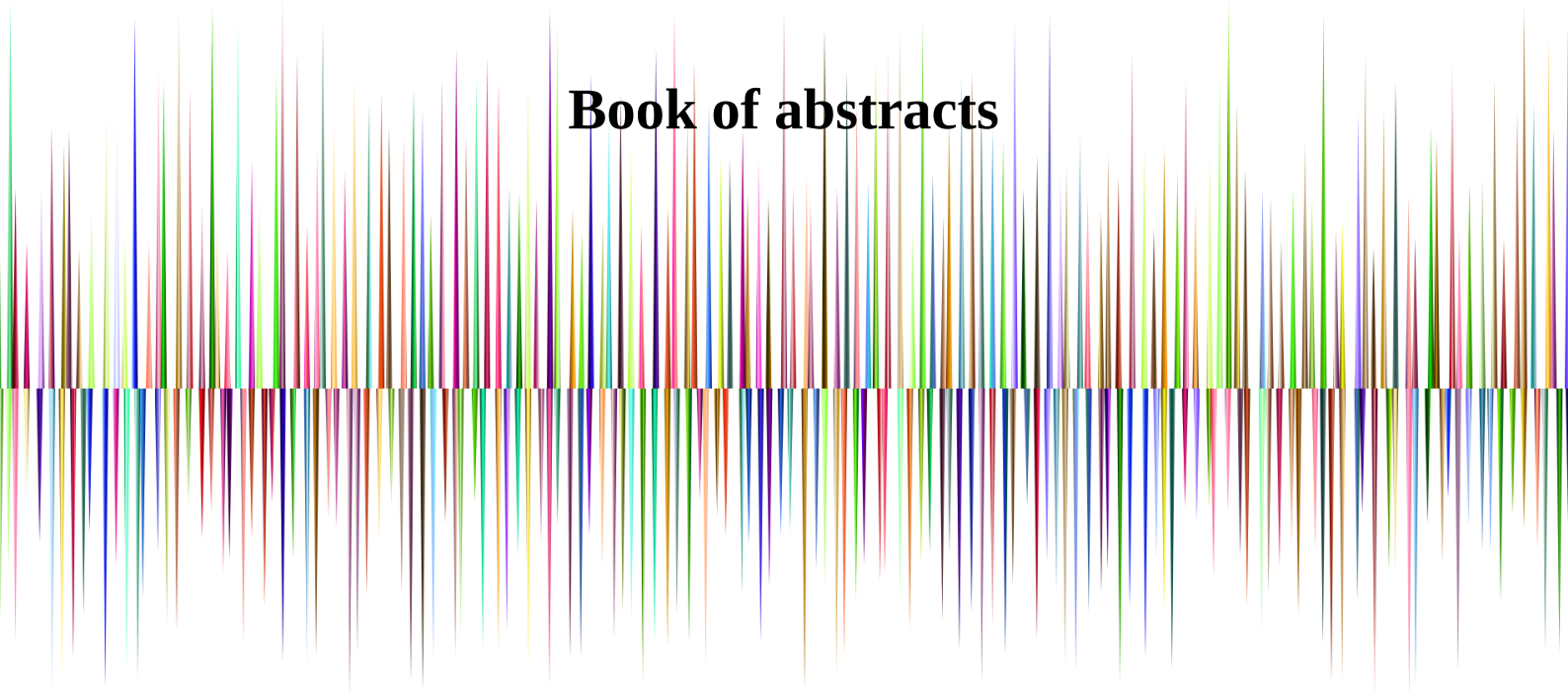
Prosody at the crossroads of disciplinary pathways

Université Grenoble Alpes, CLV building

21-22 May 2026



Book of abstracts

Table des matières

Day 1: Thursday, May 21st

08:30 Registration and welcome coffee

09:10 Opening ceremony and welcome address

09:15 Plenary 1

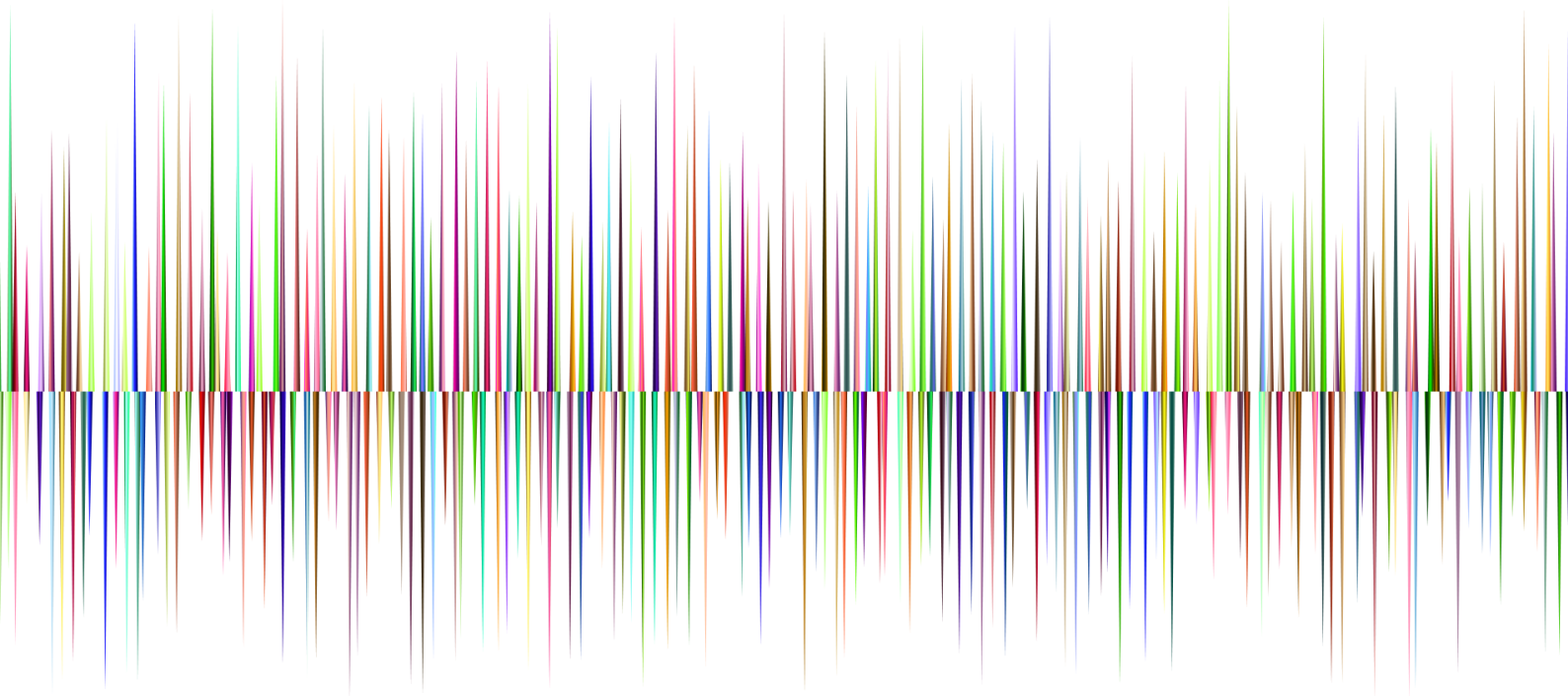
Eugenia Sansegundo (Consejo Superior de Investigaciones Científicas, Madrid) *Voice Quality in Forensic Phonetics*

10:25 Parallel sessions 1 (Rooms 1, 2, 3)

Room 1: Anna Kohári & Fruzsina Csébfalvi (ELTE Research Centre for Linguistics, Hungary / Radboud University, Netherlands) *Boundary marking by creaky voice across sentence types*

Room 2: Vieri Samek-Lodovici (University College of London, United Kingdom) *On the Role of Givenness in Prosodic Prominence*

Room 3: Bert Remijsen (The University of Edinburgh, United Kingdom) *Voice quality and tone as independent dimensions of contrast in Dinka*



Keynote speaker

Voice Quality in Forensic Phonetics: When, How, and Why?

Eugenia San Segundo Fernández

(Consejo Superior de Investigaciones Científicas, Madrid)

Forensic phonetics applies phonetic knowledge to solve legal problems, analyzing speech as evidence with a focus on speaker identity. Since voice quality is a powerful carrier of identity, it is particularly valued in tasks such as speaker profiling, the design of voice line-ups, and forensic voice comparison. This keynote introduces the field of forensic phonetics and its main areas of application, with a focus on forensic voice comparison, where recordings of an unknown offender and a suspect are analyzed to determine whether they belong to the same speaker. Forensic practitioners draw on both auditory and acoustic methods, considering not only segmental features but also suprasegmental and prosodic characteristics, including voice quality.

The second part of the talk centers on voice quality, a key forensic parameter that has received increasing attention in recent years. For instance, it is included in the Best Practice Manual for Forensic Speaker Comparison by the European Network of Forensic Science Institutes (ENFSI). Two complementary perspectives are considered regarding voice quality: a narrow definition, focusing on vocal fold activity (typically analyzed through sustained vowels), and a broader one, where voice quality emerges from the interaction of laryngeal and supralaryngeal features. Particular attention is given to one commonly used protocol for the perceptual evaluation of voice quality: the Vocal Profile Analysis (VPA) scheme and its simplified versions, which have proven especially valuable in applied contexts and computer-assisted analyses. Drawing on an international survey of practitioners, the keynote highlights how voice quality is currently conceptualized and used in forensic casework.

