F1 frequency, with F2 playing a lesser, yet non-negligible part in /ɛ–i/ and sometimes in /ɔ–u/.

Our findings show that BL_BG and ML_BG have very similar vowel reduction and neutralisation patterns. JSp contrast reduction is weaker in nonfront vowels than in Bulgarian, but stronger in /ɛ–i/, which may be linked to three of the speakers being originally from eastern Bulgaria. JSp and BL_BG attach greater weight to duration in marking stress than ML_BG.

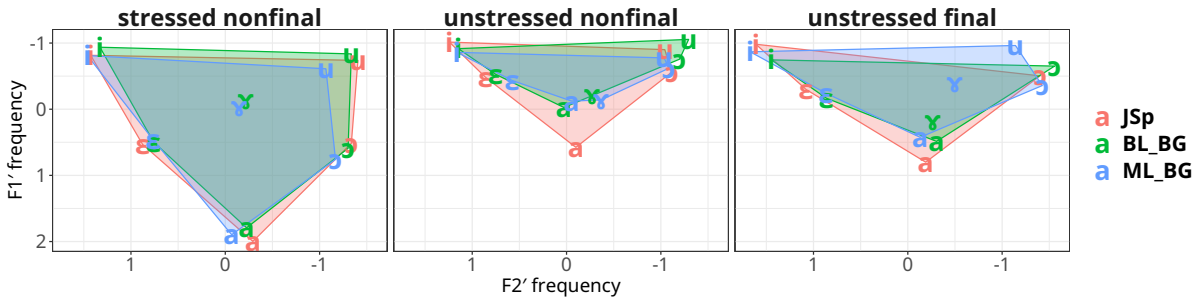


Figure 1. *Mean normalised F1–F2 vowel space across the varieties by prosodic condition.*

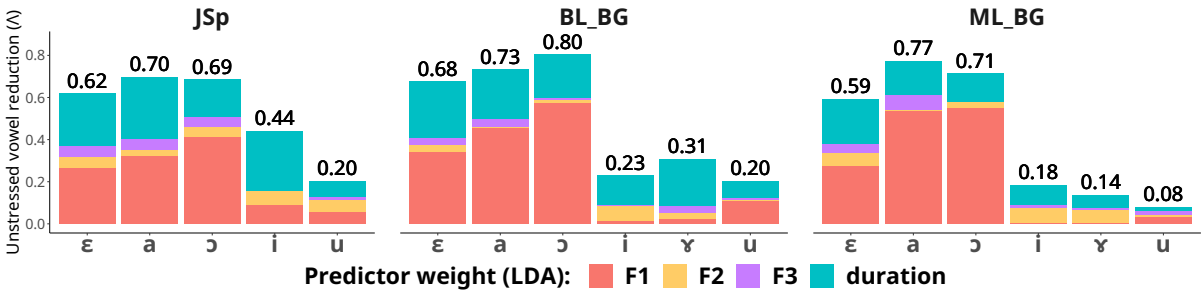


Figure 2. *Pillai's trace (Λ) & discriminant function weights for stressed vs unstressed vowels.*

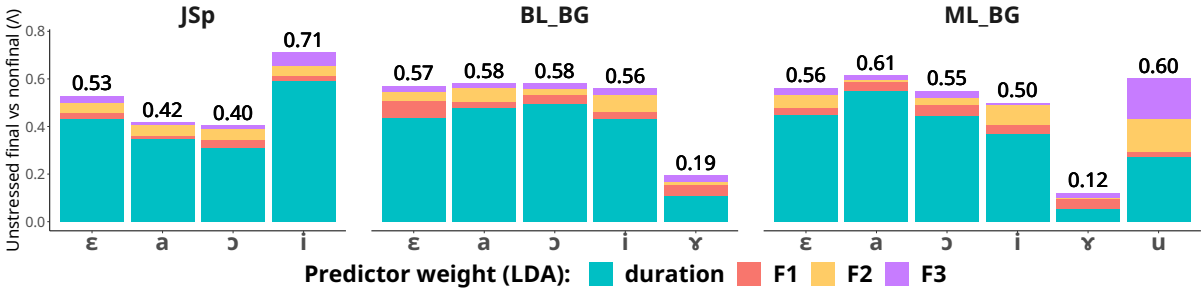


Figure 3. *Pillai's trace (Λ) & discriminant function weights for final vs nonfinal vowels.*

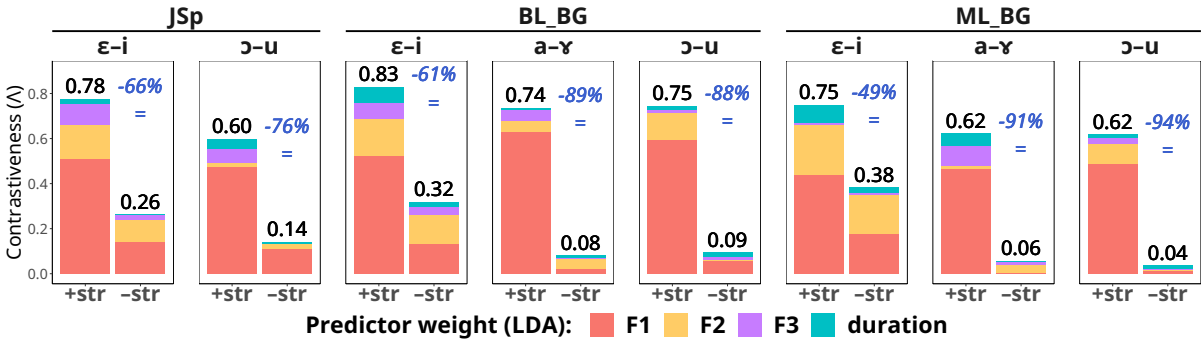


Figure 4. *Stressed (+) & unstressed (–) contrast (Λ, weights); unstressed contrast reduction (%).*

References

[1] Gabriel, C., Kireva, E. 2014. Speech rhythm and vowel raising in Bulgarian Judeo-Spanish. In: Campbell, N., Gibbon, D., Hirst, D. (eds), *Proc. Speech Prosody* 2014. Dublin, 728–732.

[2] Andreeva, B., Dimitrova, S., Gabriel, C., Grünke, J. 2019. The intonation of Bulgarian Judeo-Spanish spontaneous speech. In: Calhoun, S., Escudero, P., Tabain, M., Warren, P. (eds), *Proc. 19 ICPhS*. Melbourne, 3827–3841.

[3] Andreeva, B., Barry, W., Koreman, J. 2013. The Bulgarian stressed and unstressed vowel system. In: Bimbot, F., Fougeron, C., Pellegrino, F. (eds), *Proc. InterSpeech 2013*. Lyon, 345–348.

[4] Sabev, M. under review. Unstressed vowel reduction and contrast neutralisation in western and eastern Bulgarian: A current appraisal. *JPhon*.

Focus types and the prosody-gesture link in Catalan and German: A production study

Paula G. Sánchez-Ramón[1,2], Alina Gregori[2], Pilar Prieto[1,3], Frank Kügler[2]
Universitat Pompeu Fabra[1], Goethe University Frankfurt[2], Institució Catalana de Recerca i Estudis Avançats[3]

In the last decades, research has shown that gesture and speech are highly interconnected (e.g. [1], a.o.), and that information structure and prosody correlate in terms of prominence ([2] for German; [3] for Romance), but the role of focus types on the attraction of gesture use and the prosody-gesture link in adult speech has rarely been considered. For French-speaking children, [4] found that head gestures rather than prosodic features were used to indicate the information status of discourse referents, suggesting that those gestures with no referential connection to speech may play a linguistic structural role in communication (similar to prosody).

We investigate the impact of focus conditions on prosodic prominence, on gestures, and on their co-occurrence. Following [5], focus conditions are classified as: information focus (most important information), contrastive focus (overt presence of alternatives) or corrective focus (disagreement to a previous statement). Contrastive and corrective conditions have a stronger prosodic prominence than information focus conditions across languages [6]. Thus, we hypothesize that a stronger prosodic and gestural marking will be associated with corrective and contrastive focus constituents rather than with information focus.

A production study is currently being conducted relying on an adaptation of the elicitation method by [4] (Figure 1), which is suitable to investigate the synchrony between prosody and gestures in different focus types. The method consists of pictures prompted in a digital board game. Participants, video-recorded, communicate with an animated conversation partner who is blindfolded and their task is to request certain objects from the digital "speaker". The focus types can be controlled by the responses of the animated "speaker". For data coding, prosody will be analyzed using ToBI adaptations and prominence levels for each language ([7] for German; [8] for Catalan) by assuming that pitch accents are associated with different levels of prominence (in accordance with [9]; [10]). Regarding gesture labeling, head nods and hand gestures will be collected. Thus, strokes and their prominence levels will be annotated using the M3D labeling system [11]. We expect to analyze the data in the upcoming weeks for both German and Catalan.

Keywords: non-referential gestures; focus; methodology; prosody-gesture link

(1) 1a. “Agafa la maleta TARONJA”
1b. “Nimm den ORANGENEN Koffer”
Take the ORANGE suitcase

Figure 1. *Example of the elicitation method by Esteve-Gibert et al. (2021). Contrastive focus: Bag contains more items, and two suitcases differ in color.*

References

[1] McNeill, D. (1992). Hand and mind: what gestures reveal about thought. University of Chicago Press.
[2] Féry, C. & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. Journal of Phonetics 36(4). 680–703.
[3] Dufter, A. & Gabriel, C. (2016). Information structure, prosody, and word order. In Fischer, S. & Gabriel, C. (eds.), Manual of grammatical interfaces in Romance, 419–456. Berlin, Boston: De Gruyter.
[4] Esteve-Gibert, N., Lœvenbruck, H., Dohen M. & D'Imperio, M. (2021). Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech. Developmental Science. e13154.
[5] Krifka, M. (2008). Basic notions of information structure. Acta Linguistica Hungarica 55(3). 243–276.
[6] Zimmermann, M. (2008). Contrastive focus and emphasis. Acta Linguistica Hungarica, 55, 347–360.
[7] Grice, M., Baumann, S. & Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In Jun, S. (ed.) Prosody Typology: The Phonology of Intonation and Phrasing. Oxford: OUP,55-83.
[8] Prieto, P., Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., Sichel-Bazin, R., & Vanrell, M.M. (2015). Intonational phonology of Catalan and its dialectal varieties. In S. Frota & P. Prieto (Eds.), Intonation in Romance, pp. 9-62. Oxford: OUP.
[9] Baumann, S., Grice, M. & Steindamm, S. (2006). Prosodic Marking of Focus Domains - Categorical or Gradient? Proc. Speech Prosody 2006, Dresden, 301–304.
[10] Kügler, F. & Calhoun, S. (2020). Prosodic Encoding of Information Structure: A typological perspective. In Gussenhoven, C. & Chen, A. (eds.): The Oxford Handbook of Language Prosody, Oxford: OUP, 454-467.
[11] Rohrer, P., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Esteve-Gibert, N., Ren, A., Shattuck-Hufnagel, S. & Prieto, P. (2020). The MultiModal MultiDimensional (M3D) labeling system for the annotation of audiovisual corpora: Gesture Labeling Manual. UPF Barcelona.

An MRI examination of sibilants and liquids in Lower Sorbian

Phil J. Howson
Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany

Lower Sorbian is a moribund West Slavic language spoken in Eastern Germany (Marti, 2007). Lower Sorbian has a three-way sibilant inventory (Żygis, 2003; Howson, 2015) including dental, retroflex, and alveolopalatal places of articulation (e.g., /s̪, ʂ, ɕ/). Additionally, an alternation between the retroflex and post-alveolar occurs when the following segment is a high vowel (i.e., /ʂ/ > [ʃ]). The liquid inventory includes a trill, palatalized trill, and a lateral, (i.e., /r, rʲ, lʲ/; Howson, 2018). The primary aim of this study is to better define the articulatory properties of the Lower Sorbian sibilants and liquids using MRI technology.

For this study, a single male speaker of Lower Sorbian (age: 87 years) participated. Data were recorded with a Siemens 3T Trio with a pixel size of 1.2 mm x 1.2 mm, and a sagittal slice thickness of 1.8 mm. To generate 3D images of the vocal tract, 44 slices were taken. 3D image composition required the participant to produce static articulation for 14 seconds. Dicom files were converted into 44 bitmap files for each segment, which were then loaded into Image3D (Birkholz, 2023). Image 3D uses Catmull-Rom splines to trace tongue, lips, palate, and pharyngeal wall. Splines were traced for the mid-sagittal slice of the target segments, /s̪, ʂ, ɕ, lʲ, r, rʲ/ and [ʃ]. Splines were then exported to scalable vector graphics (SVG) file and imported to Inkscape for presentation.

The results revealed that for the sibilants, the dental fricative had a tongue tip facing up with the constriction in the dental region. /s̪/ also had a low tongue body with retraction into the pharyngeal region. The retroflex, despite its name, did not have the tongue tip facing up, rather the constriction was formed in the post-alveolar region, with a raised tongue body. And a slightly advanced tongue dorsum, at least compared to the dental. The alveolopalatal had the highest tongue body, forming a constriction along the hard palate and into the post-alveolar region. The allophonic post-alveolar had a constriction near the hard palate and a bunched tongue shape that formed an additional constriction in the post-alveolar region. The midsagittal traces for the sibilants are presented in Figure 1a.

The analysis of the liquids revealed a strong degree of palatalization for the lateral and a central constriction along the post-alveolar region. The tongue body was advanced with a slight pharyngeal constriction being produced with the tongue root. The rhotic was produced with the tongue tip near the front of the alveolar ridge and a high degree of tongue dorsum retraction. This resulted in a narrow constriction in the pharyngeal cavity. The palatalized rhotic had a more retracted place of articulation with a raised tongue body, compared to the unpalatalized rhotic. The tongue dorsum was also retracted, but not as retracted as for the unpalatalized rhotic. Nonetheless, there was still a noticeable constriction in the pharyngeal cavity. The midsagittal traces for the sibilants are presented in Figure 1b.

The data here presents evidence that there are three phonemic sibilants in Lower Sorbian and that the allophonic post-alveolar has a distinctly different shape from the other sibilants. The retroflex and alveolopalatal primarily differed in the length of the constriction and the height of the tongue body. The analysis also suggests that small variations in constriction formation with respect to place and length can result in robust differences in acoustic cues (Shadle, 1985; Badin, 1989) that are used to maintain delicate balances in a three-way system.

The analysis of the liquids revealed important information about palatalization and liquids: the tongue body and dorsum can freely advance to articulate a palatalized lateral, but that in the case of the trills, the articulatory constraints that require tongue dorsum retraction (Kavitskaya, Iskarous, Noiray, & Proctor, 2009) does not permit the same kind of advancement for rhotics. The high degree of freedom for the tongue dorsum, along with the lateral side channel, may be one of the primary features distinguishing the laterals as a class from rhotics.