The final part briefly addresses the growing challenge of voice deepfakes in forensic phonetics. As synthetic speech becomes more realistic, voice quality must be reconsidered as a cue to identity, calling for closer integration between phonetic expertise and emerging technologies.

REFERENCES

de Jong-Lendle, G., Nolan, F., McDougall, K., & Hudson, T. (2015). Voice lineups: a practical guide. *Proceedings of the International Congress of Phonetic Sciences* (pp. 10-14).

ENFSI (2021). Best Practice Manual for the Methodology of Forensic Speaker Comparison. European Network of Forensic Science Institutes.

Jessen, M. (2020). Speaker profiling and forensic voice comparison: The auditory-acoustic approach. In Coulhard, May & Sousa-Silva (Eds.), *The Routledge Handbook of Forensic Linguistics*. Routledge.

Kreiman, J., & Sidtis, D. (2011). Foundations of voice studies: An interdisciplinary approach to voice production and perception. John Wiley & Sons.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press.

San Segundo, E. (2021). International survey on voice quality: Forensic practitioners versus voice therapists. *Estudios de Fonética Experimental*, 30, 9-34.

San Segundo, E., Foulkes, P., French, P., Harrison, P., Hughes, V., & Kavanagh, C. (2019). The use of the Vocal Profile Analysis for speaker characterization: Methodological proposals. *Journal of the International Phonetic Association*, 49(3), 353-380.

San Segundo, E., & Mompeán, J. (2017). A simplified Vocal Profile Analysis Protocol for the assessment of voice quality and speaker similarity. *Journal of Voice*, 31(5), 644.e11–644.e27.

San Segundo, E., & Skarnitzl, R. (2021). A computer-based tool for the assessment of voice quality through visual analogue scales: VAS-Simplified Vocal Profile Analysis. *Journal of Voice*, 35(3), 497.e1-e9.

San Segundo, E., López-Jareño, A., Wang, X., & Yamagishi, J. (2025). Human perception of audio deepfakes: The role of language and speaking style. *arXiv preprint arXiv:2512.09221*.

Boundary marking by creaky voice across sentence types

Anna Kohári & Fruzsina Csébfalvi

(ELTE Research Centre for Linguistics, Hungary / Radboud University, Netherlands)

Introduction. Non-modal voicing can have a wide range of purposes in speech, one of which is signalling intonational phrase boundaries (Garellek, 2019; Davidson, 2021). For instance, Gonzalez (2022) found that non-modal voice quality commonly appears at intonational phrase boundaries in Spanish declaratives, and Markó (2013) reported that Hungarian speakers likewise use creaky voice to mark boundaries. More recent studies suggest that different types of voice qualities can function as acoustic cues signalling the illocution type of different sentences. It has been shown that Icelandic speakers tend to use breathy voice in rhetorical questions and exclamatives, as also observed for German and English speakers (Dehé & Braun, 2020a; Dehé & Wochner, 2024). This raises the question of whether non-modal voicing generally marks phrase boundaries across different sentence types with different illocutionary acts. The boundary-marking function of creaky phonation is primarily examined irrespective of sentence types. Markó (2013) found that similarly to declarative sentences, the last syllable of yes/no questions was typically produced with creaky voice. However, interrogatives and declaratives are generally characterized by a phrase-final low boundary tone in Hungarian, while exclamatives typically end with a mid-boundary tone (Gyuris & Mády, 2014). The aim of this study is to determine whether phrase-final creaky voice systematically occurs in Hungarian, regardless of sentence types. Our research question was whether creaky voice occurs at the end of exclamatives, even when they are produced with a mid-boundary tone and prototypical creaky voice (cf. Garellek, 2019) cannot be realized.

Methods. The study included 10 native Hungarian speakers (5 male, 5 female), aged between 19 and 26 years. Participants produced 10 sentences of each sentence type (declarative, imperative, exclamative) in appropriate linguistic contexts (Alberti et al., 2021). This method of the experiment was inspired by the study of Dehé & Braun (2020b). Each sentence set contained the same number of syllables and matched exactly in the final word (CVCVCVC). The last three vowels of every target sentence (900 vowels) were perceptually and visually classified in Praat as modal, breathy, or creaky, following Zahner et al. (2020). Interrater agreement for phonation-type annotations was 96%, based on a 10% sample reviewed by an independent expert. We also measured the f0 of the vowels automatically. Mixed-effects regression models, Fisher's exact test and post hoc analyses were performed in R.

Results. Declaratives and imperatives exhibited similar patterns regarding the voice quality of the final syllables (Figure 1). While declarative and imperative sentences both tended to have a creaky vowel in the final syllables, this pattern was not observed in exclamatives. Exclamatives were realized with a modal vowel in the final syllables in more than half of the cases, while creaky voice occurred in less than half of the phrase-final vowels. The final vowels in exclamatives showed a statistically significant difference in voice quality relative to both declaratives and imperatives based on the post hoc tests ($p < .001$). Exclamatives typically exhibited higher f0 than the other sentence types, in line with expectations, when accounting for speaker gender and phonation type.

Discussion. According to our results, using creaky voice on the last syllable as a marker of intonational phrase boundaries varies depending on sentence type in Hungarian and is not universally present across all sentence types. In those sentence types characterized by low phrase-final f0, the final vowels are typically glottalized, whereas in exclamatives, which exhibit higher phrase-final f0, most speakers do not mark the boundary with glottalization.

These findings highlight that creaky voice's function as a prosodic boundary marker in Hungarian is modulated by the interplay of sentence type and boundary tone.

REFERENCES

Alberti, G., Dóla, M., Kárpáti, E., Kleiber, J., Viszket, A., & Szeteli, A. (2021). Lehetséges lehetséges világaink. *Jelentés és Nyelvhasználat*, 8(1), 105–145.

Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews: Cognitive Science*, 12(3), e1547.

Dehé, N., & Braun, B. (2020a). The prosody of rhetorical questions in English. *English Language & Linguistics*, 24(4), 607–635.

Dehé, N., & Braun, B. (2020b). The intonation of information-seeking and rhetorical questions in Icelandic. *Journal of Germanic Linguistics*, 32(1), 1–42.

Dehé, N., & Wochner, D. (2024). Voice quality and speaking rate in Icelandic rhetorical questions. *Nordic Journal of Linguistics*, 47(1), 111–120.

Garellek, M. (2019). The phonetics of voice. In W. Katz, & P. Assmann (Eds.), *Handbook of phonetics* (pp. 75–106). Routledge.

González, C., Weissglass, C., & Bates, D. (2022). Creaky voice and prosodic boundaries in Spanish: An acoustic study. *Studies in Hispanic and Lusophone Linguistics*, 15(1), 33–65.

Gyuris, B., & Mády, K. (2014). Approaching the prosody of Hungarian wh-exclamatives. In *VLlxx: Papers in Linguistics* (pp. 333–349).

Markó, A. (2013). *Az irreguláris zönge funkciói a magyar beszédben*. ELTE Eötvös Kiadó.

Zahner, K., Xu, M., Chen, Y., Dehé, N., & Braun, B. (2020). The prosodic marking of rhetorical questions in Standard Chinese. *Proceedings of Speech Prosody* (Vol. 10, pp. 389–393).

On the Role of Givenness in Prosodic Prominence

Vieri Samek-Lodovici

(University College of London, United Kingdom)

This talk will identify some new strengths and weaknesses in Schwarzschild's 1999 (Sch99) analysis of how discourse givenness shapes the distribution of pitch accents (PAs). Sch99 proposes that given a free distribution of F-marks on syntactic heads and phrases, the constraints in (1) determine all observed PA-distributions provided that F-marked nodes undominated by other F-marked nodes count as Foc-marked (i.e. foci) and hence attract a PA via the FOC constraint.

(1)

GIVENness: A constituent that is not F-marked is given.

AVOIDF: Do not F-mark

FOC: A Foc-marked phrase contains an accent.

HEADARG: A head is less prominent than its internal argument.

As for strength, I'll show that Sch99 correctly predicts the minimal PA-pattern alternation in (2) and (3) (discourse context in curly brackets. PAs in capitals. Relevant corresponding F-marking shown on the right). The alternation hinges on the presence/absence of a conditional. The fact that givenness accounts for it, demonstrates its fundamental role in shaping prosody.

(2) {They will hire Mary but} they will FIRE SUE. (F-marking: $[_{VP}\text{fire}_F\ \text{Sue}_F]$)

(3) {What will they do if John hires Mary?} They'll fire SUE. (F-marking: $[_{VP}\text{fire}_F\ \text{Sue}_F]_F$)

For (2), Sch99 explains that (i) *fire* and *Sue* are new, hence F-marked to not violate *GIVENness*; (ii) the VP <$fire_F\ Sue_F$>, however, is not F-marked because its existential F-closure <∃P, ∃Y, P(Y)> (roughly *something is done to someone*) is entailed by <*they will hire Mary*>, and hence given; (iii) the F-marked $fire_F$ and Sue_F thus count as foci for FOC and get a PA each.

By contrast in (3), unexamined in Sch99 though conditionals in another construction are, *fire* and *Sue* are again F-marked because new, but the VP <$fire_F\ Sue_F$> is also F-marked because not entailed and hence not given. Specifically, the conditional blocks <what will they do> from entailing the VP's existential F-closure <∃P, ∃Y, P(Y)> because whether something will be done is dependent on whether John hires Mary. It follows that the VP is F-marked, and, therefore, that the F-marked fire and Sue do not count as focused because dominated by an F-marked VP. Only the VP counts as focused, thus receiving the single PA on *SUE* via the HEADARG constraint.