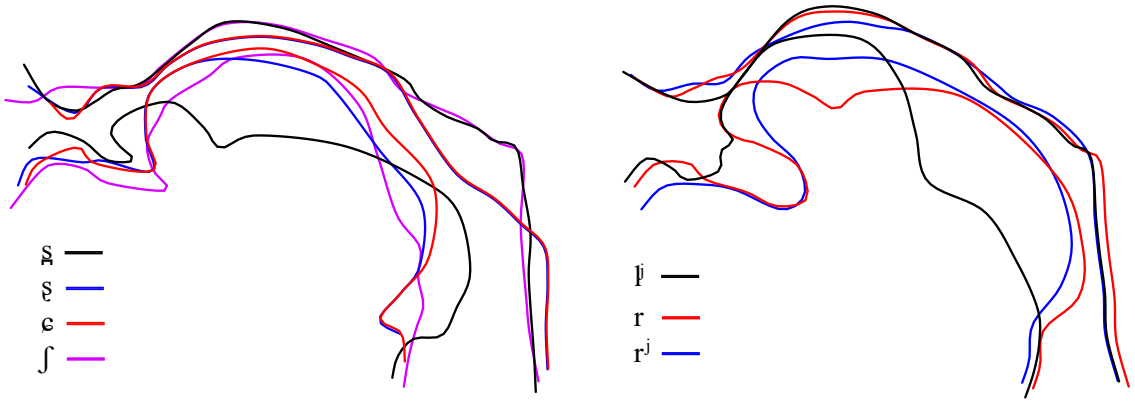


Figure 1. a.

References

[1] Badin, P. (1989). Acoustics of voiceless fricatives: production theory and data. *Quarterly Progress and Status Report, STL-QPSR '89*, pp. 33-55. Stockholm: Speech Transmission Laboratory, Royal Institute of Technology (KTH).

[2] Birkholz, P. (2023). *Image 3D*. https://www.vocaltractlab.de/index.php?page=image3d-download.

[3] Howson, P. (2015). An acoustic examination of the three-way sibilant contrast in Lower Sorbian. In, *Proceedings of Interspeech 2015*, pp. 2670-2674. Dresden: International Speech Communication Association.

[4] Howson, P. (2018). Rhotics and palatalization: an acoustic examination of upper and lower Sorbian**.** *Phonetica*, 132-150.

[5] Kavitskaya, D., Iskarous, K., Noiray, A., & Proctor, M. (2009). Trills and palatalization: consequences for sound change. In Reich, J., Babyonyshev, M., & Kavitskaya, D., (eds): *Proceedings of the Formal Approaches to Slavic Linguistics 17*, pp. 97-110. Ann Arbor: Michigan Slavic Publications.

[6] Marti, R. (2007). Lower Sorbian – Twice a Minority Language. *International Journal of the Sociology of Language*, 31-51.

[7] Shadle, C. (1985) The acoustics of fricative consonants. *RLE Technical Report 50*. Cambridge, M. A.: Massachusetts Institute of Technology.

[8] Zygis, M. (2003). Phonetic and phonological aspects of Slavic sibilant fricatives. In Hall, T. A. & Silke Hamann, S. (eds.), *Papers in Phonology and Phonetics* (ZAS Papers in Linguistics 32), pp. 175-213. Berlin: ZAS.

Cue Re-weighting in Production and Perception of Vowel Length Contrast in Cantonese

Yue Yin and Szeching Sin
Peking University

Background. For a phonological contrast, coarticulation often introduces co-varying cues, and sound change occurs via the cue re-weighting process, in which the coarticulated cue takes over as the primary cue. An important and controversial issue is how this change progresses in both perception and production. Two families of theories address this issue. Listener-driven theories [1, 2]suggest that sound change occurs earlier in perception than in production. while speaker-driven theories [3, 4] indicate that cue-reweighting happens in production first. Empirical studies also have inconsistent results [5, 6, 7], which shows that new data, preferably from different languages, are needed to further clarify this issue.

The vowel length contrast of [aː] and [ɐ] in Cantonese [8] provided a new type of sound change to tap into this issue. For the [aː ɐ] contrast, vowel length is the primary cue while the spectral cue (F1, F2) is the secondary coarticulated cue [9, 10], but their relevant importance varies in the production of young speakers [11], which demonstrates a cue-reweighting process. This study aims to shed further light on this issue by looking into the production and perception of the [aː ɐ] contrast in the younger generation.

Methods. We first conducted acoustic analysis on 5 minimal pairs of [aː ɐ] contrast with different initials and stop codas (/tap-tɐp/, /lap-lɐp/, /sap-sɐp/ /kʰat-kʰɐt/, /mak-mɐk/) produced by 16 native speakers of Cantonese (8 male, 8 female; aged 18-25). And then an identification experiment and a goodness rating experiment on another 7 native speakers (2 male, 5 female; ages 18-25) are done to tap into the perception aspect. The stimuli were drawn from the audios of /taːp/ (踏) and /tɐp/ (碓), which are modified in five steps of duration: the maximum(232 ms) and the minimum(125 ms) durations are the average duration of the original audios of /taːp/ (踏) and /tɐp/ (碓) respectively, and then take 3 points equally spaced between the maximum and minimum values to get 5 duration steps in total. In order to examine the contextual effect, we set one-morpheme and two-morpheme(X+/ tʃɐi/ (仔), a diminutive suffix) conditions for the above experiments.

Results. The acoustic analysis shows significant differences in both vowel duration and formants (F1, F2), with the long vowel ([aː]) lower and more fronted than the short vowel ([ɐ]), but the minimal pairs with /k/ coda show a tendency that /mak/ is merging into /mɐk/. The results of the identification experiment demonstrate that the duration cue is more salient than the formant cue when the vowel duration is short, but as the vowel duration increases, the formant cue becomes more critical than the duration cue. When the vowel duration is short, participants are more likely to identify the stimuli as /tɐp/ (碓) regardless of the formant cues, and this is even more prominent in the two-morpheme condition. With longer duration, participants' judgments relied mainly on formant cues, tending to identify stimuli with /ɐ/ quality as /tɐp/ (碓)[1] and stimuli with /aː/ quality as /taːp/ (踏) (see Figure 1). The same tendency can also be found in the results of the goodness rating experiment.

Discussion. The results show that although speakers use both cues in production, via perception experiments, it turns out that the formant cue rises in prominence, gradually taking over as the primary cue. These findings have implications for the theories of sound change in terms of the interaction of perception and production, and the choice between cue-reweighting and merging. On the one hand, the results suggest that the cue re-weighting process of [aː ɐ] contrast in Cantonese is more advanced in perception than production, which supports the

[1] Although when the vowel duration reaches the maximum, the frequency of identifying stimuli with /ɐ/ quality as /tɐp/(碓) has slightly decreased, it maintains higher than 70%.

listener-driven theories of sound change. On the other hand, this shift in perception is also conditioned by production. The fact that the formant cue becomes more dominant as the vowel duration increases, suggests that this cue shifting can only happen under the condition of sufficient vowel duration, otherwise, it may lead to the merge of the contrast.

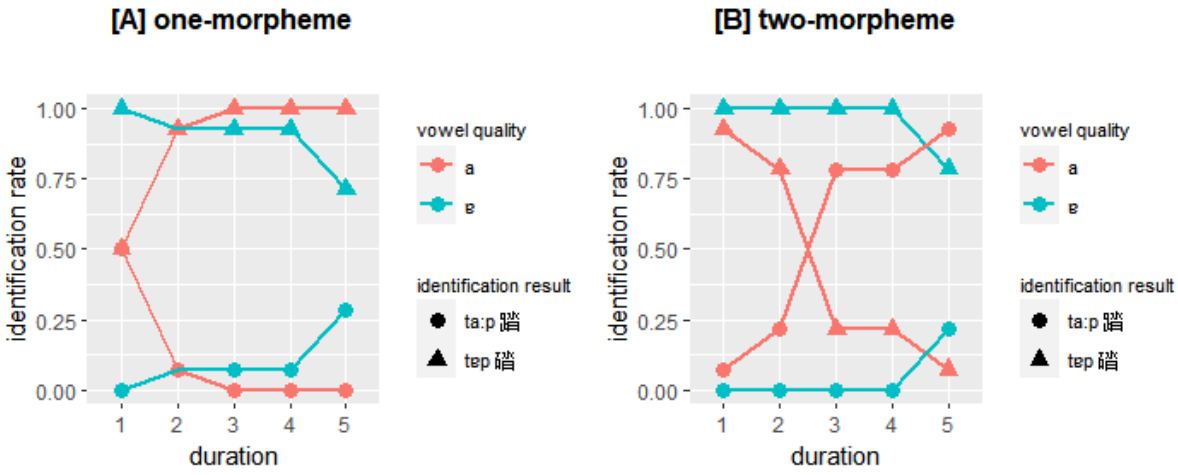


Figure 1. *Results of the identification experiment. The participants were asked to identify the stimuli or* /ta:p/*(踏) or* /tɐp/*(磕), therefore, for the same stimuli, the sum of the identification rates of* /tɐp/*(磕) and* /ta:p/*(踏) equals 1(100%).*

References

[1] Ohala, J. J. 1993. The phonetics of sound change. In Charles Jones (Ed.), *Historical Linguistics: Problems and Perspectives*. London: Longman, 237-278.

[2] Beddor, P. S. 2009. A coarticulatory path to sound change. *Language* 85(4), 785-821.

[3] Kang, Y. 2014. Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, 45, 76-90.

[4] Kirby, J. (2013). The role of probabilistic enhancement in phonologization. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization*. Oxford: Oxford University Press, 228-246.

[5] Harrington, J. (2012). The coarticulatory basis of diachronic high back vowel fronting. In M. J. Solé & D. Recasens (Eds.), *The initiation of sound change. Perception, production, and social factors*. Amsterdam: John Benjamins Publishing Company, 103-122.

[6] 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.

[7] Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66, 185-216.

[8] Kao, D. L. 1971. *Structure of the syllable in Cantonese*. The Hague: Mouton.

[9] Li, X. 1985. Guangzhouhua yuanyin de yinzhi ji changduan duili [Opposition of vowel quality and length in Cantonese]. *Fangyan [Dialect]* 1, 28-38.

[10] Zhang, L. 2010. Guangzhouhua changduan yuanyin de yuyin shiyan xintan [Long and short vowels in Cantonese: a revisit with new experiments]. *Fangyan [Dialect]* 2, 134-144.

[11] Jin, L., & Zhang, M. 2013. Guangzhou fangyan changduan yuanyin tongji fenxi [A Statistic Analysis on Long and Short Vowels in Guangzhou Dialect]. *Yuyan yanjiu jikan [Bulletin of Linguistic Studies]*, 179-98+327-328.

Comparing conversational alignment of Articulation Rate and Speaking Rate in typical talkers and talkers who stutter

Lotte Eijk, Stefany Stankova and Sophie Meekings
Department of Psychology, University of York, United Kingdom

Conversational alignment is the process during which talkers change their speech to become more similar to their interlocutor. This phenomenon - also referred to as e.g., entrainment, convergence - has been observed on both local and global time scales in measures related to rhythm in typical populations [1, 2]. However, for people who stutter, rhythmic speech production is more challenging. A characteristic feature of stuttering disfluency is silent 'blocks' which affect Speaking Rate (e.g., [6]), but not Articulation Rate which only includes short pauses. If talkers align on Speaking Rate (SR) - which could be affected in conversations with people who stutter - instead of Articulation Rate (AR) to rhythmically align their productions, this potentially has negative implications for intelligibility and conversational success [3]. In this study, we investigated whether typical talkers align differently (locally or globally) to stuttering partners compared to typical speakers, contrasting SR with AR.

Twenty dyads participated over Zoom: ten typical-typical (T-T; *M* age = 31.3, *SD* = 12.6) and ten typical-stuttering pairs (T-S; *M* age = 32.8, *SD* = 12.2). Participants performed a picture description by themselves, and on a different day, participated in a spot-the-differences Diapix task [7] with another speaker. They described three pictures and had 10 minutes to spot 12 differences per picture. AR and SR were calculated per IPU (inter-pausal unit; pauses of over 0.5s e.g., [1]). IPUs of under 5 syllables were excluded. Stutters on a word were included in an IPU even if the pause was of over 0.5s. A script [8] was used in Praat [9] to calculate AR and SR, using the standard settings. Local alignment was investigated by predicting the AR and SR in one speaker's IPU (dependent variable) by the other speaker's previous IPU (independent variable) using a LMER model. For global alignment, ARs and SRs were calculated by dividing the number of syllables per picture per participant by the phonation time or duration, respectively. Difference scores were calculated by subtracting one participant's value from the other, and taking the absolute value. For both local and global alignment, we ran two LMER models in R [10], one including and one excluding an interaction with group (T-T vs T-S).

Figure 1 shows difference scores per picture per group. Difference scores for the pairs with the stuttering participant seem to be higher than for typical-typical dyads, with greater differences for SR. Statistically, for both AR and SR, we did not find any group effects, neither locally, nor globally, nor any global effects. However, we did find local AR alignment for the whole group of 20 pairs. In contrast, there was no evidence of local SR alignment.

This study shows the potential of using Zoom data in populations that may be more challenging to test in a lab. We successfully demonstrated local alignment to articulation rate in both typical-typical and typical-stuttering pairs despite the relatively poor audio quality of this mode of interaction compared to in-person conversation. Although we found no significant alignment at the global scale, the results show considerable individual variation in patterns of alignment over time, which may be attributable to differences in stuttering severity and auditory environment, which could potentially have an effect as the physical presence of a conversational partner has previously been found to strengthen effects, e.g., [11]. Future work will investigate factors affecting alignment over time on both the local and global level.

People who stutter have more difficulty controlling their speech rate. However, our data indicates that talkers align locally on AR, which is not affected by stuttering, instead of SR. We furthermore found no differences in the effects between groups. In conclusion, therefore, our results demonstrate that typical speakers align their AR locally to both stuttering and neurotypical interlocutors. Thus, it seems that stuttering blocks that affect speaking rate may not affect rhythmic alignment when people who stutter converse with typical talkers.

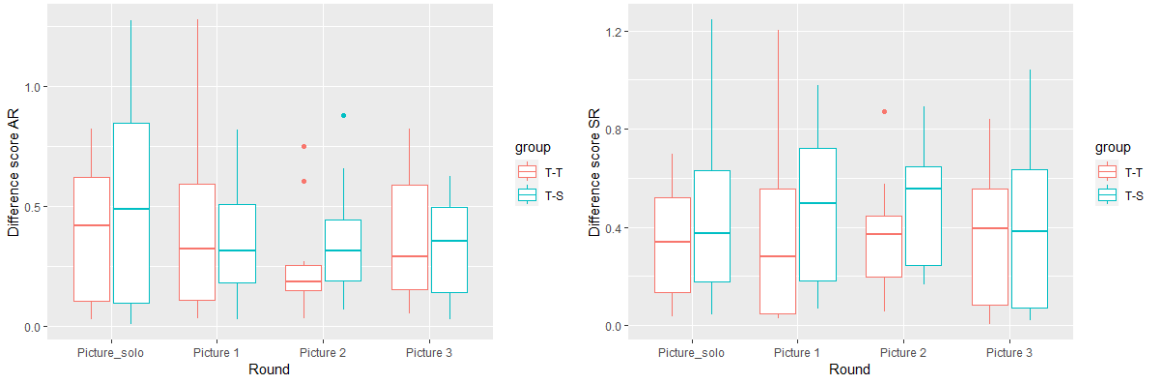


Figure 1. *Difference scores per round per group. The left plot shows the Articulation Rate difference scores, and the right plot shows the Speaking Rate difference scores. Colours indicated the different groups (T-T = typical-typical, T-S = typical-stuttering)*

References

[1] Levitan, R. and Hirschberg, J.B., 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, OR.

[2] Eijk, L., Ernestus, M., Schriefers, H., 2019. Alignment of pitch and articulation rate. *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia (2019), pp. 2690-2694.

[3] Borrie, S.A. and Liss, J.M., 2014. Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research, 57*(3), pp.815-824.

[4] Branigan, H.P., Pickering, M.J. and Cleland, A.A., 2000. Syntactic co-ordination in dialogue. *Cognition, 75*(2), pp. B13-B25.

[5] Brennan, S.E. and Clark, H.H., 1996. Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition, 22*(6), p.1482.

[6] de Andrade C.R., Cervone L.M., Sassi F.C., 2003. Relationship between the stuttering severity index and speech rate. *Sao Paulo Med J. 121*(2):81-4. doi: 10.1590/s1516-31802003000200010.

[7] Baker, R. and Hazan, V., 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods, 43*(3), pp.761-770.

[8] De Jong, N.H., Pacilly, J., & Heeren, W., 2021. PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically, *Assessment in Education: Principles, Policy & Practice,* 28:4, 456-476, DOI: 10.1080/0969594X.2021.1951162

[9] Boersma, P. and Weenink, D., 2022. Praat: Doing phonetics by computer (Version 6.2.23).

[10] 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/.

[11] Schoot, L., Hagoort, P., & Segaert, K. 2019. Stronger syntactic alignment in the presence of an interlocutor. *Frontiers in psychology, 10*, 685.

The intonation of yes-no questions in Guinea-Bissau Portuguese: features of a new variety?

Gabriela Braga[1, 2], Sónia Frota[2], Flaviane Fernandes-Svartman[1]
[1]*University of São Paulo,* [2]*University of Lisbon*

Portuguese is the only official language in Guinea-Bissau, an African country in West Africa. For historical and political reasons, European Portuguese (EP) is the official norm, leading to the assumption that the standard variety (SEP) is the one that should be spoken in the country. However, the country is a multilingual space in which Kriol is the language of national unity and identity, and Portuguese is spoken as a second language.

We analyzed the intonation of neutral yes-no questions in Guinea-Bissau Portuguese (GBP), to examine whether the intonation of this sentence type resembles that described for SEP [1], [2] (among others), or even another variety of European Portuguese [2], [3], [4], Brazilian Portuguese (BP) [2], [5] (among others), or African varieties of Portuguese already described in the literature – São Tomé and Príncipe [6], Angola [7], and Mozambique [8] – or whether we are facing a Guinea-Bissau Portuguese variety in formation.

The data was collected through a reading task following the lines of the InAPoP project [9], common to all the Portuguese varieties mentioned above. It included 10 neutral yes-no questions, besides vocatives, commands, and requests, which were distractors sentences. All sentences were produced after an eliciting context that was presented to the participant through a PowerPoint file, followed by the target sentence. Three repetitions of each sentence were produced in random order. The task was performed by four Guinea-Bissau participants (three men and one woman), aged between 20-25 years old, native speakers of Kriol, and speakers of GBP as a second language. The participants were undergraduate students at UNILAB (Universidade da Integração Internacional da Lusofonia Afro-Brasileira) in Bahia, Brazil, and had arrived in Brazil less than 30 days before the date of recordings. The audio material was segmented and analyzed with Praat [10]. The data analysis used the theoretical framework of Prosodic Phonology [11], and Intonation Phonology [12], and the P-ToBI annotation [13].

The analysis shows that GBP neutral yes-no questions (Fig.1) display a high tonal density. The pitch accent associated with the first prosodic word of the utterance is the one with the highest F0 pitch. This peak is followed by a smooth fall of the intonational contour (when segmental material is available), where high pitch accents are found (H* or !H*). The nuclear contour also presents a peak, which corresponds to a high tone associated with the stressed syllable of the last prosodic word of the utterance. The end of the contour shows a low boundary tone, and thus the nuclear contour is represented by (L+)H* L%. The formation of two peaks, one at each edge of the intonation contour, seems necessary to convey this sentence type in GBP.

Additionally, the first studies about the intonation of Kriol [14] show that speakers have two ways to express the meaning of an yes-no question: using the syntactic structure of a WH, with a particle ‘ke’ at the beginning of the sentence (this seems to be to most common strategy), or through the intonation contour, that presents the highest peak associated to the first prosodic word and another peak at the nuclear contour, similar to GBP.

Our results for GBP neutral yes-no questions show aspects that distinguish it from SEP, namely the shape of the nuclear contour and the high tonal density, as well as some that are similar, such as the initial pre-nuclear pitch accent. They also resemble yes-no question intonation from the Central-Southern varieties of BP, due to the configuration of the nuclear contour. However, GBP especially resembles African Portuguese varieties, as it displays a similar nuclear configuration together with high tonal density. This outcome was previously pointed out in the literature for declarative sentences [15], supporting the suggestion that GBP is developing its own intonation grammar.

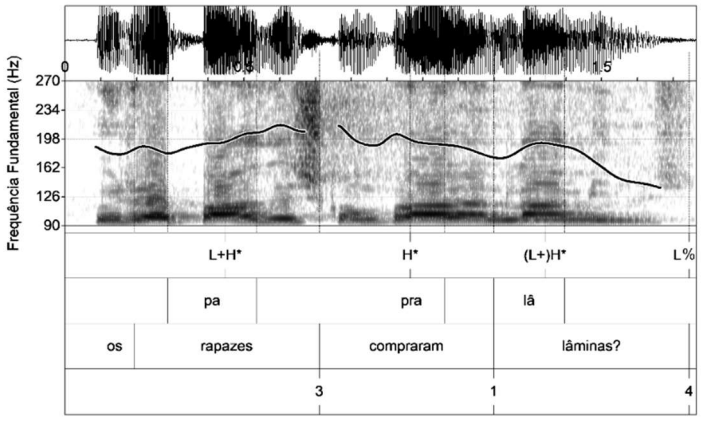


Figure 1. *GBP yes-no question "Os rapazes compraram lâminas?" (Did the boys buy razors?).*

References

[1] Frota, S. (2014). The intonational phonology of European Portuguese. In S.-A. Jun (Ed.), *Prosodic Typology II*. (pp. 6-42). Oxford University Press.

[2] Frota, S., Cruz, M., Svartman, F., Collischonn, G., Fonseca, A., Serra, C., Oliveira, P., & Vigário, M. (2015a). Intonational variation in Portuguese: European and Brazilian varieties. In S. Frota, & P. Prieto (Eds.), *Intonation in Romance* (pp. 235-283). Oxford University Press.

[3] Vigário, M., & Frota, S. (2003). The intonation of Standard and Northern European Portuguese. *Journal of Portuguese Linguistics*, *2*(spe), 115-137.

[4] Cruz, M., Crespo-Sendra, V., Castelo, J., & Frota, S. (2022). Asking questions across Portuguese varieties. In M. Cruz, & S. Frota. (Eds.), *Prosodic variation (with)in languages: Intonation, phrasing and segments* (pp. 36-70). Equinox Publishing.

[5] Castelo, J., & Frota, S. (2017). The yes-no question contour in Brazilian Portuguese: A geographical continuum. In P. P. Barbosa, M. C. de Paiva, & C. Rodrigues (Eds.), *Studies on Variation in Portuguese* (pp. 112-133) [Issues in Hispanic and Lusophone Linguistics, 14]. https:// doi.org/10.1075/ihll.14.04cas.

[6] Braga, G. (2019). Aspectos prosódicos das sentenças interrogativas globais do português de São Tomé: Uma análise inicial. *Estudos Linguísticos*, *48*(2), 688-708. https://doi.org/10.21165/el.v48i2.2323.

[7] Santos, V. G. dos. (2020). *Aspectos prosódicos do português angolano do Libolo: Entoação e fraseamento* [Doctoral dissertation]. University of São Paulo. https://doi.org/10.11606/T.8.2020.tde-03032020-174301.

[8] Serra, C., & Oliveira, I. (2022). The intonation of Portuguese spoken in Maputo, Mozambique: A case study. Special issue *Prosody and interfaces* (Ed. by C. Serra, F. Fernandes-Svartman, and M. Cruz). *DELTA, 38*(3), Article 202258880, 1-33. https://dx.doi. org/10.1590/1678-460X202258880.

[9] Frota, S. (Coord.). 2012–2015. *InAPoP – Interactive Atlas of the prosody of Portuguese*, Research project, University of Lisbon / FCT (PTDC/ CLE-LIN/119787/2010). http://labfon.letras.ulisboa.pt/InAPoP/.

[10] Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer* (Version 5.3.82). [Computer Software]. http://www.praat.org.

[11] Nespor, M., & Vogel, I. (2007). *Prosodic phonology*: With a new foreword. Walter de Gruyter.

[12] Ladd, D. R. (2008). *Intonational Phonology*. 2nd ed. Cambridge University Press.

[13] Frota, S., Oliveira, P., Cruz, M., & Vigário, M. (2015). *P-ToBI: Tools for the transcription of Portuguese prosody*. Laboratory of Phonetics and Phonology, University of Lisbon, ISBN: 978-989-95713-9-6. http:// labfon.letras.ulisboa.pt/InAPoP/P-ToBI/.

[14] Braga, G. "Prosódias em contato: O guineense e o português falado na Guiné-Bissau" (provisory title), Ph.D Dissertation, University of São Paulo, Brazil, in progress.

[15] Santos, V. G. dos, & Braga, G. (2017). Associação tonal em sentenças declarativas neutras do português de Bissau e de São Tomé. PAPIA, 27(1), 7-32.

Prosodic marking of allegedly attractive vs. unattractive objects in child-directed speech

Katharina Zahner-Ritter[1], Luisa Geib[2], Bettina Braun[2]
[1]*University of Trier,* [2]*University of Konstanz*

Child-directed speech (CDS) differs from adult-directed speech in a number of linguistic aspects, among others in prosody ([1]–[3]): CDS typically shows slower speaking rate, higher overall f0, and more variability in f0 and voice quality. So far, we know very little about the factors that predict prosodic modification *within* child-directed speech. In this study, we address visual attractiveness as one potential factor and test whether the *type of visual rendition* in a picture book (colourful picture vs. black-and-white line-drawing) affects parental prosody. Since children like bright colours (even more than adults, [4]), we expect parents to compensate for the reduced attractiveness of line-drawings by a stronger prosodic marking. From related research on motivational speech we know that motivational speech is characterized by higher and more variable f0, faster tempo and a lower amount of non-modal voice quality (resulting in more periodic signals, [5], [6]). Along these lines, we expect parents to produce targets depicted in black-and-white with shorter durations, higher and more variable f0, and more modal voice quality as compared to colourful pictures to keep the child engaged in the task.

To test this hypothesis, we recorded 11 German mothers in a picture-book scenario via Zoom while they were talking about colourful vs. black-and-white drawings to their child (manipulated within-subjects). Twelve high-frequent disyllabic words (that are part of children's early productive vocabulary [7]) were used and paired with freely available pictures. We created two PowerPoint "picture-books", one that had half of pictures in colour and the second half in line-drawings, the other with reversed rendition-order. Parents received both orders with a delay of 24 days on average. Their children were 1-2 years old (7 boys, 4 girls, mean age at first testing = 19 months, SD = 3.9 months). After the second recording session, parents filled in a questionnaire, indicating for every target word whether or not their child knew the word, and whether or not they think their child finds the respective object interesting.

The first mention of the target by the parent was labelled for beginning and end of the word, intonation and voice quality (labelled as modal, breathy, glottalized); maximum f0 and f0 range were automatically extracted in the target word (from 100 to 650 Hz). We used linear mixed-effects regression models [8] to analyse a) the duration of the target word (in ms, for 191 items that were produced as intended and occurred in final position), b) the maximum f0 (f0max) and c) the range in f0 (f0range) in the target word (in Hz, for 217 intonation-phrase final items). Voice quality was re-coded as modal vs. non-modal and analysed in a generalized mixed-effects model (217 phrase-final items). *Visual rendition* was entered as a fixed factor, parental report on child's *interest* (yes/no) and *familiarity* of the word (yes/no) as control predictors, *speaker* and *item* as crossed-random intercepts (models with random slopes did not converge).

Figure 1 shows the predicted effects for the fixed factor *visual rendition*. As expected, target words in black-and-white drawings were significantly shorter than in colourful rendition ($p < 0.05$), showed a higher and more variable f0 (maxf0, $p = 0.06$; rangef0, $p < 0.01$), and were more frequently produced with modal voice ($p < 0.05$). Together these prosodic adaptations, i.e., more f0 variation in shorter duration, increase a word's prosodic salience [9], and might hence increase a child's motivation for the activity ([5], [6]). Analyses further revealed intriguing interactions, e.g., between *rendition* and *familiarity* for the variables f0 range and voice quality (such that the rendition effect was stronger for unknown targets). We are currently collecting more data to test whether these interactions generalize. Taken together, our results reveal visual rendition to be one of the factors that predicts parental prosodic modifications within CDS. We will also discuss the implications that arise for early word learning [10].

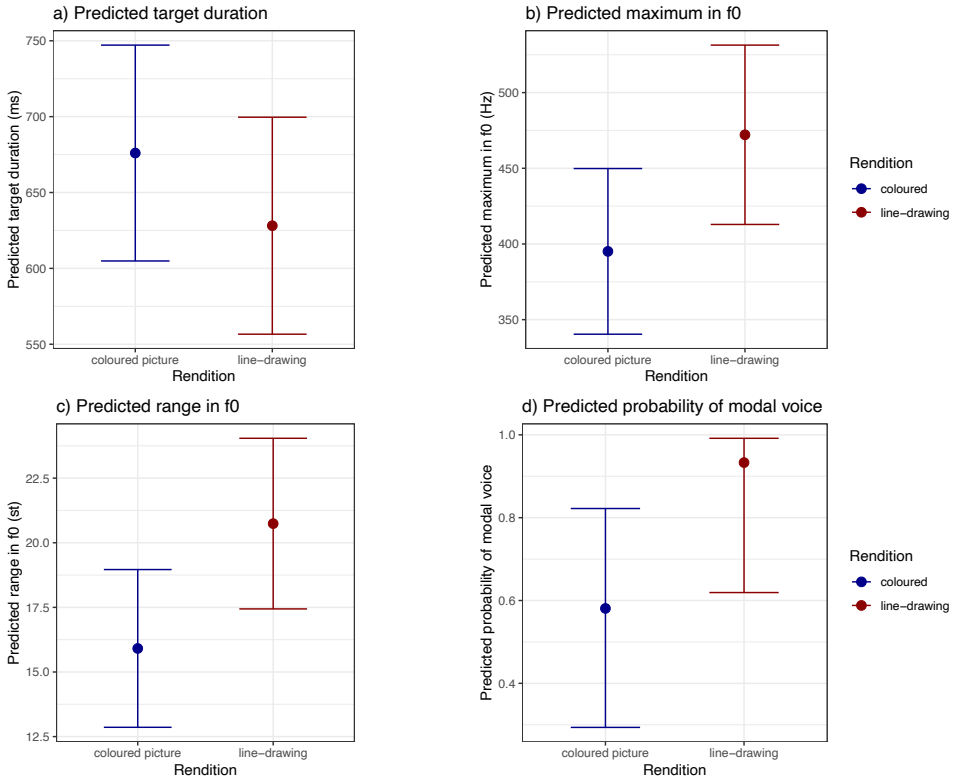


Figure 1. Overview of predicted effects of *rendition* for the different dependent variables.