On the weakness side, I'll show that the analysis incorrectly predicts both PA patterns in (3) and (4) to be acceptable, even though (4) is not.

(3) {John called Mary, and then} MARY called SUE. (F-marking: $[_{TP}\text{Mary}_F\ \text{called}\ \text{Sue}_F]$)

(4) {John called Mary, and then} Mary called SUE. (F-marking: $[_{TP}\text{Mary called Sue}_F]_F$)

For (3), Sch99 shows that (i) *Sue* is not given, and is thus F-marked to not violate *GIVENness*; (ii) the VP is not F-marked because its existential F-closure <∃X, *someone called* X> is entailed by <*John called Mary*>; (iii) Mary is given, but leaving it F-unmarked produces the TP <Mary called Sue$_F$> which violates *GIVENness* because it is not given (nothing entails that Mary called someone). F-marking Mary fixes this because the VP's resulting existential F-closure <∃Y, ∃X, Y called X> is entailed by <*John called Mary*> and hence given, thus letting the TP node remain F-unmarked without violating GIVENness. Since *Mary* and *Sue* are F-marked and undominated by F-marked nodes, they are Foc(used) and get a PA via FOC. Left undiscussed in Sch99, however, is the structure that does F-mark the TP node [*Mary called SUE*$_F$]$_F$ leaving Mary unmarked. Like (3), this structure violates AVOIDF twice and satisfies *GIVENness* (since its existential F-closure, roughly something happened, is entailed by the context). Therefore, it, too, with its single PA on Sue, is optimal, incorrectly predicting the PA on *Mary* to be optional.

The talk will describe the above reasoning in a much clearer, step-by-step, way, and examine what changes are necessary to capture the alternation in (1)-(2) without getting (3)-(4).

REFERENCES

Schwarzschild, R. (1999). GIVENNESS, AvoidF, and Focus. Natural Language Semantics, 7(2), 141-177.

Voice quality and tone as independent dimensions of contrast in Dinka

Bert Remijsen

(The University of Edinburgh, United Kingdom)

This paper presents an acoustic analysis of voice quality in the Bor dialect of Dinka, a Nilotic language spoken in South Sudan. Bor Dinka presents a binary contrast between modal voice and breathy voice, which is orthogonally crossed in the phonology with both a four-way tone contrast and a three-level vowel length contrast (Andersen 1987, Remijsen 2013). The phonemic voice quality contrast means that speakers of Dinka shift between modal and breathy voice qualities from one syllable to the next within an utterance, as a function of the specification for voice quality of the morphemes.

Languages with independently contrastive voice quality and tone are rare, and the evidence on them is limited. Against the background of this evidence base, it is worthwhile to conduct a production study in which voice quality, tone, and vowel quality are orthogonally crossed. We carried out a comprehensive acoustic study on this configuration in Dinka. The study is based on 29 four-member minimal sets for voice quality (Modal vs. Breathy) and tone (Low vs. High), across all seven of the Dinka vowels (/i, e, ɛ, a, ɔ, o, u/). One such four-member set is presented in Table 1. These materials were elicited from eight speakers of Bor Dinka. The results indicate that voice quality, tone, and vowel quality each have their own primary correlate: phonation, F0, and formants, respectively. In addition, each distinction influences other phonetic parameters to a lesser extent. Importantly, the voice quality contrast is realized saliently on vowels in Low- and-High-toned syllables alike, and across the vocalic domain.

Table 1 – A minimal set for voice quality and tone in Dinka.

Voice quality	**Low**	**High**
Modal	*tòook* ‘light:3sg’	*tóook* ‘light:nf’
Breathy	*tọ̀ook* ‘crack:3sg’	*tọ́ook* ‘crack:nf’

Such a phonemic contrast offers a range of opportunities to researchers of phonation. First, it offers an effective testing ground for acoustic measures of phonation (cf. Garellek 2019). When speakers consistently realise a contrast in phonation, we are in a good position to evaluate the relative discriminatory potential of acoustic measures. In this context, we recommend the use of our data to researchers who want to test phonation measures. Second, phonemic voice quality enables us to examine between- and within-speaker variation in phonation, and it raises interesting questions in relation to a speaker’s baseline.

REFERENCES

Andersen, T. (1987). The Phonemic System of Agar Dinka. *J. of African Languages and Linguistics* 9, 1-27.

Garellek, M. (2019). The phonetics of voice. In W. F. Katz & P.F. Assmann (eds.) *The Routledge Handbook of Phonetics*, 75-106.

Remijsen, B. (2013). Tonal alignment is contrastive in falling contours in Dinka. *Language* 89, 297-327.

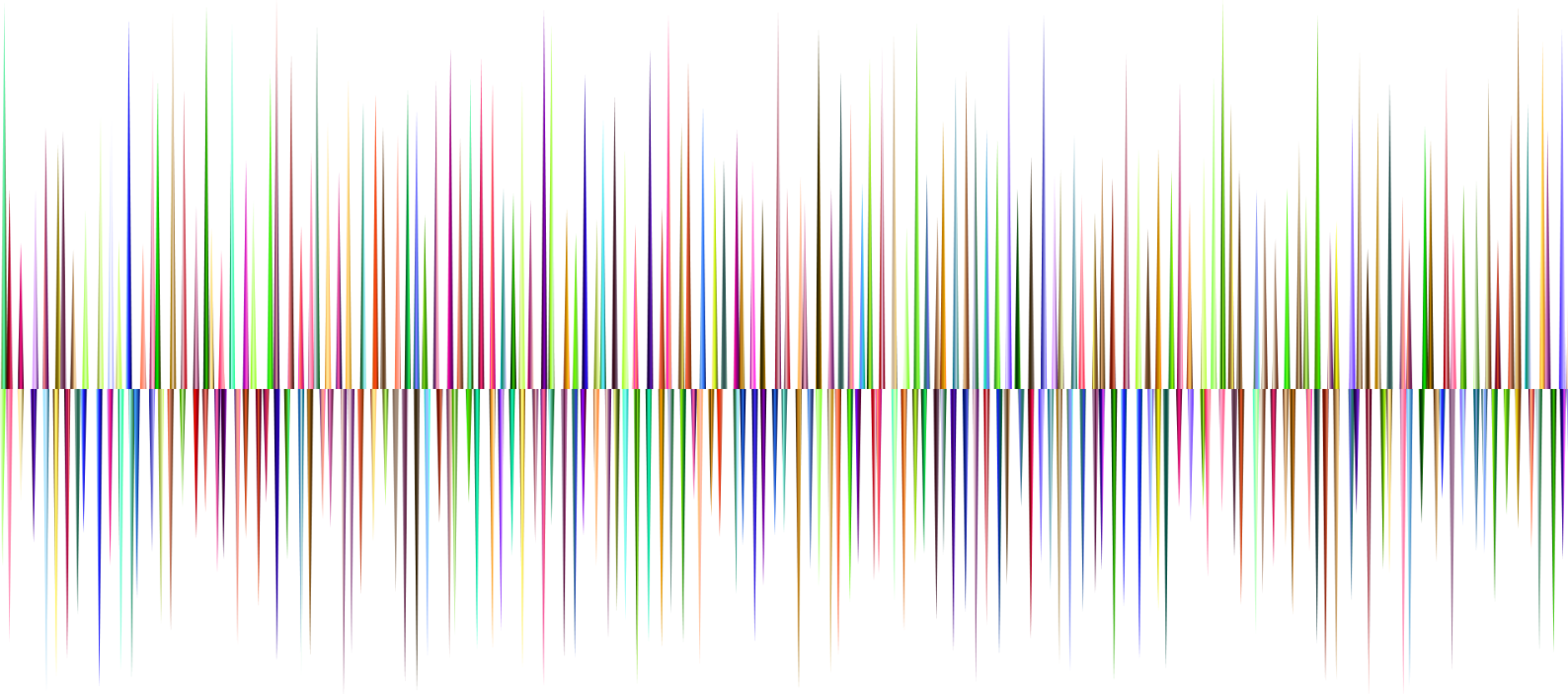
11:10 Coffee break

11:40 Parallel Sessions 2 (Rooms 1, 2, 3)

Room 1: **Stephan Wilhelm** (Université Grenoble Alpes, France) *From layering to source: redefining the laryngeal gesture as the core prosodic matrix*

Room 2: **Jorik Geutjes et al.** (Utrecht University / University Medical Centre Utrecht, Netherlands) *Early Prosodic Boundary Perception: Innate Biases in Preterm Newborns*

Room 3: **Farhat Jabeen & Catherine Lai** (Universität Bielefeld, Germany / University of Edinburgh, United Kingdom) *Pitch, please: Evaluating TTS models for simulating human intonation*



From layering to source: redefining the laryngeal gesture as the core prosodic matrix

Stephan Wilhelm

(Université Grenoble Alpes, France)

Traditional prosodic models have historically operated under a 'superposition' paradigm, where pitch (F0), duration, and intensity are conceived of as discrete functional layers superimposed onto static segments (e.g. Lehiste 1970). In the context of this additive framework, Voice Quality (VQ) is generally relegated to the margins of linguistics, dismissed as an ancillary, static indexical marker of identity or a paralinguistic vehicle for emotion. This paper proposes a theoretical reanalysis of prosodic architecture, arguing for the re-integration of VQ – the acoustic output of the laryngeal gesture – as the core pillar of a source-based matrix.

Drawing on Hjelmslev's (1943) principle of biplanarity, this presentation suggests that VQ can be modelled as a semiotic entity in its own right, where the dynamics of laryngeal activity maps a signifier (the formal phonatory expression) onto a signified that spans pragmatic nuances as well as semantic and morphosyntactic content. In accordance with Ní Chasaide and Gobl (2004, 2010) and Yanushevskaya et al. (2016) on the attitudinal and overtly linguistic potential of the voice source, evidence from speech synthesis (Ward, 2019) and language typology (Remijsen & Ladd, 2008; Remijsen et al., 2025) demonstrates that this semiotic coupling relies on high-resolution control of the laryngeal source. This challenges traditional views of VQ as a mere affective overlay, establishing it instead as an autosegmental prosodic operator.

Crucially, this moves beyond Lehiste's (1970) superpositional framework by questioning whether segmental and prosodic organisation can truly be treated as fully independent layers of speech organisation.

Following Van Heuven's (1994) identification of the segment as the minimal prosodic domain, it is argued that the laryngeal gesture allows for sub-segmental modulation. This goes beyond mere temporal resolution. This implies a dynamic control of the source spectrum that shapes the signal's internal structure during segmental realisation. By demonstrating that VQ transforms the acoustic fabric from the inside out, this talk proposes that it is not a secondary layer, but the primary matricial substance in which segmental entities are embedded. This ontological precedence is mirrored in early language development, where infants are sensitive to prosodic and voice-source cues predating the stabilisation of discrete segmental categories (e.g. Mehler et al., 1988; De Boysson-Bardies, 1999; Polzehl et al., 2024).

This research thus posits that the laryngeal output is speech's constitutive acoustic substrate, not a peripheral adjunct. Within this integrated architecture, the segment is redefined as a localised spectral differentiation of the laryngeal continuum rather than a bundle of inherent features acting as a carrier for superimposed prosodic properties. The source is not an external variable; it is the foundational substance through which meaning is instantiated. The vocal tract modulates a pre-structured laryngeal flux, giving rise to segmental contrasts as focal refinements of the matrix. Yet, prior to this filtering stage, VQ already incorporates meaning into the material texture of the signal by shaping the source spectrum, revealing the traditional segmental-suprasegmental dichotomy as essentially an analytical artefact.

Finally, this research highlights the broader implications of a source-based model, noting that VQ serves as a primary vector for pragmatic effectiveness, as evidenced by its role in

enhancing the interactional clarity of synthetic agents (Lameris & Ward, 2025). By contrast with additive frameworks that marginalise spectral dynamics, the laryngeal output provides the core substance for linguistic communication. Ultimately, the voice source must be embraced as the generative environment through which semiotic content is realised.

REFERENCES

De Boysson-Bardies, B. (1999). *How Language Comes to Children: From Birth to Two Years*. MIT Press.

Gobl, C., & Ní Chasaide, A. (1997). "The voice source in speech production". In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 427-481). Blackwell.

Gobl, C., & Ní Chasaide, A. (2004). «The role of voice quality in communicating emotion, mood and attitude". *Speech Communication*, 40(1-2), 189-212.

Gobl, C., & Ní Chasaide, A. (2010). Voice quality. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 388-441). Wiley-Blackwell.

Hjelmslev, L. (1943/1961). *Prolegomena to a Theory of Language*. University of Wisconsin Press.

Lameris, H., & Ward, N. G. (2025). Creakiness, breathiness, and nasality contribute to the perceived suitability of synthesized speech in a pragmatically-rich domain. *Proceedings of the Speech Synthesis Workshop (SSW)*.

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition. *Psychological Science*, 1(1), 148-153.

Ní Chasaide, A., & Gobl, C. (2004). Voice quality and f0 cues for affect expression: Implications for synthesis. *Proceedings of the 15th International Congress of Phonetic Sciences*.

Polzehl, T., Herzig, T., Wicke, F., Wermke, K., Khamsehashari, R., Dahlem, M., & Möller, S. (2024). Towards Classifying Mother Tongue from Infant Cries - Findings Substantiating Prenatal Learning Theory. *Interspeech 2024*, 4199-4203. doi:10.21437/Interspeech.2024-345.

Remijsen, B., & Ladd, D. R. (2008). The interaction of tone and vowels in Dinka. *Journal of Phonetics*, 36(2), 294-313.

Remijsen, B., Blum, M. L., & Noè, U. (2025). Voice quality and tone as independent dimensions of contrast in Dinka. *Glossa: A Journal of General Linguistics*, 10(1), 1-41.

Van Heuven, V. J. (1994). What is the smallest prosodic domain? In P. A. Keating (Ed.), *Phonological Structure and Phonetic Form* (pp. 76-98). Cambridge University Press.

Ward, N. G. (2019). *Prosodic Patterns in English Conversation*. Cambridge University Press.

Yanushevskaya, I., Ní Chasaide, A., & Gobl, C. (2016). The interaction of long-term voice quality with the realisation of focus. *Proceedings of the 8th International Conference on Speech Prosody (Speech Prosody 2016),* 936-940. doi:10.21437/SpeechProsody.2016-193.

Early Prosodic Boundary Perception: Innate Biases in Preterm Newborns

Jorik Geutjes et al.

(Utrecht University / University Medical Centre Utrecht, Netherlands)

For newborns acquiring language, segmenting continuous speech into meaningful linguistic units is an important step. Before they acquire the vocabulary and grammar of their native language, infants must depend on non-lexical cues to identify word and phrase boundaries. The prosodic structure of speech may assist them in this task. Major speech units, e.g. Intonational Phrases (IPs), are generally marked by three types of prosodic cues: pitch change, pre-boundary syllable lengthening, and pauses [1]. Although infants are known to be sensitive to prosodic structure early on [2,3], the mechanisms underlying this sensitivity remain unclear.

We hypothesised that infants earliest processing of prosodic boundaries is driven by innate perceptual biases, defined as physiologically-motivated or cross-species perceptual mechanisms, in the form of the Respiratory Code (RC) and the Iambic-Trochaic-Law (ITL). The RC links respiratory physiology to prosodic structure, associating high pitch with phrase beginnings and low pitch with phrase endings, with inhalation-related pauses typically occurring between phrases [4]. According to the ITL, listeners universally tend to perceive lengthened and low-pitched elements as phrase-final [5], a pattern that has also been observed in non-human species, suggesting an evolutionary origin [6]. Together, these biologically-motivated principles highlight pitch changes, duration and pauses as potential cues for prosodic boundaries and may underlie infants early sensitivity to such cues, providing innate tendencies to detect transitions between IPs even before substantial linguistic exposure. Therefore, we predicted that preterm newborns are able to process IP boundaries through these biases using the individual cues, despite their minimal linguistic exposure since the onset of hearing.

Methods

To test this prediction, we examined the processing of IP boundaries in preterm newborns, born to Dutch-speaking parents between 28 and 34 weeks of gestation. Within the first week after birth, 40 preterm newborns listened to short coordinated name sequences in Dutch, containing or lacking an utterance-medial IP boundary ([Moni and Lilli and Manu] vs. [Moni and Lilli] [and Manu]). The stimuli were presented to each infant in six conditions: a no-boundary baseline; an all-cues condition combining phrase-final pitch rise, final lengthening, and a pause; and four single-cue conditions in which the boundary was marked by a pitch fall, a pitch rise, final lengthening, or a pause alone (see Figure 1). Using EEG, we measured the neurophysiological response to boundary processing, the Closure Positive Shift (CPS), in each condition versus the no-boundary baseline.

Results

Linear mixed effects modelling showed a significant positive difference in ERP amplitude in the right-frontotemporal region for the **pause-only condition**, relative to the no-boundary condition (\beta = 1.85, SE = 0.52, t = 3.58, p = 0.002). This amplitude difference is manifested as a CPS, emerging approximately 600ms after preboundary syllable onset (see Figure 2). This finding suggests preterm newborns initially process major prosodic boundaries based on solely on pauses, partially supporting our hypothesis. However, no significant effects were observed in any region for the remaining boundary conditions, including the all-cues condition, which also featured a pause. The absence of an effect in the all-cues condition may

be the result of conflicting information provided by the phrase-final pitch rise. This high final pitch contrasts with the low phrase-final pitch described by both the RC and the ITL, potentially overriding or weakening the boundary cue provided by the pause, and consequently preventing boundary perception.

Discussion

Overall, our results suggest that preterm newborns rely on pauses as reliable indicators of major prosodic boundaries, reflecting an innate bias. This experience-independent reliance likely extends to fetuses at a comparable gestational age, providing a prenatal mechanism supporting early speech segmentation. Future research could examine whether a larger degrees of final lengthening and pitch fall, potentially exceeding those typically used in Dutch adult-directed speech, may be required to elicit neural responses reflecting boundary perception in preterm newborns, to further clarify the role of the innate biases related to these cues.