References

[1] M. Kalashnikova, Ch. Carignan, and D. Burnham, 'The origins of babytalk: smiling, teaching or social convergence?', *Royal Society Open Science*, vol. 4, pp. 1–11, 2017.

[2] M. Weirich and A. Simpson, 'Effects of gender, parental role, and time on infant- and adult-directed read and spontaneous speech', *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 11, pp. 4001–4014, 2019.

[3] K. Zahner, M. Schönhuber, J. Grijzenhout, and B. Braun, 'Konstanz prosodically annotated infant-directed speech corpus (KIDS corpus)', *Proceedings of the 8th International Conference on Speech Prosody*, Boston, USA, 2016, pp. 562–566.

[4] M. R. Zentner, 'Preferences for colours and colour-emotion combinations in early childhood', *Developmental Sci*, vol. 4, no. 4, pp. 389–398, Nov. 2001.

[5] J. Voße, O. Niebuhr, and P. Wagner, 'How to motivate with speech. Findings from acoustic phonetics and pragmatics', *Front. Commun.*, vol. 7, p. 910745, 2022.

[6] J. Voße and P. Wagner, 'Investigating the phonetic expression of successful motivation', *Paper at the 9th Tutorial and Research Workshop on Experimental Linguistics*, Dec. 2019, pp. 117–120.

[7] M. C. Frank, M. Braginsky, D. Yurovsky, and V. A. Marchman, 'Wordbank: an open repository for developmental vocabulary data', *J. Child Lang.*, vol. 44, no. 3, pp. 677–694.

[8] R. H. Baayen, D. J. Davidson, and D. M. Bates, 'Mixed-effects modeling with crossed random effects for subjects and items', *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, 2008.

[9] S. Baumann and B. Winter, 'What makes a word prominent? Predicting untrained German listeners' perceptual judgments', *Journal of Phonetics*, vol. 70, 2018.

[10] N. Mani and L. Ackermann, 'Why do children learn the words they do?', *Child Dev Perspect*, vol. 12, no. 4, pp. 253–257, Dec. 2018.

The effect of embodied and non-embodied shadowing on L2 English pronunciation

Ting Yao[1], Patrick Louis Rohrer[1,2], Sharon Gutierrez Metelli[1], Valeria Tamburini[1], Bianca Regina Held[1], Pilar Prieto[1,3]
[1]*Department of Translation and Language Sciences, Universitat Pompeu Fabra,*
[2]*Laboratoire de Linguistique de Nantes, Université de Nantes*
[3]*Institució Catalana de Recerca i Estudis Avançats*

Recent evidence suggests that the shadowing technique (i.e., the immediate vocalization of speech stimuli; see Lambert, 1992. p.266) may be an effective tool to improve L2 English learners' pronunciation (e.g., Foote & McDonough, 2017; Sugiarto et al. 2020). However, little is known about the effectiveness of embodied shadowing (i.e., the immediate vocal and physical reproduction of speech accompanied by rhythmic hand gesture stimuli) and the role of the individual speech latency during training (i.e., the time-lapse between the model speech and the learners' reproduction, see Oki, 2010). This training study first investigates the impacts of embodiment during shadowing on improving L2 English pronunciation, and secondly, it assesses the potential effects of the learners' speech latency during pronunciation training on comprehensibility, accentedness, and fluency gains.

Fifty-four adult Catalan learners of English, with an intermediate to advanced level of English proficiency, were trained during six 30-minute sessions either in non-embodied shadowing or embodied shadowing conditions. The training was conducted through six online videos, in which participants immediately shadowed a speech while watching a native speaker narrating a story. The participants from the embodied group were requested to imitate the beat gestures (i.e., rhythmic hand gestures of up-and-down movements, see Fig.1) of the model meanwhile shadowing the speech, while the non-embodied group did not view any gesture, nor receive any instructions in terms of their own gesture production. Furthermore, in training sessions two, four, and six, a total of 48 pitch-accented syllables from the midsection of each learner's training speech were selected for analysis. Following Oki's (2010) description of shadowing latency, we coded with Praat (Boersma & Weenink, 2022) the speech latency (or temporal distance) between the onsets of target pitch-accented syllables produced by the native speaker and the corresponding learner's onsets.

Participants' pronunciations at pre-test and post-test reading-aloud tasks were assessed by three native English speakers on a Likert scale from one to seven (higher the score better the pronunciation) in terms of comprehensibility, accentedness, and fluency. Three linear mixed-effects models showed that, in contrast to the embodied shadowing, the non-embodied shadowing significantly improved speech accentedness and fluency scores after training (see Fig. 2, right and left panels). The null effects observed in the embodied shadowing training could be attributed to the increase in cognitive load caused by the incorporation of complex visual information, which might distract learners' attention from stimuli pronunciation. Further analysis of linear regression models showed that speech latency during training was a key predictor of the students' gains in comprehensibility and fluency in the embodied condition (see Fig. 3, right and left panels), which indicates that the value of embodied shadowing techniques for pronunciation learning may depend directly on learners' shadowing performance during training, with shorter latencies associated with greater improvement. Overall, this study provides pedagogical implications, suggesting the beneficial effects of non-embodied shadowing on L2 global pronunciation. It also emphasizes the importance of considering cognitive load and learners' shadowing performance during training when using embodied shadowing techniques for pronunciation learning.

Keywords: *shadowing, embodied shadowing, L2 pronunciation, speech latency, comprehensibility, accentedness, fluency*

Figure 1. *Beat gesture.*

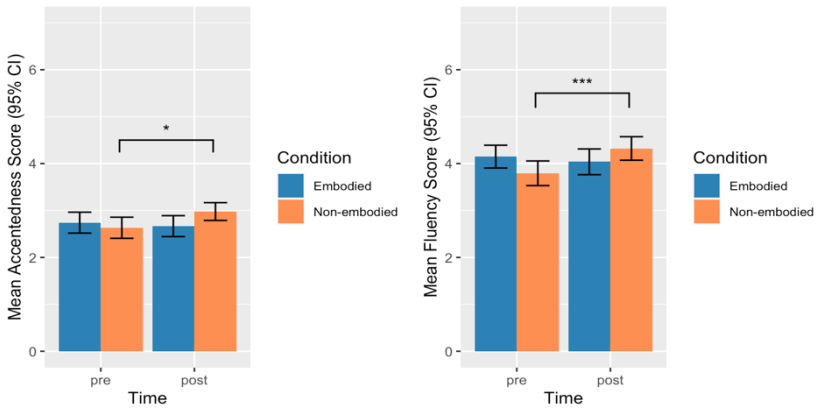


Figure 2. *Mean accentedness and fluency scores separated by Time (Pre- and Post-Test) and by Condition (Embodied and Non-Embodied Shadowing conditions).*

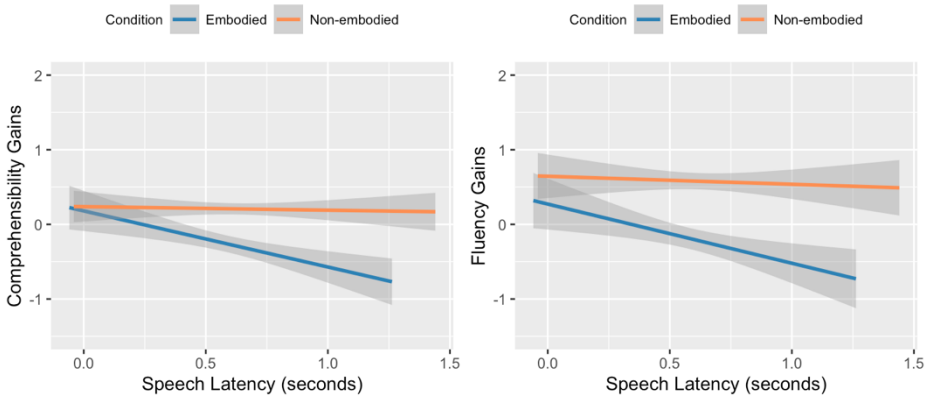


Figure 3. *Mean pronunciation gains as a function of speech latency in seconds.*

References

[1] Boersma, P. & Weenink, D. (2022). *Praat: doing phonetics by computer* [Computer program]. Version 6.2.12, retrieved 17 April 2022 from http://www.praat.org/

[2] Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, *3*(1), 34-56.

[3] Lambert S (1992) Shadowing. *Meta* 37(2): 263–73.

[4] Oki, T. (2010). The role of latency for word recognition in shadowing. *ARELE: Annual Review of English Language Education in Japan*, *21*, 51-60.

[5] Sugiarto, R., Prihantoro, P., & Edy, S. (2020). The Impact of Shadowing Technique on Tertiary Students' English Pronunciation. *Linguists: Journal Of Linguistics and Language Teaching*, *6*(1), 114-125.

Hiatus as phonetically re-articulated vowels in Italian

Valentina De Iacovo[1], Paolo Mairano[2], Bianca Maria De Paolis[1, 3], Anna Anastaseni[1, 4]
[1]*Laboratorio di Fonetica Sperimentale "Arturo Genre",* [2]*Université de Lille,* [3]*Laboratoire SFL (CNRS/Université Paris 8)*, [4] *GIPSA-Lab (CNRS/Université Grenoble-Alpes)*

Italian has a distinctive length opposition for consonants (e.g., *pala* /ˈpala/ 'shovel' vs. *palla* /ˈpalla/ 'ball'), but not for vowels. In effect, vowel length in Italian is considered to be allophonic: vowels are short when unstressed and also in stressed position within a closed syllable (*Marte* /ˈmar.te/ [ˈmar.te] 'Mars'), but long in stressed position in an open syllable (*mare* / ˈma.re / [ˈmaː.re] 'sea'). Nevertheless, there exist words (*corte* /ˈkorte/ 'court' vs. coorte /koˈorte/ 'cohort', as well as *re* /ˈre/ 'king' vs. ree /ˈree/ 'guilty') whose orthographic form suggests a phonological distinction [1, 2], which could be driven by length. Italian speakers consider the vowels in 'coorte' and 'ree' as heterosyllabic, as reflected in standard syllabication patterns (co-orte, re-e), and are therefore hiatuses.

The only study concerning such vowels [3] points out diverging articulatory strategies for their production. More specifically, the simple short vowel is articulated with greater apical peripherality than the hiatus. Because of this, and because of the presumed heterosyllabicity, we hypothesise that such vowels may best be described phonetically as re-articulated vowels. The situation may not be dissimilar to fake consonantal gemination in English, such as in 'dim morning' or 'ship partner', where the consonant can either simply be lengthened or re-articulated. In the case of Italian vowels, we posit that such re-articulation happens via glottal mechanisms that create a perceived syllabic boundary within the vowel itself. For this experiment, a corpus of non-sense words was therefore set up, so as to have the same structure for all pairs contrasting a phonetically long vowel and a hiatus: CVˈCVCV vs CVˈCVVCV ([taˈtaːpa] vs [taˈtaapa]) and CVCVˈCV vs CVCVˈCVV (e.g., [tataˈpa], [tataˈpaa]). We recruited 20 L1 Italian speakers, aged between 20 and 30, who were asked to read the target words within the frame sentence '*Ho detto la parola X tranquillamente*' ('I said word X calmly'). Speakers sat in a sound-proof booth, and the acoustic and electro-glottographic (EGG) signals were recorded, thereby obtaining a total of 480 target realisations. An initial inspection of the data with Praat [4] showed the use of various strategies by speakers; beyond duration, which is systematically longer for hiati than for long vowels, hiati generally also have different patterns for intensity and f_0 and in some cases glottalization (see figure 1). EGG data were not yet analysed, but will hopefully confirm the existence of laryngeal mechanisms used to differentiate hiati and long vowels, thereby confirming the status of such hiati as re-articulated vowels in Italian.

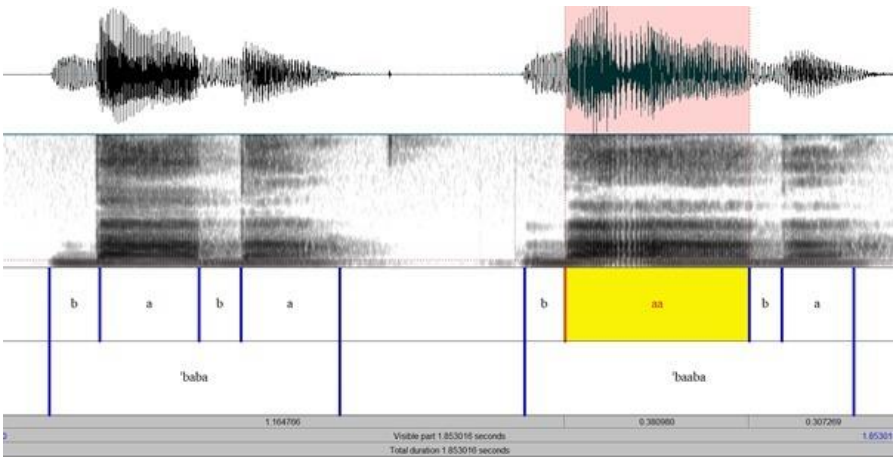


Figure 1. Example of duration differences between long and double vowels in Italian.

References

[1] Marotta, G., Sorianello, P. 1998. Vocali contigue a confine di parola. In Bertinetto PM., Cioni, L. (a cura di), *Unità fonetiche e fonologiche: produzione e percezione*, Atti delle VIII Giornate di Studio del Gruppo di Fonetica Sperimentale, Scuola Normale Superiore di Pisa, Pisa 18-20 dicembre 1997, Pisa, Stamperia SNS, pp. 101-113.

[2] Gili Fivela, B., Bertinetto, P. M. 1999. Incontri vocalici tra prefisso e radice (iato o dittongo?). *Archivio glottologico italiano*, 84(2), 129-172.

[3] Santini, D. 2018. Sara vs. Sahara: uno studio sperimentale di fonetica articolatoria sulle vocali singole vs. ripetute nella lingua italiana, condotto mediante sonda ecografica. Tesi di Laurea Specialistica, Università degli Studi di Pisa.