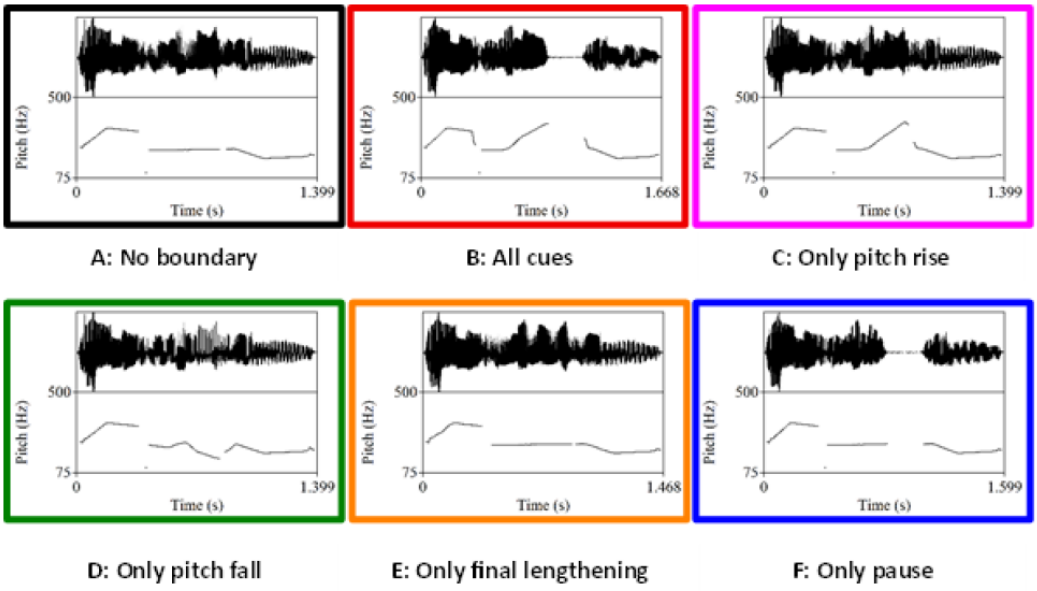


Figure 1. Example waveforms and pitch contours for the coordinated name sequence across stimulus conditions

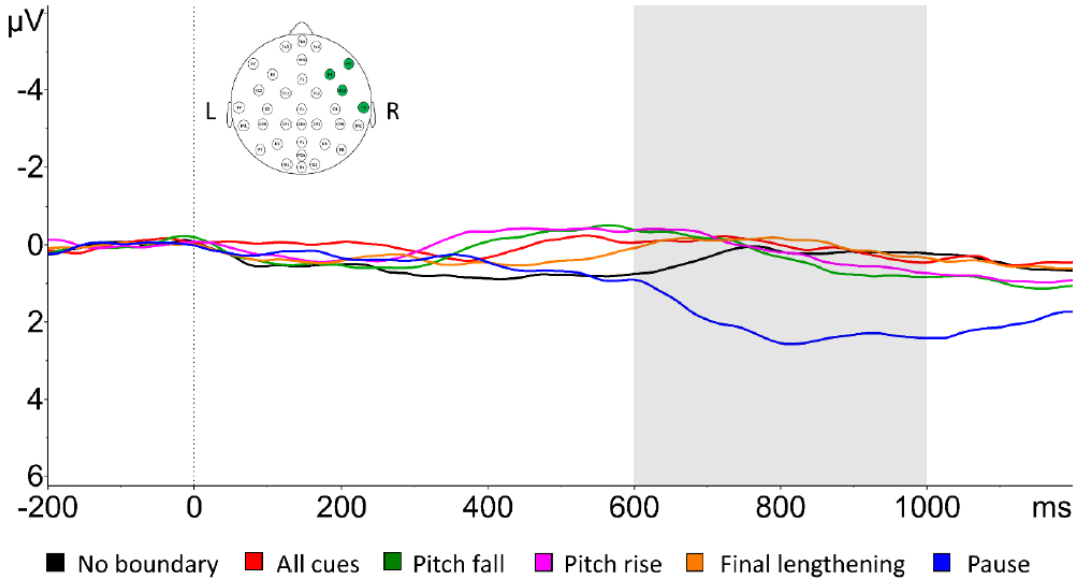


Figure 2. ERP waveforms for frontotemporal electrodes on the right hemisphere, timelocked to the onset of the preboundary syllable. Typically, CPS is a positive (downward-facing) deflection observed on frontal electrodes between 500-800ms after detecting a prosodic boundary. This response may be delayed in preterm newborns due to incomplete myelination of the brain (e.g. 600-1000ms, indicated by the grey area).

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] Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*, 95(3), 1570-1580.

[3] Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, 2(3), 385-394.

[4] Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science*, 8(2), 425-434.

[5] Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago press.

[6] Spierings, M., Hubert, J., & Ten Cate, C. (2017). Selective auditory grouping by zebra finches: testing the iambic–trochaic law. *Animal cognition*, 20(4), 665-675.

Pitch, please: Evaluating TTS models for simulating human intonation

Farhat Jabeen & Catherine Lai

(Universität Bielefeld, Germany / University of Edinburgh, United Kingdom)

This study evaluates the ability of TTS systems to simulate probabilistic intonation observed in human data. We compare the output of two non-deterministic text-to-speech (TTS) systems with human intonation in Hindi, an under-studied language. While prior work demonstrates that TTS output diverges from human speech (Hu et al., 2024), we argue that different TTS models vary in their ability to approximate human-like intonation. To this end, we compare a commercial TTS system (ElevenLabs) and an open-source multilingual system (Toucan), assessing their performance against intonational patterns derived from human data.

As a test case, we examine the Hindi discourse particle '*-hii*'. The particle exhibits variable prosodic attachment to preceding words (hosts), reflected in the alignment of the upstepped F0 peak (Jabeen and Patil, 2025). The i-ii indices in (1) illustrate this variation. Crucially, '*-hii*' cannot associate with words on its right edge (1-*iii) (Sharma, 1999; Jabeen and Patil, 2025). We therefore hypothesise that a model approximating human speech should capture (i) variability in the attachment of '*-hii*' on the left edge, denoted by the alignment of upstepped F0 peak and (ii) a falling or flat F0 contour on its right edge. We further test whether systematic differences emerge in the output of the TTS systems, indicating their ability to model human speech.

Methodology Human baseline was obtained from a female native speaker of Hindi, who recorded 50 sentences extracted from Bollywood film scripts. To generate TTS output, the same test sentences were provided in the Devanagari script to ElevenLabs and Toucan systems. Each system was configured to produce a male Standard Hindi voice at normal tempo and stability settings. Time-normalised F0 contours (five samples per word) were extracted from five words on either side of '*-hii*' (where available). These contours were then normalised relative to each speaker's mean F0. Variability in F0 was analysed using Functional Principal Component Analysis (FPCA) followed by Linear Mixed-Effects Regression on the first three principal components (PCs).

Results The FPCA model accounted for 74% of the overall variance in the data (PC1: 30%, PC2: 26%, PC3: 18%). Figure 1 illustrates variability in the intonation of '*-hii*', which optionally carries an upstepped F0 peak. Regression analysis revealed no difference between human and TTS output for PC1 ($p=0.3$) or PC3 ($p=0.4$) scores. Indicating partial divergence in intonational realisation, PC2 showed a significant difference between the human baseline and TTS systems:

- **Human-ElevenLabs** (B: -0.24, SE = 0.09, t = -2.4, p = 0.01)
- **Human-Toucan** (B: -0.21, SE = 0.09, t = -2.1, p = 0.03)

No difference was found in the output of TTS systems. Significant differences in PC2 highlight that specific aspects of intonation are not equally well modelled. To detect outliers in the output, the data was subjected to qualitative inspection. F0 contours (Figure 1, bottom) demonstrate that an identical sentence rendered by ElevenLabs and Toucan can exhibit substantial variation in both F0 range and contour.

Table 1: Variability in human baseline and TTS models' output.

Feature	Human	ElevenLabs	Toucan
Lack of upstepped peak on host of '-hii'	10%	12%	16%
Lack of flat/falling F0 following '-hii'	12%	4%	16%

Conclusion While all systems deviate from previous claims, ElevenLabs and human baseline align more closely on the left edge of '*-hii*', whereas Toucan better approximates human baseline on the right edge. Overall, these results suggest that TTS systems are valuable tools for generating speech data, though deviations from human speech should be carefully considered.

(1) tʊm.hɑ.re^{ii} b^{h}ɑi=ko^{i} -hi k^{h}et^{*iii} dʒot.nɑ pə.ɽe gɑ
your brother-Acc -hii field plow lie would
'Your brother would have to plow the fields.'

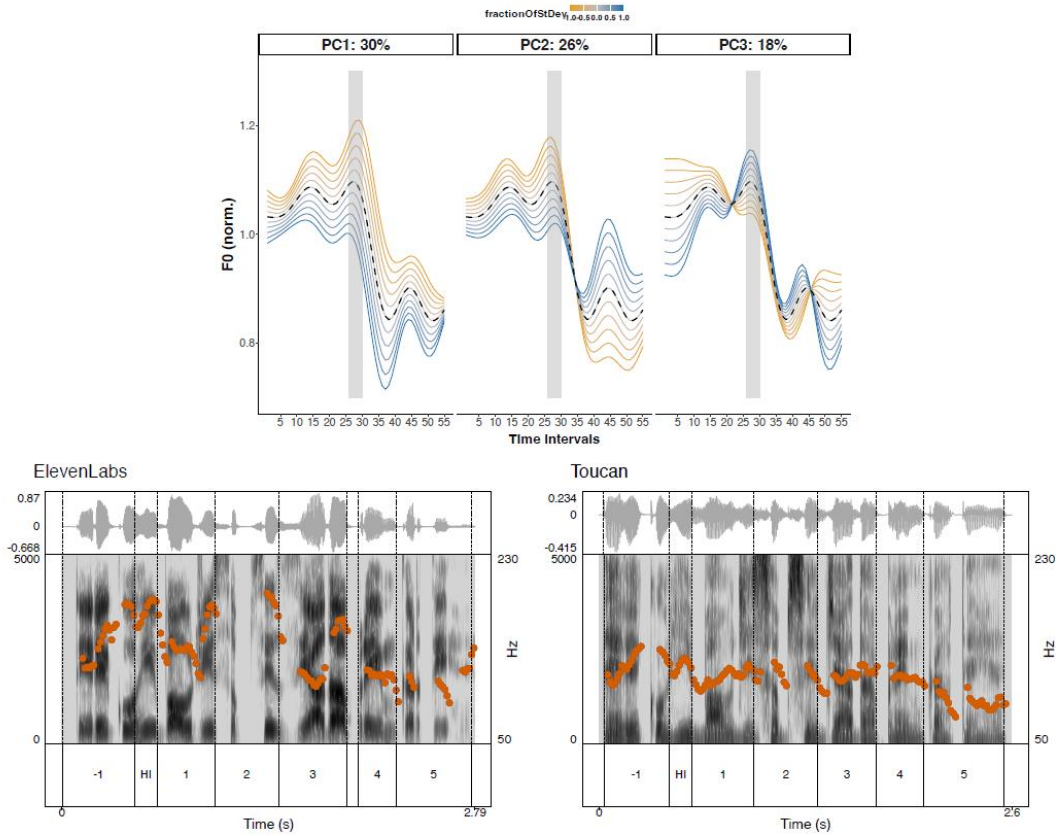


Figure 1: Top: FPCA curves of output generated by human speaker, ElevenLabs, and Toucan models. The gray rectangle denotes the position of '-hii'.
Bottom: F0 contour of an utterance produced by ElevenLabs and Toucan systems.