[4] Boersma, P., Weekink D. 2023. Praat: doing phonetics by computer [Computer program]. Version 6.3.09, retrieved 2 March 2023 from http://www.praat.org/

Do prosody and gestures help children with Developmental Language Disorder process pragmatic meanings? An eye-tracking study

Albert Giberga[1], Alfonso Igualada[1], Nadia Ahufinger[1], Mari Aguilera[2], Ernesto Guerra[3], Núria Esteve-Gibert[1],

[1]*Universitat Oberta de Catalunya,* [2]*Universitat de Barcelona,* [3]*Universidad de Chile*

Prosody is a key component in pragmatic processing and acquisition. Prosodic cues are used by children to understand pragmatic information, such as speech act ambiguities (Zhou et al., 2011), politeness (Hübscher et al., 2017) or irony (Li et al., 2012). Prosodic breaks can also disambiguate prepositional attachments (Wiedmann & Winkler, 2015). Crucially, when prosodic information is combined with gestures (i.e., multimodal prosody), children boost their understanding of complex pragmatic meanings (Kirk et al., 2011). Non-representational gestures such as beats also improve their discourse comprehension (Llanes-Coromina et al., 2018). Multimodal cues highlighting the speaker's pragmatic intent might be especially beneficial for children with Developmental Language Disorder because they might help overcome their deficits in structural language. In this study, we investigate the children's ability to infer pragmatic and discourse meanings through prosody and non-representational gestures, comparing Typically Developing children (TD) to children with DLD.

39 children with DLD and 39 TD children aged 5 to 10 were evaluated for their linguistic and cognitive abilities. Then, they participated in a visual-world eye-tracking experiment that assessed whether and when they process prosody and gestures to infer target pragmatic meanings. Upon hearing and watching a video with a speaker producing a sentence with a specific pragmatic meaning, they were asked to point at the image representing the meaning of the utterance they heard. We manipulated pragmatic intent (within-subjects: interrogativity; indirect requests; discourse structure - one experimental block per intent) and the multimodal cues present in the video (within-subjects: prosodically-enhanced, multimodally-enhanced, and non-enhanced). Each pragmatic intent was evaluated through 12 experimental trials, preceded by three familiarisation trials. We hypothesized that the presence of multimodal cues would facilitate the speed and accuracy of target image selection for children with DLD, as compared to typically developing children. Additionally, we anticipated that this advantage would become more pronounced as children with DLD encounter increasingly complex linguistic contexts throughout their developmental trajectory.

The findings from the offline task revealed that the presence of prosodic and multimodal enhancements had a positive impact on all children's ability to process interrogative and discourse meanings (χ^2= 36.96, p > 0.001; χ^2= 37.66, p > 0.001, respectively), regardless of their age or language ability (see Figure 2). Interestingly, a three-way interaction emerged for indirect requests, indicating that multimodal-enhancement was particularly beneficial for older children with DLD (χ^2= 16.99, p > 0.001). We are currently analysing the eye-tracking data (proportion and timing of fixations to the target image), with results expected by the end of April 2023. The present study will demonstrate the potential contribution of prosody and gesture to the comprehension of complex pragmatic meanings in children with language impairments, shedding light on the types of cues that facilitate successful communication interactions for these individuals.

Figure 1: *Visual display for an **indirect request** in the 'multimodally-enhanced' condition. Upon seeing the main character in the centre of the screen saying '**It is too dark in here…**' with the specific prosodic and gesture cues of indirect request, children are expected to select the image representing the speaker's intention behind that indirect request (bottom right) but not the image representing a literal statement (top left) nor any of the two distractors (top right and bottom left).*

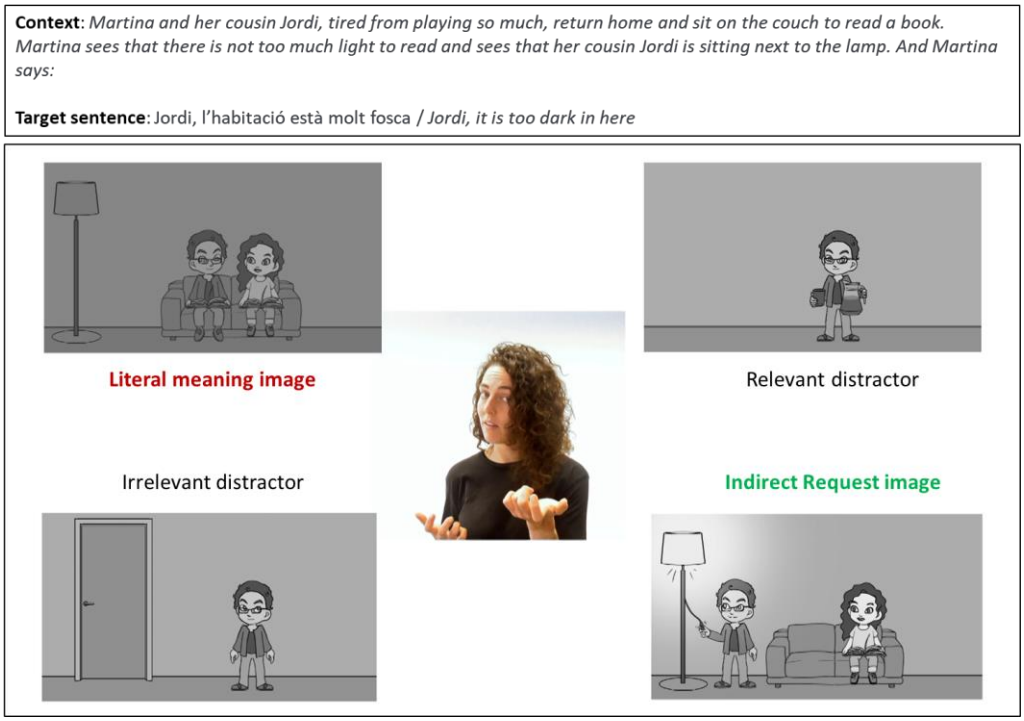
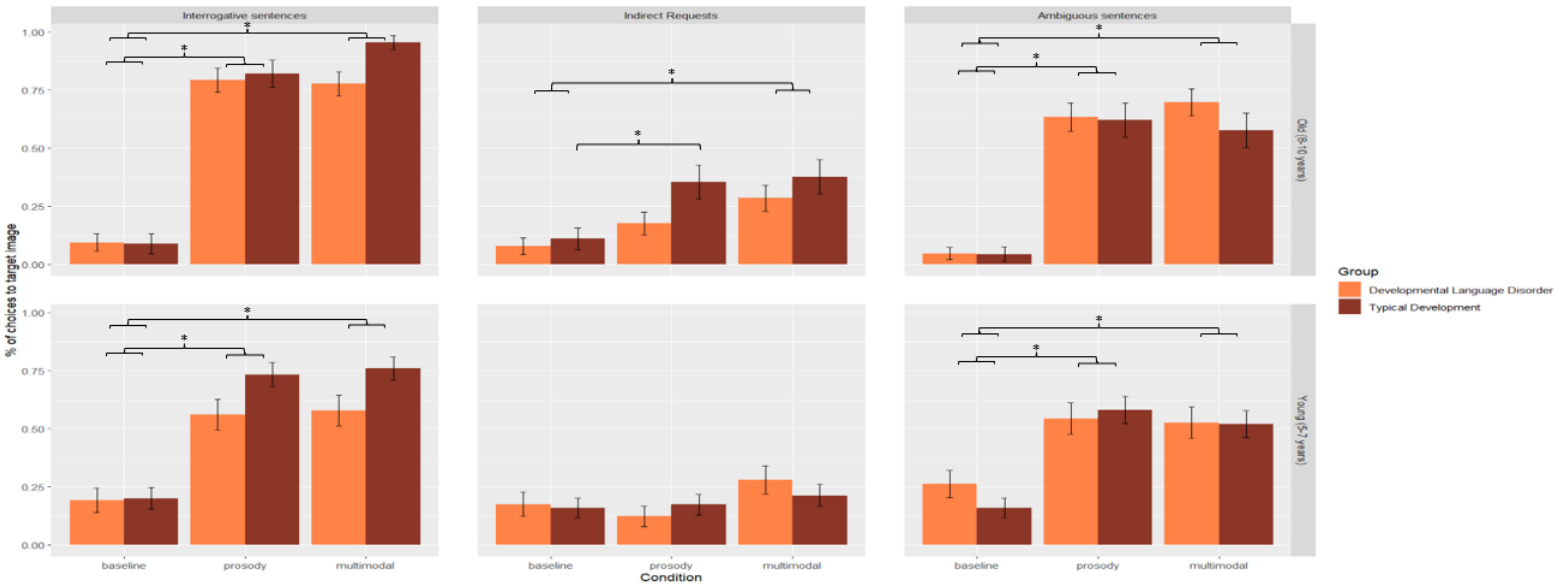


Figure 2: *Offline results for the three experimental blocks and two age groups. Y axis shows the proportion of selections to the target image for the three pragmatic meanings: interrogative sentences, indirect requests and syntactically ambiguous sentences. X axis shows the three experimental conditions: no-enhancement (baseline), prosodically-enhanced and multimodally-enhanced. * **represents a significance of p < 0.05.***



References

[1] Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. First Language, 37(1), 24–41. https://doi.org/10.1177/0142723716673953

[2] Kirk, E., Pine. K. & Ryder, N. (2011) I hear what you say but I see what you mean: The role of gestures in children's pragmatic comprehension, Language and cognitive processes, 26:2, 149-170, DOI: 10.1080/01690961003752348

[3] Li, J. P., Law, T., Lam, G. Y., & To, C. K. (2012). Role of sentence-final particles and prosody in irony comprehension in Cantonese-speaking children with and without autism spectrum disorders. *Clinical Linguistics & Phonetics*, *27*(1), 18–32. https://doi.org/10.3109/02699206.2012.734893

[4] Llanes-Coromina, J.; Vilà-Giménez, I.; Kushch, O.; Borràs-Comes, J.; Prieto, P. Beat gestures help preschoolers recall and comprehend discourse information. J. Exp. Child Psychol. 2018, 172, 168–188. [CrossRef] [PubMed]

[5] Trott, S., Reed, S., Kaliblotzky, D., Ferreira, V., & Bergen, B. (2022). The role of prosody in disambiguating English indirect requests. *Language and Speech*, 002383092210877. https://doi.org/10.1177/00238309221087715

[6] Wiedmann, N., & Winkler, S. (2015). The influence of prosody on children's processing of ambiguous sentences. *Ambiguity*. https://doi.org/10.1515/9783110403589-009

[7] Zhou, P., Crain, S., & Zhan, L. (2012). Sometimes children are as good as adults: The pragmatic use of prosody in children's on-line sentence processing. *Journal of Memory and Language*, *67*(1), 149–164. https://doi.org/10.1016/j.jml.2012.03.005

Regional variability in Russian polar question intonation

Margje Post
University of Bergen, Norway

Most polar (yes/no) questions in Russian are realised with a rising pitch accent, consisting of a high rise on the nuclear syllable, followed – if postnuclear syllables are available – by a fall to low level [1]. Recent studies [2, 3, 4] suggest that the high turning point of this rising pitch accent is moving from the nuclear to the first postnuclear syllable in the speech of younger Russians. Even small alignment differences can be perceptually relevant [5, 6] and lead to misunderstandings between the generations [2]. This difference in peak alignment might be both generational [2, 3, 4] and regional [4].

The rising pitch accent appears to dominate in polar questions also in regional varieties of Russian [7], but with variation in phonetic implementation. The rise-fall predominates even in the majority of questions with lowered questionhood, which are abundant in the interviews represented in dialect corpora. These are not pragmatically "neutral", information-seeking questions, but clearly express a different pragmatic stance. Even though most Russian polar questions are formed with the same pitch accent, we find considerable variation in, among others, pitch excursion, slope and alignment of the pitch targets, in addition to variation in cues other than F0.

With the aim to explore both regional variation in modern urban Russian speech and the role of utterance structure in the tonal configuration of polar questions, we performed a reading task among 33 adolescent male and female speakers in the capital Moscow and in Perm (Ural), a city that is known for a comparably strong local colouring in its speech. We analysed 3 questions with varying segmental and suprasegmental structures.

In our data, most renderings from both cities reach the pitch peak on the first posttonic syllable, and the average pitch peak is posttonic for all speakers (Fig. 1), like in earlier studies of today's youth, but a statistical analysis showed that on average, the young male speakers from Perm align the high targets later and use a lower F0 maximum and a narrower pitch excursion in the rise (Fig. 2) than their peers in Moscow. No significant differences could be found between the intonation of the girls from Moscow and Perm. This suggests that the female speakers use an intonation closer to a supra-local norm than the men, who tend to favour more local forms, a trend seen in other languages [8].

In order to explore the nature of the gender-specific differences, we analysed more questions and had a closer look at the interspeaker variability. Most male speakers from Moscow cluster on one end of the scale for both pitch peak alignment (Fig. 1, blue columns) and pitch span of the rise (Fig. 2), whereas their female classmates show much more interspeaker variation. Some of the Moscow girls use a large pitch range and relatively early peak alignment, with values close to those of the teacher we recorded (SP17, grey columns in Fig. 1 and 2), whom we assess to speak traditional Central Standard Russian, whereas other girls from Moscow show the later alignment values that have been recorded among today's youth, and which are also common among our speakers in Perm. This could be a sign that some Moscow girls use more conservative speech than others.

Our questions differ in segmental and suprasegmental structure, which leads to considerable differences in pitch peak alignment between the utterances. Two of them had long postnuclear stretches. Their effect on peak alignment in Russian has, to our knowledge, not been studied earlier. Possible factors influencing the alignment in our data are the presence of an unvoiced stop after the nuclear vowel [9], the quality of the nuclear vowel (long, low /a/ vs. short, high /i/), the number of postnuclear syllables and tonal repulsion, i.e., the distance between stressed syllables. The latter appears to overrule the number of syllables, but more data are needed to confirm the last finding.

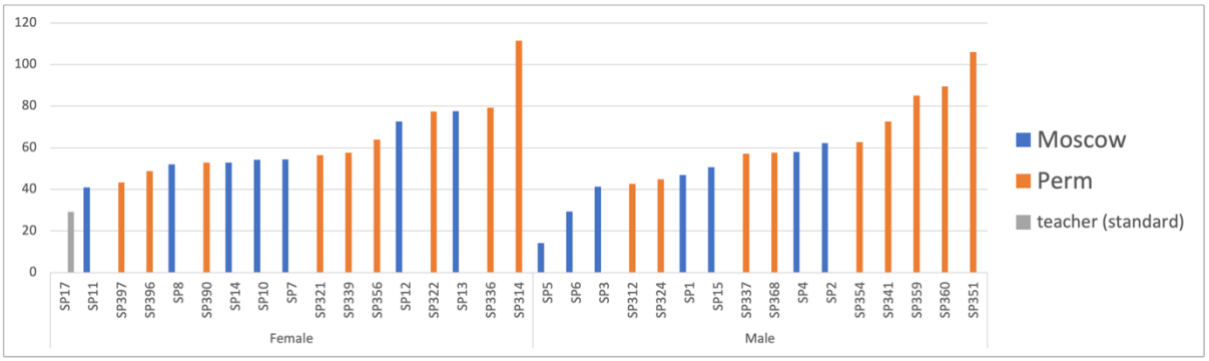


Figure 1. Pitch peak alignment in the rising pitch accent in Moscow (blue) and Perm (red columns), 16 female (left) and 16 male (right) individual adolescent speakers (mean distance in 5 questions in ms. between the onset of the first posttonic syllable and the high target). The teacher (SP17, grey column) represents traditional Central Standard Russian.