	Human	ElevenLabs	Toucan
Lack of upstepped peak on host of '-hii'	10%	12%	16%
Lack of flat/falling F0 following '-hii'	12%	4%	16%

Table 1: Variability in human baseline and TTS models' output.

REFERENCES

Hu, N., Kim, J., Orrico, R., Gryllia, S. & Arvaniti, Amalia (2024). Can OpenAI's TTS model convey information status using intonation like humans? *Proceedings of SpeechProsody*, 32–36.

Jabeen, F. & Patil, S. (2025). Intonation and prosodic phrasing of particle '-hii' in Hindi/Urdu dialogues. *Proceedings of 29th Workshop on the Semantics and Pragmatics of Dialogue*.

Sharma, D. (1999). Nominal clitics and constructive morphology in Hindi. *Proceedings of LFG99*, 1–21. CSLI Publications.

12:35 Lunch break

14:00 Plenary 2

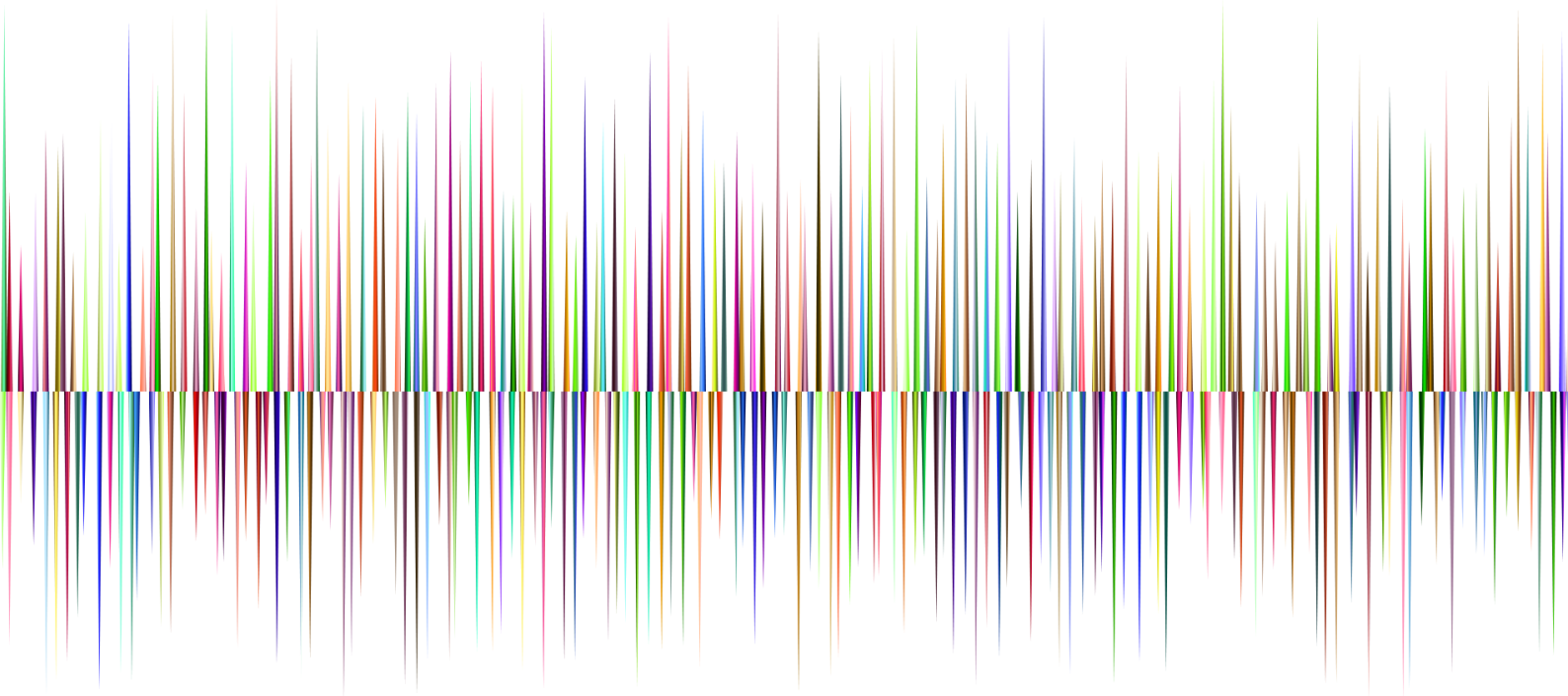
Radek Skarnitzl (Charles University, Prague) *Phrasal prosody as a cornerstone for L2 English*

15:10 Parallel Sessions 3 (Rooms 1, 2, 3)

Room 1: Sophie Herment, Julia Bongiorno & Laetitia Leonarduzzi (Aix-Marseille Université, France) *Rising contours in Ireland: perception and interpretation*

Room 2: Mark Campana (Kobe City University of Foreign Studies, Japan) *Prosody at the gate of cognition*

Room 3: Nadav Matalon & Eyal Weinreb (Weizmann Institute of Science, Israel) *Evidence for a language-like organization of prosody*



Keynote speaker

Phrasal prosody as a cornerstone for teaching the pronunciation of L2 English

Radek Skarnitzl

(Charles University, Prague, Czech Republic)

Prosodic aspects of speech like stress, intonation, phrasing and rhythm are typically regarded as difficult to teach. That is also one of the reasons why they are rarely targeted in foreign language classrooms, despite their proven importance for spoken communication. First, they serve several essential functions in everyday communication: we use them to indicate the grammatical structure of utterances, to highlight important information, to express our stance to our interlocutors and to what we are saying. In this respect, prosodic aspects of speech have been described as "signposts" which help navigate listeners around the speaker's meaning. Second, it is largely these characteristics of speech that affect the speaker's comprehensibility – not necessarily only in a foreign language. In other words, unpredictable prosodic patterns are likely to make speakers more difficult to understand, requiring greater cognitive effort on the part of the listener. This also means that it is based on prosodic aspects of speech that listeners form (usually implicit) judgements about speakers, and inadequate "prosodification" of utterances may thus negatively impact the way speakers are perceived. Crucially, however, there is ample evidence that prosodic aspects are both teachable and learnable.

In the talk, I will approach the teaching of prosody from the perspective of the prosodic phrase. I will introduce the prosodic phrase as the central unit of speech in terms of both production and perception, and particularly in relation to our cognition. As such, the phrase is a unit which plays an essential role in fluent spoken communication and whose "signature" may be observed throughout the sound patterns of English. Based on that, I propose a phrase-centric approach to teaching the pronunciation of L2 English, which consists in **four pillars**; these address four fundaments of English phrasal prosody.

I will also demonstrate the effectiveness of a short pronunciation training which was centred around the prosodic phrase. Learners listened to their own speech whose melodic and rhythmic aspects were computer-manipulated to approach native-like prosodic patterns. This manipulated speech served as a model for repetition. This short, individual training resulted in significant improvements: in a delayed post-test six weeks after the training, the learners manifested improved melodic patterning, better phrasing, and higher scores of perceived competence as compared to pre-training recordings.

The talk will conclude with some pedagogical suggestions. I am convinced that the basics of prosody in the form of the four pillars are straightforward to teach; the vital requirement is that phrasal prosody work is integrated into language classrooms. I will therefore offer specific examples on how to make learners aware of the specifics of English phrasal prosody and on how to target phrasing while focusing on grammatical structures or on vocabulary, while practicing reading or listening, and on how to pay attention to it while speaking.

REFERENCES

Chun, D. M., & Levis, J. M. (2020). Prosody in L2 teaching: Methodologies and effectiveness. In: C. Gussenhoven & A. Chen (Eds.), *Oxford handbook of language prosody* (pp. 619–630). Oxford University Press.

O'Brien, M. G. (2022). Making the teaching of suprasegmentals accessible. In: J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 85–106). Wiley Blackwell. https://doi.org/10.1002/9781394259663.ch5

Sanderman, A. A., & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and Speech*, 40(4), 391–409. https://doi.org/10.1177/002383099704000405

Skarnitzl, R., & Bořil, T. (2024). Training of English prosody with acoustically modified voices. *Journal of Second Language Pronunciation*, 10(3), 375–403. https://doi.org/10.1075/jslp.24041.ska

Rising contours in Ireland: perception and interpretation

Sophie Herment, Julia Bongiorno & Laetitia Leonarduzzi

(Aix-Marseille Université, France)

Irish English is known for its geographical variation, and in particular its North / South division. We focus in this paper on the main varieties spoken in the northern and southern parts of Ireland, namely Mid-Ulster English in the North (Gregg, 1972) and Hiberno-English in the South of the island (Harris, 1985). The two varieties have been widely documented (see Wells, 1982 or Trudgill et al., 2005 amongst others). One of the most striking characteritics of Mid-Ulster English is its rising intonation used not only for questions (including Wh-questions) but also for declarative sentences and commands, being therefore part of the intonational system as the default, unmarked contour (Turcsan & Herment, 2015). These rises are referred to as UNB rises (Cruttenden, 1997), and can be distinguished from what is known as HRT (high rising terminals) or 'uptalk' (Warren, 2016), *i.e.* non-systematic rising intonations typical of certain varieties such as American, Australian and New Zealand English, and which also extend to Great Britain (cf. Bradford, 1997; Cruttenden 1994, 1997; Ladd, 2008) and Ireland (Warren, 2016; Bongiorno, 2021).

Our aim is to better understand the correlation between segmental and suprasegmental features: if a speaker is characterized as a northern Irish speaker, will their rising intonations be perceived as UNB rises, *i.e* indicating completeness? In the same way, if a speaker is characterized as a southern Irish speaker, will their rising intonations be perceived as uptalk? In order to answer these questions, we set up a perception test. We selected 3 female speakers (2 young women and an older one) from the PAC-Donegal corpus (Turcsan & Herment, 2015) and from the PAC-Dublin corpus (Bongiorno, 2021). Each speaker pronounces falling declarative sentences and rising declarative sentences in context. The Donegal speakers typically pronounce UNB rises, while the Dublin speakers realize HRTs. 50 Irish participants recruited on the Prolific platform are asked first to complete a questionnaire with personal questions such as place of residence, gender, age, etc. and then to listen to the randomized extracts and tick boxes corresponding to what they interpret:

- i) the speaker sounds confident;
- ii) the sentence is complete;
- iii) the sentence denotes a particular attitude from the speaker;
- iv) the speaker is ready to let the other person speak;
- v) the speaker is checking if the listener is following.

These statements are taken from two former experiments performed on Dublin HRTs (Bongiorno, 2021) and Newcastle English (Herment et al., 2020). Finally, the participants listen to a longer passage by each of the six speakers and are asked to guess where they live and how old they are.

The test is ongoing and the results will be presented at the conference. Our hypothesis is that the recognition of accents will guide listeners in interpreting rising contours. The test will also enable us to examine whether age, place of residence or other social characteristics from the listeners' side have an influence on the interpretation of the rises, and if the presumed age and place of residence on the speakers' side also play a role on the results.

REFERENCES

Bongiorno, J. (2021). *Etude du système intonatif de l'anglais parlé à Dublin : Focus sur les montées stylistiques* [Unpublished PhD thesis]. Aix-Marseille Université.

Bradford, B. (1997). Upspeak in British English. *English Today*, 51(13.3), 29-36.

Cruttenden, A. (1994). Rises in English. In J. W. Lewis (Éd.), *Studies in General and English Phonetics : Essays in honour of Professor J.D. O'Connor* (p. 155-173). London: Routledge.

Cruttenden, A. (1997). *Intonation* (2nd éd.). Cambridge University Press.

Gregg, R.J. (1972). The Scotch-Irish Dialect Boundaries in Ulster, in M. Wakelin (ed), *Patterns in the Folk Speech of the British Isles*: 109-139, London; Athlone.

Harris, J. (1985). Phonological variation and change. *Studies in Hiberno-English*, Cambridge: Cambridge University Press.

Herment, S. Leonarduzzi, L. & Bouzon, C. (2020). The various rising tones in Newcastle English: a phonological distinction? *Anglophonia - French Journal of English Linguistics* 29/2020, Presses Universitaires du Midi. http://journals.openedition.org/anglophonia/3297

Ladd, D. R. (2008). *Intonational Phonology* (2nd edition). Cambridge University Press.

Trudgill, P., Hughes, A. & D. Watt, 2005, *English Accents and Dialects*, Hodder Arnold.

Turcsan, G., & Herment, S. (2015). L'anglais du Nord de l'Irlande. In I. Brulard, J. Durand, & P. Carr (Éds.), *La prononciation de l'anglais contemporain* (p. 183-198). Presses Universitaires du Midi.

Warren, P. (2016). *Uptalk*. Cambridge University Press.

Wells, J. C. (1982). *Accents of English*. Cambridge University Press.

Prosody at the gate of cognition

Mark Campana

(Kobe City University of Foreign Studies, Japan)

Prosody as we know it can be divided into four parts: pitch/key; tunes (melody); timed patterns of rhythm, meter and tempo, and "voice qualities": vocal sounds driven by changing of articulator settings (Laver 1980). Moreover, any comprehensive theory of these phenomena must also take into account their interactions with the brain and mind, especially those that relate to emotion. Here we outline a path by which prosodic features link to specific cognitive functions, which together communicate emotional states and expressions.

Our inspiration comes from Wierzbicka (1999), who proposed that cross-linguistically emotion words are best defined as 'cognitive scenarios' made up of simple propositions with a universal vocabulary. Among a class of mental predicates are wanting (or not wanting), feeling (typically good or bad), knowing (or not knowing), and thinking. Wierzbicka's definition of the emotion word apprehension goes roughly as follows: Somebody thinks that something bad can happen. They don't want it to happen, and don't even know if it will. Still, they want to do something about it if possible. When they think like this, they feel something (slightly) bad."

We propose that mental predicates that contribute to the meaning of emotion words have correlates in the physical world as well, both in the form of brain activity (cognitive functions) and in sound: prosodic features. Voice qualities, for example, play a larger-than-expected role in wanting something, whether it be internally (drawing attention to oneself) or externally, as in the discourse or environment. Similar correspondences can be heard in pitch with feeling, melody with knowing, and timing patterns with thinking. On the cognitive side, wanting corresponds to the function of volition; feeling to that of empathizing; knowing to memory access, and thinking to the planned deployment of utterances.

The data are drawn from stance-final utterances, where relinquishing the floor provides speakers with an opportunity to demonstrate their vocal (as well as verbal) skills more succinctly (Jaffe 2009). These will be enacted for the audience using reiterant speech (Nooteboom 2000) to illustrate the path between each element of prosody and mental predicate/cognitive function.

REFERENCES

Jaffe, Alexandria (2009). *Stance: Sociolinguistic Perspectives*. Oxford University Press.

Laver, John (1980). *The Phonetic Description of Voice Quality*. Cambridge University Press.

Nooteboom Nooteboom, Sieb. (2000). 'The prosody of speech: Melody and Rhythm'. MS, Research Institute for Language and Speech, Utrecht.

Wierzbicka, Anna (1999). *Emotion Across Languages and Cultures*. Cambridge University Press.

Evidence for a language-like organization of prosody

Nadav Matalon & Eyal Weinreb

(Weizmann Institute of Science)

Prosody is often treated as dependent upon text—a layer that is "added" after the words are selected and concatenated. However, from an evolutionary perspective, prosody predates segments and words, both ontogenetically—during language acquisition (Davis et al., 2000)—and likely phylogenetically, in the development of human language. This discrepancy may stand in the way of understanding the structure and function of prosody. This is especially true in spontaneous speech, where text is produced on the fly and is significantly less structured than in written language. In this study (Matalon et al., 2025), we approach prosody as a linguistic system in its own right, drawing on a large-scale analysis of spontaneous conversations in English.

We segment speech into Intonation Units (IUs) (Biron et al., 2021; Chafe, 1994) and characterize the resulting IUs solely on the basis of pitch and intensity. We use an autoencoder neural network to reduce dimensionality, and cluster the standardized IU vectors without pre-specifying the number of clusters. Across datasets, this analysis identifies approximately 200 recurring intonation patterns, distinguished by duration, shape, and register. This set of patterns resembles a lexical vocabulary both in its frequency distribution (Zipf, 1949) and information content (Shannon, 1948). Manual examination of the conversational data shows that an intonation pattern typically conveys a single attitudinal meaning (e.g., enthusiastic, calm, opinionated) and contributes to one of several interactional functions (e.g., yes-no question, contrastive statement, affirmation), determined at the intersection of prosody, text and context. An investigation of IU sequences identifies pairs of patterns that occur in conversation significantly more frequently than expected by chance, but no longer sequences are found. Manual analysis further shows that IU pairs often serve distinct compound functions.

In summary, our data-driven approach identifies signs of "vocabulary", "semantics", and "syntax" in the use of intonation in spontaneous conversation, and shows that IUs constitute a central time scale for coordinating linguistic content across lexical, syntactic, and interactional domains. We further show that, in the process of meaning-making, prosody establishes a frame of interpretation for words, and vice versa, pointing to a non-hierarchical relation between verbal and nonverbal information. These findings can inform and refine existing theoretical concepts (Barnes & Shattuck-Hufnagel, 2022), outlining a language-like organization of prosody that is not reducible to textual structure. The methodology we propose could guide future research aimed at refining these findings and expanding them to additional contexts, such as other languages, genres, populations, and individual speakers.

REFERENCES

Barnes, J., & Shattuck-Hufnagel, S. (Eds.). (2022). *Prosodic theory and practice*. The MIT Press.