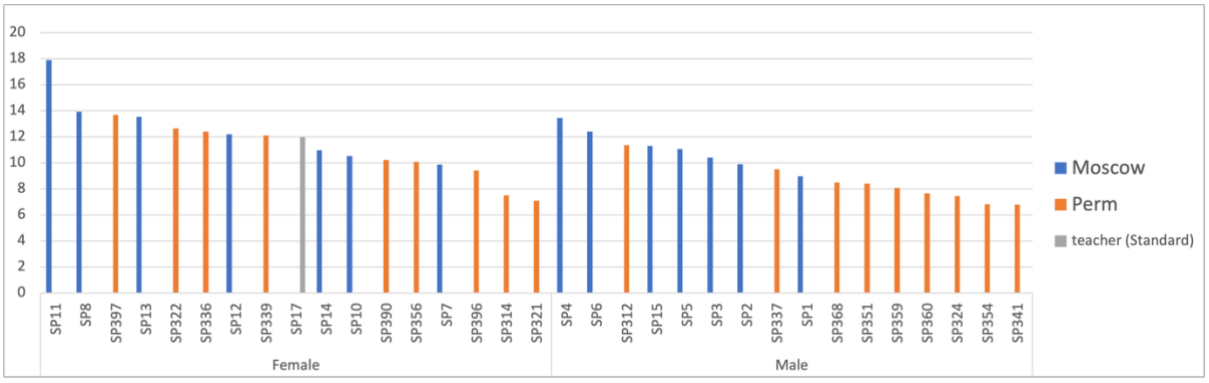


Figure 2. Pitch span between the low and high tonal target of the nuclear pitch accent in Moscow and Perm, female and male speakers (mean values of 5 questions, in semitones).

References

[1] Svetozarova, N. 1998. Intonation in Russian. In Di Cristo, A. & Hirst, D. (Eds.), *Intonation systems: A survey of twenty languages*. Cambridge University Press, 264-277.

[2] Volskaja, N. 2013. Konflikt pokolenij v zerkale russkoj intonacii. In Arxipova, A. (Ed.), *Aktual'nye voprosy teoretičeskoj i prikladnoj fonetiki.* Buki-Vedi, 53-62.

[3] Kachkovskaia, T., Zimina, S., Portnova, A. & Kocharov, D. 2022. Social variability of peak alignment in Russian rise-fall tunes. *Proceedings of Speech Prosody 2022*, 862-866. DOI: 10.21437/SpeechProsody.2022-175

[4] Grammatčikova, E., Knjazev, S., Luk'janova, L. & Požarickaja, S. 2013. Ritmičeskaja struktura slova i mesto realizacii tonal'nogo akcenta v regional'nyx variantax sovremennogo russkogo literaturnogo jazyka. In Arxipova, A. (Ed.), *Aktual'nye voprosy teoretičeskoj i prikladnoj fonetiki.* Buki-Vedi, 69-90.

[5] Prieto, P. 2011. Tonal Alignment. In van Oostendorp, M. et al. (Eds.), *The Blackwell Companion to Phonology* 2. John Wiley and Sons, 1203-1221.

[6] Odé, C. 1989. *Russian intonation: a perceptual description.* Rodopi.

[7] Post, M. 2022. Spoken corpora of spontaneous speech as a source to study polar question intonation in Russian dialects. *Computational Linguistics and Intellectual Technologies* 21, 477-487. DOI: 10.28995/2075-7182-2022-21-477-487

[8] Labov, W. 2001. *Principles of linguistic change, 2: Social factors*. Blackwell.

[9] Rathcke, T. 2008. *Komparative Phonetik und Phonologie der Intonationssysteme des Deutschen und Russischen*. Herbert Utz Verlag.

Phonological Variation in Bidialectism: A Longitudinal Corpus Study on Natural Parent-Child Interaction in Alemannic and Standard German

Aaron Schmidt-Riese[1] and Martin Pfeiffer[1]
[1]*University of Potsdam*

Phonological variation occurring naturally in conversations can fulfill many functions, one of them is to foster the co-existence of two varieties in a situation of "bidialectism" (Chevrot & Ghimenton 2019).

We investigate two phonological variables (i.e. [x]-velarization and non-realization of the diphthong [aɪ̯]) that distinguish Alemannic, a dialect of German, from Standard German. In *LEKI,* a longitudinal corpus consisting of approx. 135 hours of spontaneous interactions between parents and their children aged 1;6-4;0, we line out the language-internal and language-external factors that best explain and predict the observed phonological variation.

Unlike in Standard German, which has complementary allophones requiring a velar fricative after back low vowels (*Buch* [buːx] 'book'), but otherwise a palatal one (*ich* [ɪç] 'I'), in High Alemannic there is only the velar option (*ich* [ɪx] 'I'). Moreover, for certain frequent and grammatical lexemes there is an optional alternative of fricative deletion in Alemannic.

Similarly, the Early New High German diphthongization of long high vowels did not happen in Alemannic, which leads to realizations like [viːs] compared to Standard German [vaɪ̯s] for the word *weiß* 'white'.

In general, the results show that the contextual variable 'activity type' is the strongest predictor for the choice of either the dialectal realization or its Standard German pendant. For the parents, *reading aloud* significantly increases the probability of the Standard German variant. However, some variables like the fricative velarization seem so robust that they cannot always be suppressed successfully. For children, *pretend play* (cf. Bateman 2018) is a comparable factor for promoting standard variants. Imitating characters from the media might be a suitable explanation (cf. Al-Harbi 2015). Due to the longitudinal design of our corpus, we are also able to analyze the children's development in terms of phonological variation.

However, in single lexemes, language-internal factors are also crucial for the selection of a variety. For instance, *ich* 'I' is more likely to appear with dialectal fricative deletion when located in the syntactic position after the final verb, which is called *Wackernagel* position. This is in line with previous research on German sentence structure, since this position is generally known for clitics (cf. Weiß 2018, Itô & Mester 2019).

Another finding is that phonological variation can serve for functional differentiation, too. This is the case for *gleich,* a syncretism in Standard German. In our Alemannic data, it is produced with word-final velar fricative when used as an adjective ([gliːx] meaning 'equal') but with deletion when used as a temporal adverb ([gliː] meaning 'soon', 'in a minute').

Furthermore, rare cases of over-generalizations and co-occurrence violations that occur exclusively in children can be observed in the data. An example is the realization of *vielleicht* 'maybe' as [fɪ'liːçt], i.e. containing the dialectal monophthong, but lacking the then obligatory fricative velarization, leading to an unacceptable form in both target varieties, which would either be [fɪ'laɪ̯çt] in Standard German or [fɪ'liːxt] in Alemannic (cf. Auer 1997). This is especially interesting, since it gives insight into the cognitive organization of two competing varieties. Thus, adding to the growing body of research in this field (e.g. Roberts 2005, Katerbow 2013, Chevrot & Foulkes 2013, de Vogelaer & Katerbow 2017), we hope to contribute additional observations on the role sociolinguistic variation plays in first language acquisition.

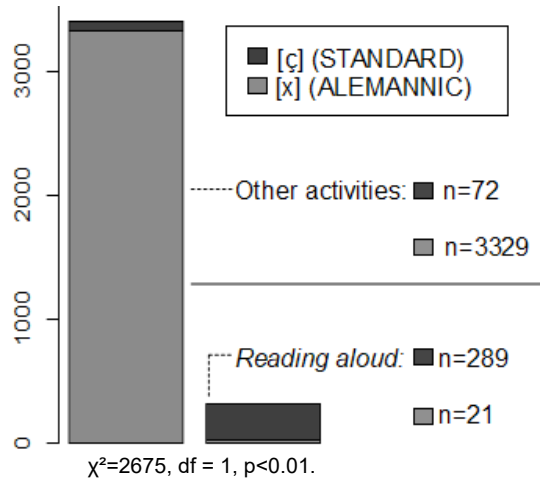


Figure 1: *Activity of 'Reading aloud' as factor for Standard in parental output.*

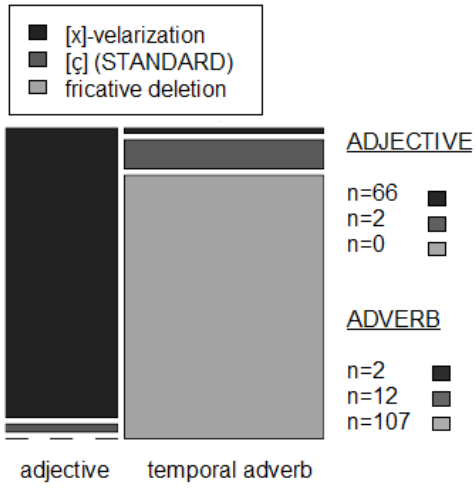


Figure 2: *Functional differentiation of German 'gleich' into [gli:] and [gli:x].*

References (in order of appearance)

[1] Chevrot, J.-P. & Ghimenton, A. 2019. Bilingualism and bidialectalism. In de Houwer, A. & Ortega, L. (eds.), *The Cambridge handbook of bilingualism.* Cambridge: Cambridge University Press, 510-523.

[2] Bateman, A. 2018. *Conversation Analysis and Early Childhood Education: The Co-Production of Knowledge and Relationships*. London: Routledge.

[3] Al-Harbi, S. 2015. The Influence of Media in Children's Language Development. *Journal of Educational and Developmental Psychology 5*, 1-5.

[4] Weiß, H. 2018. The Wackernagel complex and pronoun raising. In Jäger, A., Ferraresi, G. & Weiß, H. (eds.), *Clause structure and word order in the history of German*, 132-154.

[5] Itô, J. & Mester, A. 2019. Match as syntax-prosody MAX/DEP: Prosodic enclisis in English. *English Linguistics 36*(1). 1–22.

[6] Auer, P. 1997. Co-occurrence Restrictions between Variables. A Case for Social Dialectology, Phonological Theory, and Variation Studies. *Variation, Change and Phonological Theory 146*, 69–99.

[7] Roberts, J. 2005. Acquisition of sociolinguistic variation. In Ball, M. (ed.), *Clinical Sociolinguistics*. Oxford: Blackwell, 152-164.

[8] Katerbow, M. 2013. *Spracherwerb und Sprachvariation: Eine phonetisch-phonologische Analyse zum regionalen Erstspracherwerb im Moselfränkischen.* Berlin: deGruyter.

[9] Chevrot, J.-P. & Foulkes, P. 2013. Language Acquisition and Sociolinguistic Variation. *Linguistics 51(2)*, 251-254.

[10] De Vogelaer, G. & Katerbow, M. 2017. *Acquiring Sociolinguistic Variation.* Studies in Language Variation. Amsterdam: John Benjamins.

Corpus

Pfeiffer, M. & Anna, M. *Longitudinalkorpus Eltern-Kind-Interaktion (LEKI)*, hostet in Multimodal Oral Corpora Administration [moca] at the Hermann-Paul-Centre for Linguistics at the Albert-Ludwig University of Freiburg (Germany).

Statistical Program

R & R studio: Open Source Software. https://www.r-project.org/ (23th of November 2022).

Echoes of the past: Observations on word prosody in Kamas

Alexandre Arkhipov
Universität Hamburg

The paper examines word prosody patterns in Kamas, an extinct Samoyedic (< Uralic) language [1]. It used to be spoken in Southern Siberia; the last speaker, Klavdiya Plotnikova, died in 1989. Kamas had extensive contacts with Siberian Turkic languages, which probably lasted for a few centuries. The next major contact language was Russian; in the 20th century, the remnants of the Kamas community shifted to Russian completely. Klavdiya Plotnikova had most probably not fully acquired Kamas in her childhood and was recorded after many years of not actively using the language. Her Kamas, termed 'post-shift Kamas' by G. Klumpp [1], is affected by language attrition as well as Russian influence on all levels, however Kamas phonetics is relatively intact in her speech.

This study is initially based on the INEL Kamas Corpus [2] which contains transcripts of all known recordings of Plotnikova (ca. 14 hours total; see [3] for details). However, most of the data belong to a smaller subset of recordings, time-aligned at word and segment level within the DoReCo project [4], with annotations further adapted and manually corrected by the author.

Kamas, as the other Uralic languages, does not have lexical tone. Word stress patterns have been documented by Donner [5] in the 1910s. According to him, the primary stress would typically fall on the 1st syl. but could shift to the 2nd syl. if it was heavy (long vowel or closed). In disyllabic words, stress could fall on either syllable. Under Turkic influence, it could also go word-final, not only in Turkic loans but also in native lexemes. The phonetic substance of stress in not very clear. It is qualified by Donner [5: 126] as 'expiratory accent'; Turkic word-final stress is however generally characterized as high pitch accent [6: 781–787].

Klumpp [1] cites several factors behind vowel duration: 1) lexical vowel length (e.g. due to contractions), 2) automatic lengthening before /r/, 3) automatic lengthening of non-high ('open') vowels in open syllables (at morpheme boundaries), 4) original stress in Russian loans.

I investigate syllable prominence patterns in Plotnikova's data in terms of duration and pitch. In order to make the material relatively comparable, the main focus is on disyllabic nominal stems which are not Russian loans (see [7] for specifics of their behaviour), i.e. are either native Kamas words or well-integrated Turkic loans. Both bare and inflected forms are considered; inflected forms may contain case, number and/or possessive suffixes.

Pitch prominence (tonal accent), if present, tends to the last syllable independently of lexical vowel length. Specifically, in words with lexically long vowel in 1st syl. like *büːz'e* 'man' (Fig. 1) the long vowel typically exhibits a sustained pitch level. If there is a change in pitch, it happens on the last syllable (cf. high falling pitch in Fig. 2 and high level pitch in Fig. 3). However, a falling tone can also be found on the 1st syl., especially after voiceless (*tibi* 'man').

I suggest that the tonal contour of the accent is determined by higher-level (phrasal) prosody, while its position within the word depends on its morphological structure. Various inflectional suffixes behave differently with respect to the tonal accent (marked with acute: ˊ). The plural marker *-ʔi ~ -ʔjə* is normally accented, as in *büːz'e-ʔí* 'man-PL' (Fig. 4). When followed by a case suffix, it remains accented in ca. half of occurrences (e.g. *ine-ʔí-zi ~ ine-ʔi-zí* 'horse-PL-LAT'). The other plural marker, *-zaŋ/-zeŋ*, is only accented when used alone (*tibi-zéŋ* 'man-PL'); when it is followed by a case or possessive suffix, the accent is on the latter (*tura-zaŋ-dá* 'house-PL-POSS.3SG'). Plural suffixes, esp. *-ʔi ~ -ʔjə*, do not trigger vowel lengthening.

In turn, many case suffixes do lengthen root-final non-high vowels. Lative *-nə/-də* is often accented after a root-final high vowel (which does not undergo lengthening), e.g. *tibi-ná* 'man-LAT'. When it follows a non-high vowel (with lengthening), the accent can be found either on this vowel or on the suffix (*nükéː-nə ~ nükeː-ná* 'woman-LAT'). Instrumental *-zi* and Locative *-gən* behave similarly but are more frequently accented even after non-high vowels.

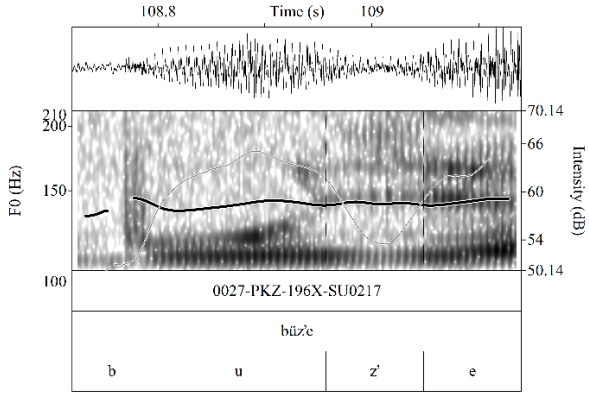


Figure 1. *büz'e* 'man': no tonal accent

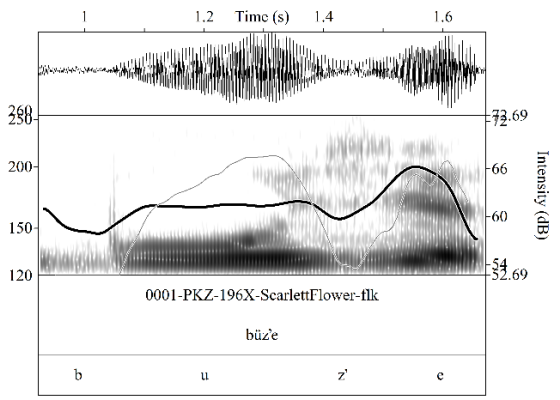


Figure 2. *büz'e* 'man': falling accent

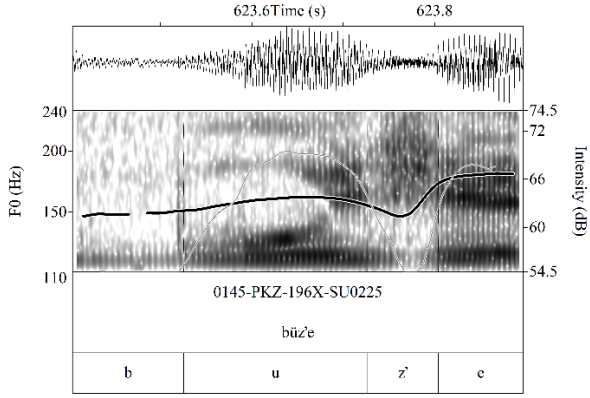


Figure 3. *büz'e* 'man': tonal upstep

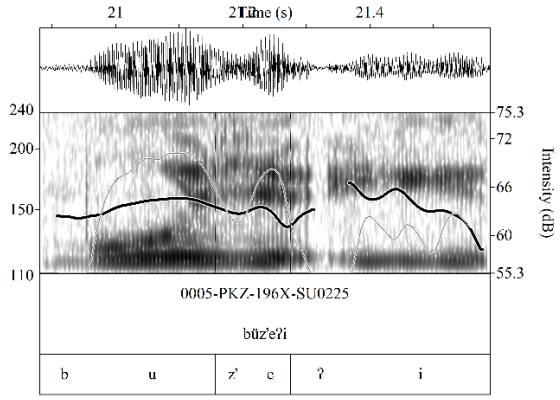


Figure 4. *büz'e-ʔi* 'man-PL': falling accent on the plural suffix

Solid black line shows pitch, thin grey line shows intensity.

References

[1] Klumpp, G. 2022. Kamas. In: Bakró-Nagy, M., Laakso, J., & Skribnik, E. (eds.). *The Oxford Guide to the Uralic Languages*. Oxford: OUP, 817–843. DOI: 10.1093/oso/9780198767664.003.0039

[2] Gusev, V.; Klooster, T.; Wagner-Nagy, B. 2019. *INEL Kamas Corpus*. Version 1.0. Publication date 2019-12-15. http://hdl.handle.net/11022/0000-0007-DA6E-9. Archived in Hamburger Zentrum für Sprachkorpora.

[3] Arkhipov A., Däbritz C. L., Gusev V. 2020. *User's Guide to INEL Kamas Corpus.* Working Papers in Corpus Linguistics and Digital Technologies: Analyses and Methodology 3. Szeged & Hamburg. DOI: 10.14232/wpcl.2020.3

[4] Gusev, V., Klooster, T., Wagner-Nagy, B., & Arkhipov, A. 2022. Kamas DoReCo dataset. In: Seifart, F., Paschen, L. & Stave, M. (eds.). *Language Documentation Reference Corpus (DoReCo) 1.1*. Berlin & Lyon. https://doreco.huma-num.fr/languages/kama1351 (Accessed on 05.12.2022). DOI: 10.34847/nkl.cdd8177b

[5] Joki, A. J. 1944. *Kai Donners Kamassisches Wörterbuch nebst Sprachproben und Hauptzügen der Grammatik*. Lexica Societatis Fenno-Ugricae VIII. Helsinki: Suomalais-Ugrilainen Seura.

[6] Johanson, L. 2021. *Turkic*. (Cambridge Language Surveys). Cambridge: CUP. DOI: 10.1017/9781139016704

[7] Arkhipov, A. (in press) Two morphologies, two stress systems, shaken, not stirred: Number marking on Russian borrowings in Kamas.

Exploring individual strategies in prosodic marking of information status in German

Janne Lorenzen[1], Simon Roessig[2] and Stefan Baumann[1]
[1]IfL Phonetik, University of Cologne, [2]Cornell Phonetics Lab, Cornell University

The ubiquity of individual variability in prosody production and perception has been acknowledged for some time. Recent studies have shown that speakers differ in which phonetic parameters they use to distinguish focus types [1] or encode informativity [2]. At the same time, listeners choose to pay attention to different cues when interpreting prosody in terms of information structure [1] or prominence [3]. Information status constitutes another pragmatic property that is encoded in German via multiple prosodic parameters. Generally, it is assumed that *new* referents are produced prosodically most prominently (i.e., with longer durations, more extensive f0 movements and higher intensity), while *given* referents are least prominent with *accessible* referents falling in between these two extremes [4].

In the present study we ask whether individual speakers follow the same or different strategies when encoding information status prosodically. To address this question, we collected data in a reading task. Participants read eight short stories with two target words in direct and indirect object roles that were systematically varied to be either *discourse-new* or *accessible* through the context but always in broad focus (e.g., in (1), where the waiter becomes *accessible* through the mention of the restaurant in the preceding sentence). To create an interactive scenario and thus prevent monotonous speech, a confederate was present, who was tasked to sort picture cards corresponding to the stories into the correct order. We recorded 32 native speakers of German (24f/8m, aged 20-38 years) producing 512 (accented) target words.

For each target word, we measured word and syllable duration, tonal onglide, periodic energy mass, Delta F0, and synchrony (the latter three from the ProPer toolbox [5]). Duration was measured only in (the accented syllable of) non-phrase-final target words to avoid effects of final lengthening. Periodic energy mass quantifies the area under the periodic energy curve integrating duration and intensity. Synchrony captures the extent and shape of the f0 movement within a syllable, while Delta F0 measures the difference in f0 between an accented syllable and the preceding one. Tonal onglide was measured as the difference in semitones between the tonal target of a pitch accent and the preceding "non-accent tone", based on annotated labels following DIMA [6]. Each parameter was entered into a separate Bayesian mixed effects model as the dependent variable using information status as predictor, random intercepts for word and speaker and by-speaker random slopes for the effect of information status. To investigate different strategies, we extracted the random slope coefficients from each model and performed a hierarchical cluster analysis on those. This way, we analyzed individual patterns predicted by the models and identified groups of similarly behaving speakers in an objective manner.

Figure 1 shows the mean random slope coefficients for every parameter across the members of each of the three clusters as well as the average values across all speakers. Positive values indicate that the cue in question has higher values in *new* than *accessible* referents and thus is used by this group of speakers to mark information status in the expected way. Across all speakers, Delta F0 is the most successful cue in distinguishing *new* from *accessible* referents, followed by tonal onglide. Periodic energy does not appear to be a reliable cue. The cluster analysis reveals three main strategies: Cluster 1 (11 speakers) marks information status only weakly, with Delta F0 constituting the best cue. Cluster 2 (4 speakers), by contrast, makes clear use of all cues, with an emphasis on (word) duration. Cluster 3 (14 speakers) again relies mostly on f0-based cues, while duration and mass seem to not or only barely mark information status.

While the overall averages capture the behavior of many speakers well (e.g., those in Cluster 3), other viable strategies, such as the one employed by Cluster 2 in which speakers prefer duration over most f0-based cues, are obscured. It is thus important to consider speaker-specific strategies to gain a deeper understanding of the prosodic marking of information status.

(1) *Claudia hatte einen erfolgreichen Tag im Geschäft.*
'Claudia had a successful day at the shop.'
Sie hat einige Kollegen vom Restaurant nebenan für viele ihrer Waren begeistern können.
'She was able to interest some colleagues from the restaurant next door in many of her goods.'
*Unter anderem hat sie dem **Kellner** eine **Vase** verkauft.*
'Among other things she sold the **waiter** a **vase**.'
Jetzt geht sie nach Hause und legt sich sofort hin.
'Now she is going home and lying down immediately.'

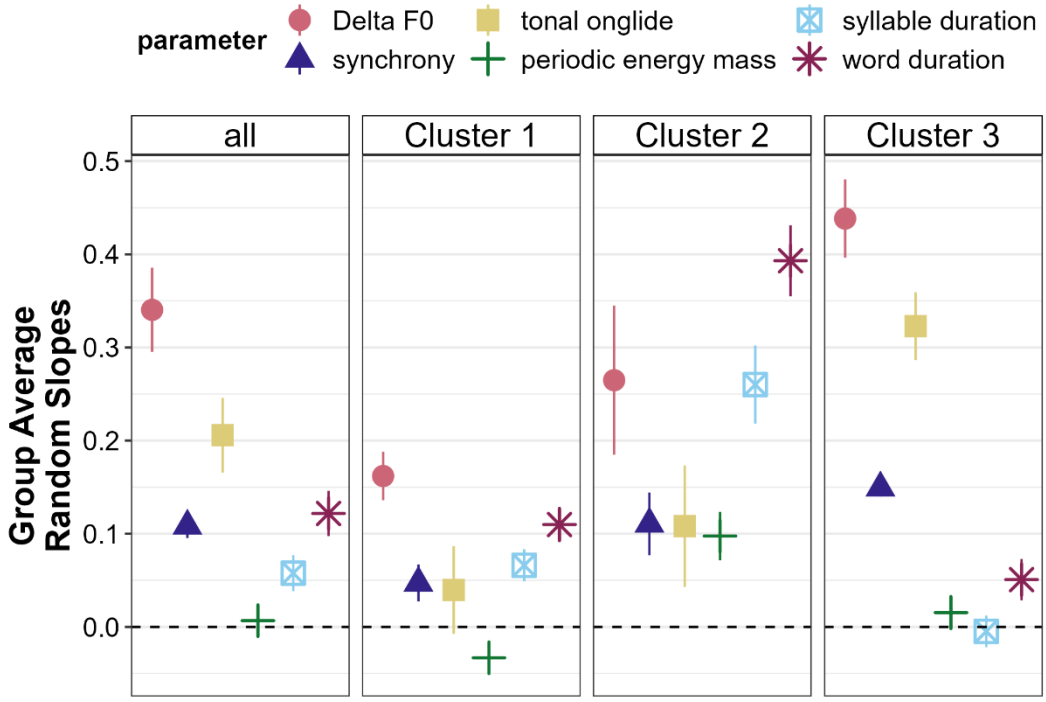


Figure 1. *Average random slopes from the Bayesian models in clusters (speaker groups) with standard errors. The leftmost plot shows the data for all speakers combined.*

References

[1] Cangemi, F., Krüger, M. & Grice, M. 2015. Listener-specific perception of speaker-specific productions in intonation. In Fuchs, S., Pape, D., Petrone, C., & Perrier, P. (Eds.), *Individual differences in speech production and perception.* Peter Lang, 123–145.

[2] Ouyang, I. C. & Kaiser, E. 2015. Individual differences in the prosodic encoding of informativity. In Fuchs, S., Pape, D., Petrone, C., & Perrier, P. (Eds.), *Individual Differences in Speech Production and Perception*. Peter Lang, 147-189.

[3] Baumann, S. & Winter, B. 2018. What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics* 70, 20–38.

[4] Baumann, S. & Riester, A. 2013. Coreference, lexical givenness and prosody in German. *Lingua* 136, 16–37.

[5] Albert, A., Cangemi, F. & Grice, M. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Proc. 9th International Conference on Speech Prosody* (Poznań), 804-808.

[6] Kügler, F., Baumann, S. & Röhr, C. T. 2022. Deutsche Intonation, Modellierung und Annotation (DIMA) – Richtlinien zur prosodischen Annotation des Deutschen. In Schwarze, C. & Grawunder, S. (Eds.), Transkription und Annotation gesprochener Sprache und multimodaler Interaktion. Tübingen: Narr, 23–54.

On the role of glottal stop: from boundary marker to correlate of focus. An experimental study on native and non-native Italian and French

Bianca Maria De Paolis, Bianca Abbà, Valentina De Iacovo, LFSAG, Italy

Glottalization of word-initial or word-final segments has been documented as a common marker of prosodic boundaries in many languages across the world [1]. Even though glottal consonants are not part of the phonological inventory of Italian and French, several acoustic analyses have been conducted on the presence of glottalization in these two languages as well (see [2, 3] among others), often linking the presence of non-modal phonation to the presence of phrase boundaries or pauses ([4] for Italian, [5] for French). In addition to that, some authors have stated that glottalization can occur in presence of pitch accents and "emphatic words" ([6] for English, [7] for French and Italian). These observations seem to suggest a link between *in-situ* focalization and its traditionally-described prosodic correlates (i.e. pitch accent), and glottalization. To our knowledge, though, no study has verified this hypothesis by systematically comparing the two situations, focus vs non-focus, and their respective co-occurrence with word-final or word-initial glottalization.

Our purpose is thus to test this hypothesis, namely on the aforementioned languages (Italian and French), using an experimental and data-driven approach. The analysis is conducted on a corpus of task-elicited speech, based on [8] collection model, in which participants are shown two short comic strips and then asked questions about the scene illustrated. The questions elicit answers in three informative contexts: neutral broad focus, narrow identification focus and corrective focus. The choice to use semi-spontaneous speech is motivated by the fact that focus is a phenomenon that strictly belongs to communicative interaction, and no real communication between speakers can take place in fully controlled speech. The lack of a precise script certainly implies a less than perfect balance between the occurrences of the target structures; despite this, the number of observations that can be made from our data corpus is largely sufficient to enable reliable statistics. The corpus currently contains a total of 2175 utterances, produced by 75 French and Italian native and non-native speakers aged between 18 and 39.

We will show that in our data of native speech word-initial glottalization is much more likely to occur for constituents under narrow focus (see Figure 1). This relation could be expressed in the following way:

(1) occurrence of GL = phrase level < phrase level + focus

Moreover, we will argue that, unlike [6] and [7], in our data glottalization does not occur as a "side-effect" of pitch accents: instead, it often surfaces as the only phonetic correlate of *in-situ* focalization, even in absence of major f0 movements (see Figure 2). In our data, this is especially valid for utterance-final phrases, in which the presence of glottal stops or creaky voice could be due to physiological reasons (see [9]): for example, the maintenance of subglottal pressure, or as a "signal of hard work". To further support our claim, we will give insights from an ongoing work, dealing with the same interplay of glottalization and focus in non native speech (Italian L2 and French L2).

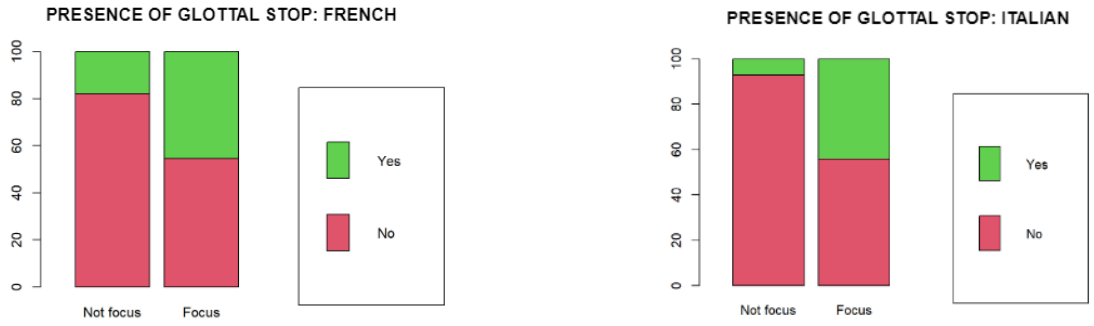


Figure 1. *Proportion of occurrences of word-initial glottalization in the two contexts, non-focus vs focus, for the French (left) and Italian (right) groups.*

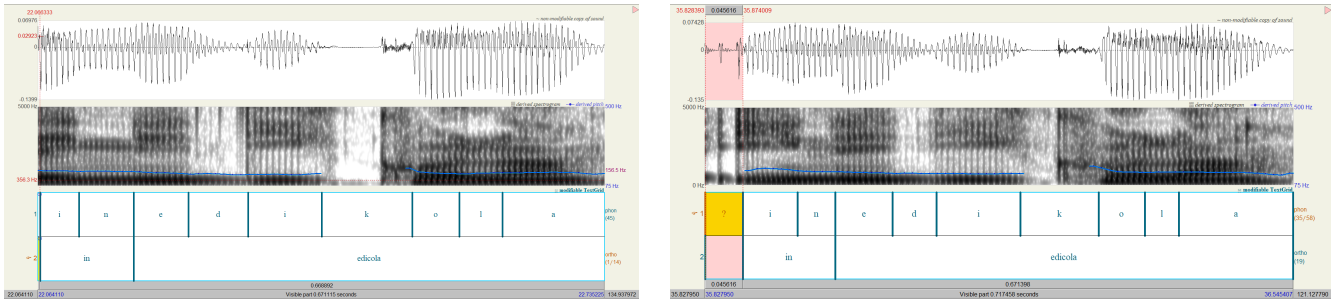


Figure 2. *Spectrograms of the phrase "in edicola" ("at the newsstand") as uttered by an Italian speaker in a neutral broad focus context (left) and corrective focus context (right).*

References

[1] Gordon M., Ladefoged P. 2001. Phonation types: a cross-linguistic overview, *Journal of Phonetics*, 29: 383–406.

[2] Delattre P. 1971. Pharyngeal features in the consonants of Arabic, German, Spanish, French and American English. *Phonetica*, 54: 93–108.

[3] Contini M., Carpitelli E., Romano A. 2005. Des occlusives glottales dans l'espace roman, *Estudios Ofrecidos a A. Quilis*, 1 : 127–145.

[4] Stevens M., Hajek J. 2006. Blocking of word-boundary consonant lengthening in Sienese Italian: some auditory and acoustic evidence, *Proceedings of the 11th Australasian International Conference on Speech Science and Technology 2006*: 176–181.

[5] Fougeron C. 2001. Articulatory properties of initial segments in several prosodic constituents in French, *Journal of Phonetics*, 29: 109–135.

[6] Cho T., Keating, P. 2009. Effects of initial position versus prominence in English, *Journal of Phonetics*, 37 , 466–485.

[7] van Santen J., D'Imperio, M. 1999. Positional effects on stressed vowel duration in Standard Italian, *Proceedings of the 14th ICPhS, San Francisco*, 1: 241–244.

[8] Gabriel C. 2010. On focus, prosody, and word order in Argentinean Spanish: a minimalist OT account, *ReVEL*, 4, 183–222

[9] Lennes M., Aho E., Toivola M., Wahlberg L. 2006. On the use of glottal stop in Finnish conversational speech, *Fonetiikan päivät - The Phonetics Symposium*, 93–102.

In L2 phonetic training, learners benefit more with hand gestures cueing visually salient articulatory information

Xiaotong Xi[1], Peng Li[2] and Pilar Prieto[3,1]

[1]*Universitat Pompeu Fabra,* [2]*University of Oslo,* [3]*Institució Catalana de Recerca i Estudis Avançats*

Previous studies have shown that hand gestures encoding phonetic features can help L2 segmental learning (e.g., [1]). However, mixed effects have been found on the role of hand gestures encoding articulatory features of segments [2]. Moreover, the role of visibility of the target articulatory features (e.g., a non-visible tongue position versus a visible mouth shape) remains largely unknown. To assess whether the degree of visual saliency of the articulatory features encoded by hand gestures may influence the phonetic training effects, the present study trained Catalan learners of English on the learning of L2 English vowel pairs /æ-ʌ/ and /i-ɪ/. While /æ/ is produced with a larger lip aperture and a more fronted tongue position than /ʌ/ [3], /i/ is pronounced with a slightly wider lip spreading and a higher tongue position than /ɪ/ [4]. With respect to visibility, while the difference in tongue position between the two vowels within each pair is non-visually accessible, the difference in lip aperture and lip spreading is visually accessible. In a pilot study, when native English speakers were asked to identify the target vowels in minimal pairs with visual-only input, the identification accuracy for /æ/-/ʌ/ was 84.62%, and for /i/-/ɪ/ was 59.49%. This difference demonstrates that the articulatory difference in lip aperture between /æ/ and /ʌ/ is more salient than the lip spreading differences between /i/ and /ɪ/. To highlight these crucial articulatory features, we designed two types of hand gestures for each vowel contrast: lip hand gesture and tongue hand gesture (see Figure 1). The lip hand gesture highlighted the visually salient difference in the lip aperture between /æ/ and /ʌ/, as well as the visually less salient difference in the lip spreading between /i/ and /ɪ/. The tongue hand gesture showed the non-visible difference in tongue position for all vowels. We hypothesized that the decrease in visual accessibility of the features encoded by hand gestures (lip hand gesture encoding lip aperture > lip hand gesture encoding lip spreading > tongue hand gestures) will affect the training outcomes, in the sense that the more visually accessible the features are, the stronger the pronunciation learning effects are observed.

Ninety-nine Catalan learners of English were trained on the two vowel pairs /æ/-/ʌ/ and /i/-/ɪ/ in two 15-minute sessions either under no gesture (NG, n = 33), lip hand gesture (LG, n =33), or tongue hand gesture (TG, n = 33) conditions by watching training videos. In the NG group, the instructor produced minimal pair words and sentences containing the target words contrasting in /æ/-/ʌ/ or /i/-/ɪ/ without any gestures. In the LG group, the instructor performed hand gestures mimicking lip aperture for /æ/-/ʌ/ and lip spreading for /i/-/ɪ/. In the TG group, the instructor produced hand gestures mimicking tongue frontness/backness for /æ/-/ʌ/ and tongue height for /i/-/ɪ/. Before, immediately after, and one week after the training, participants' production was tested through a paragraph reading, a picture naming, and a word imitation task. The pronunciation accuracy was assessed using the Pillai score which measured the acoustic overlap between two categories. Results showed that LG had more benefits in the pronunciation of /æ/ and /ʌ/ compared to NG and TG in the paragraph reading and picture naming tasks. This suggests that hand gestures encoding articulatory features can benefit more in the pronunciation learning of L2 sounds when they encode visually salient articulatory features, such as the lip aperture differences between /æ/ and /ʌ/.

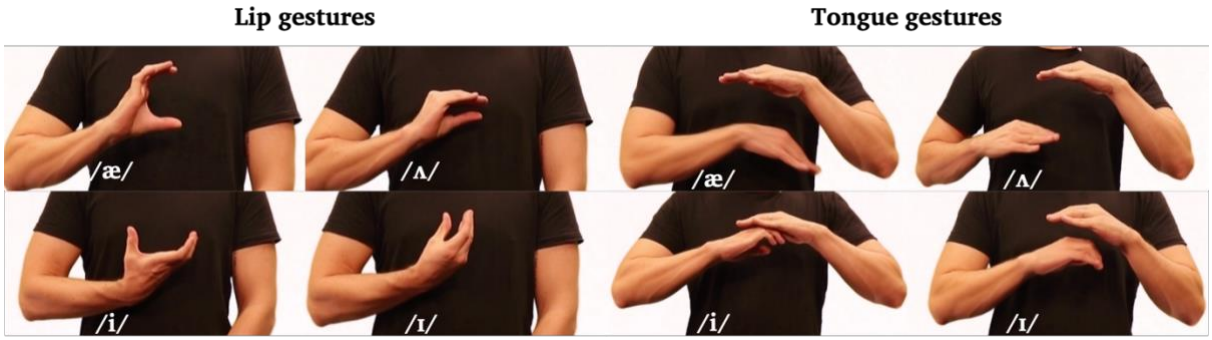


Figure 1. *Hand gestures encoding articulatory features for English vowels. From left to right in the upper panel: lip gestures for /æ/ and /ʌ/, and tongue gestures for /æ/ and /ʌ/. From left to right in the lower panel: lip gestures for /i/ and /ɪ/, and tongue gestures for /i/ and /ɪ/.*

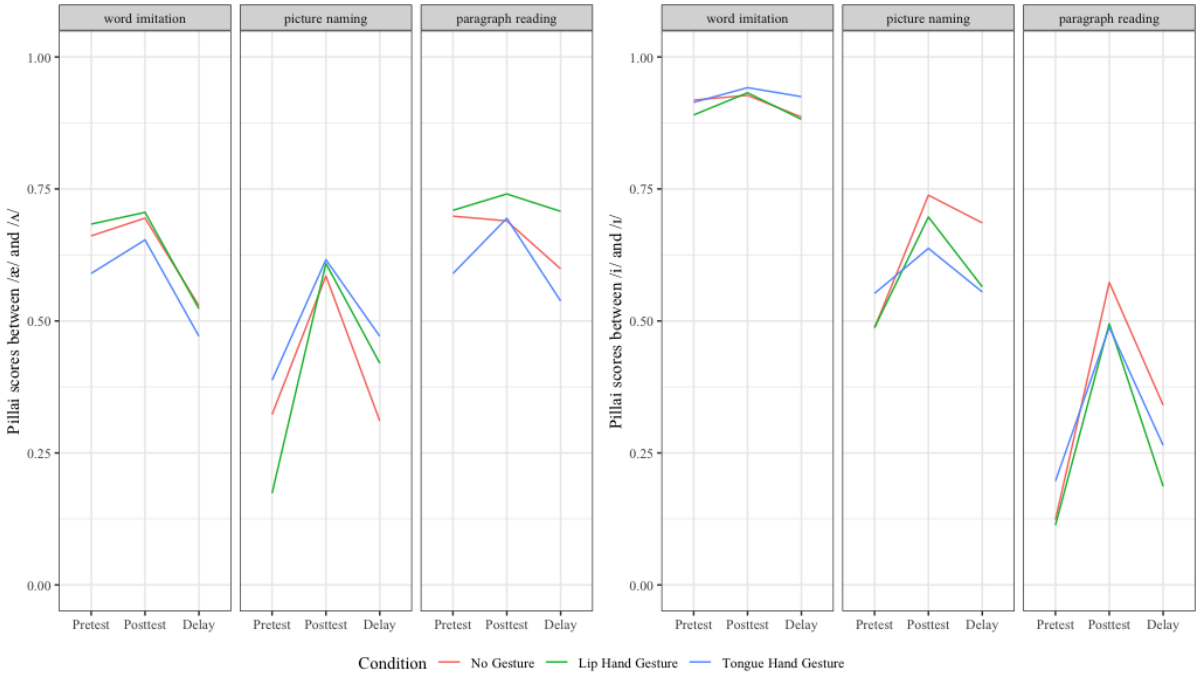


Figure 2. Pillai scores of the vowel pairs */æ/-/ʌ/ (left panel) and /i/-/ɪ/ (right panel) across conditions and tests in the three production tasks.*

References

[1] X. Xi, P. Li, F. Baills, and P. Prieto, "Hand gestures facilitate the acquisition of novel phonemic contrasts when they appropriately mimic target phonetic features," *J. Speech, Lang. Hear. Res.*, vol. 63, no. 11, pp. 3571–3585, 2020, doi: 10.1044/2020_JSLHR-20-00084.

[2] M. Hoetjes and L. van Maastricht, "Using gesture to facilitate L2 phoneme acquisition: the importance of gesture and phoneme complexity," *Front. Psychol.*, vol. 11, p. 575032, 2020, doi: 10.3389/fpsyg.2020.575032.

[3] J. P. Zerling, "Frontal lip shape for French and English vowels," *J. Phon.*, vol. 20, no. 1, pp. 3–14, 1992, doi: 10.1016/s0095-4470(19)30249-9.

[4] S. A. J. Wood, "A radiographic and model study of the tense-lax contrast in vowels," in *Phonologica 1988*, 1992, pp. 283–291.

Hosted by

Radboud Universiteit
SINCE 1923

CLS | Centre for Language Studies
Radboud University

Sponsors

International Speech Communication Association (ISCA)

Association for Laboratory Phonology (LabPhon)

lab phon association for laboratory phonology

Dutch Association for Phonetic Sciences (NVFW)

NVFW
Nederlandse Vereniging voor Fonetische Wetenschappen

Contact

Email address: pape2023.conference@gmail.com
Conference location: Berchmanianum, Houtlaan 4, Nijmegen