Biron, T., Baum, D., Freche, D., Matalon, N., Ehrmann, N., Weinreb, E., Biron, D., & Moses, E. (2021). Automatic detection of prosodic boundaries in spontaneous speech. *PLOS ONE*, 16(5), e0250969. https://doi.org/10.1371/journal.pone.0250969

Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J. K. (2000). Prosodic Correlates of Stress in Babbling: An Acoustical Study. *Child Development*, 71(5), 1258–1270. https://doi.org/10.1111/1467-8624.00227

Matalon, N., Weinreb, E., Freche, D., Volk, E., Biron, T., Moses, E., & Biron, D. (2025). Structure in conversation: Evidence for the vocabulary, semantics, and syntax of prosody. *Proceedings of the National Academy of Sciences*, 122(17). https://doi.org/10.1073/pnas.2403262122

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley Press.

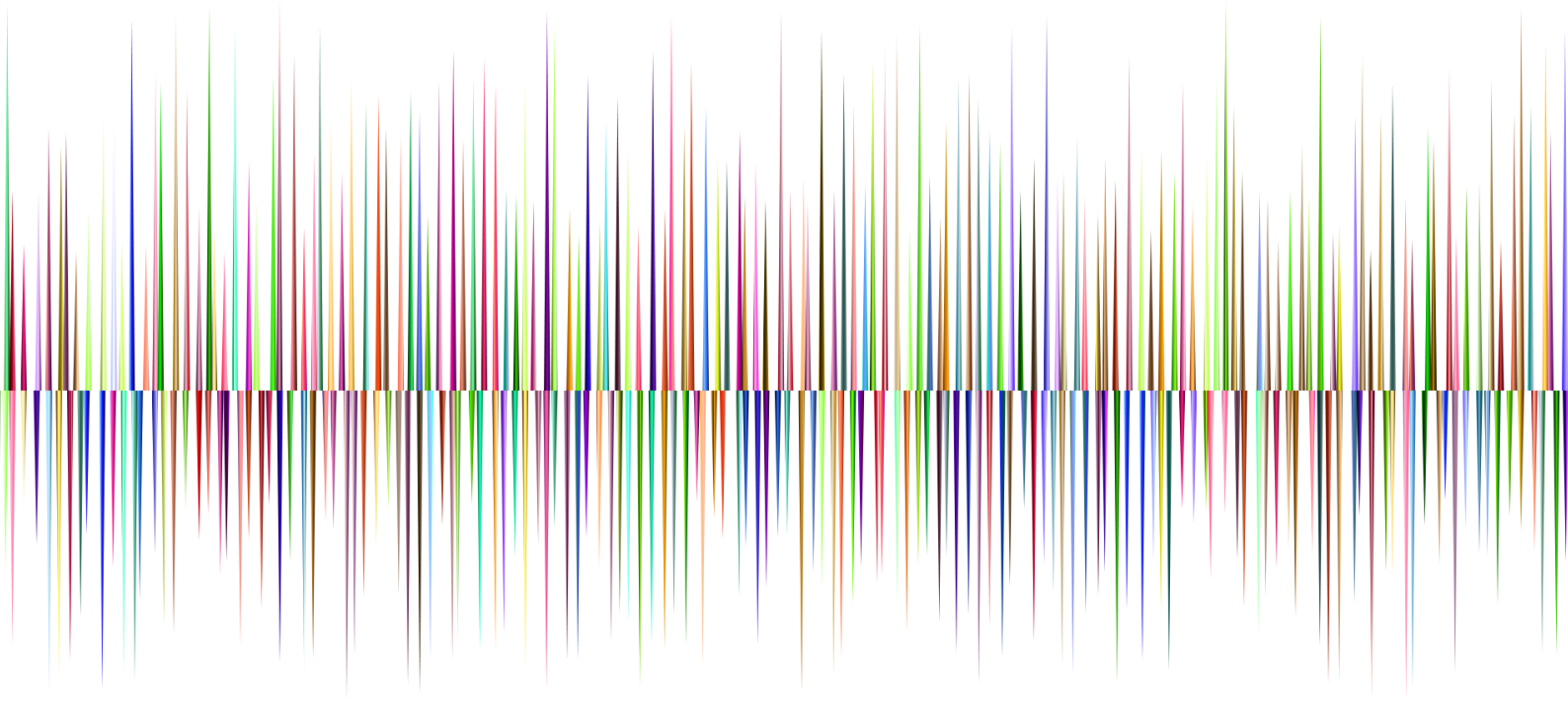
15:55 Coffee break

16:30 Parallel Sessions 4 (Rooms 1, 2, 3)

<u>Room 1:</u> Esther De Leeuw, Scott Lewis & Joséphine Dishpalli (University of Lausanne, Switzerland) *Prosodic analysis of language discrimination in bilingual speech of infants and toddlers*

<u>Room 2:</u> Dan Frost (Université Grenoble Alpes, France) Pardon my French? Identifying the triggers of English-Medium Instruction comprehension hurdles

<u>Room 3:</u> Nicole Dehé & Marieke Einfeldt (Universität Konstanz, Germany) *Intonation and voice quality in Icelandic wh-exclamatives*



Prosodic discrimination of languages in bilingual infants and toddlers 7-36 months of age

Esther de Leeuw, Scott Lewis, Joséphine Dishpalli

(University of Lausanne, Switzerland)

When and how do bilingual babies discriminate their languages? The Dual Language System Hypothesis (DLSH) assumes that bilingual children, exposed to two languages from birth, establish two separate linguistic systems from the onset of acquisition (Genesee, 1989). Alternatively, according to unitary language system hypothesis (ULSH), in the first stage of their language development, words and grammatical rules from both languages are said to be mixed, and only later can the child can be said to have two separate linguistic systems (Volterra & Taeschner, 1978). However, very few studies (Andruski et al., 2014; Maneva & Genesee, 2002; Poulin-Dubois & Goodz, 2001; Sundara et al., 2020; Vihman, 2016; Vihman, 2002) have attempted to answer this question with regard to speech production and most have done so looking at case studies. Ours is the first to include an analysis of standardised pitch and rhythm metrics, in addition to ultrasound and segmental acoustic analyses which are not reported here.

All parents of the bilingual children practiced the one-parent-one-language family language policy and data were elicited separately by each parent (total experiment = 60-90 minutes). Our preliminary analysis of the speech production data focuses on 10 children 7-37 months of age who are growing up in Lausanne, Switzerland, with different bilingual language combinations. Our prosodic measurements for the analysis of language discrimination in prosodic speech production of these infants and toddlers covers the following metrics: (1) percentage of vocalic intervals (%V); (2) standard deviation of consonantal intervals (del-C); (3) difference between 90th and 10th percentile range in semitones (ST) (80% Range); and (4) mean F0 in Hz. Additionally, we examined additional non-parental language exposure effects.

Tentative results indicate high degrees of individual variation across children as well as more language discrimination in pitch metrics than in rhythm metrics, suggesting that language discrimination in bilingual infant and toddler speech production is both child-dependent as well as variabledependent. In sum, these findings substantiate neither unequivocally the DLSH nor unequivocally the ULSH.

REFERENCES

Andruski, J. E., Casielles, E., & Nathan, G. (2014). Is bilingual babbling language-specific? Some evidence from a case study of Spanish–English dual acquisition*. *Bilingualism: Language and Cognition*, 17(3), 660–672. https://doi.org/10.1017/S1366728913000655

Genesee, F. (1989). Early bilingual development: One language or two?*. *Journal of Child Language*, 16(1), 161–179. https://doi.org/10.1017/S0305000900013490

Maneva, B., & Genesee, F. (2002). Bilingual Babbling: Evidence for Language Differentiation in Dual Language Acquisition. In B. Skarbela, S. Fish, & H.-J. Do (Eds.), *Boston University Conference on Language Development 26 Proceedings* (pp. 383–392). Cascadilla Press.

Poulin-Dubois, D., & Goodz, N. (2001). Language differentiation in bilingual infants: Evidence from babbling. In J. Cenoz & F. Genesee (Eds.), *Trends in Bilingual Acquisition* (pp. 95–106). John Benjamins Publishing.

Sundara, M., Ward, N., Conboy, B., & Kuhl, P. K. (2020). Exposure to a second language in infancy alters speech production. *Bilingualism: Language and Cognition*, 23(5), 978–991. https://doi.org/10.1017/S1366728919000853

Vihman, M. (2016). Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and Cognition*, 19(1), 69–88. https://doi.org/10.1017/S1366728914000790

Vihman, M. M. (2002). Getting started without a system: From phonetics to phonology in bilingual development. *International Journal of Bilingualism*, 6(3), 239–254. https://doi.org/10.1177/13670069020060030201

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children*. *Journal of Child Language*, 5(2), 311–326. https://doi.org/10.1017/S0305000900007492

Pardon my French? Identifying the triggers of English-Medium Instruction comprehension hurdles

Dan Frost

(Université Grenoble Alpes, France)

Academics in France are often expected not just to publish in English, but also to present their work in English and in many contexts, to teach their courses in English (English-Medium Instruction / EMI). While some universities provide targeted training for researchers in oral English, most do not (Jiménez Muñoz & González-Álvarez, 2020), yet we know that foreign-accented speech impacts intelligibility (Kang et al., 2018), and increases listeners' cognitive load. Beyond just making the listener work harder, these linguistic barriers can dampen student motivation (Roussel et al., 2017) and hinder overall learning outcomes (Roussel et al., 2022). Furthermore, research indicates that accented English frequently triggers listener bias, resulting in diminished perceptions of the speaker's credibility (Stocker, 2017).

While it was a long-held belief that age was a key factor in successful language learning, more recent studies with adult learners have shown that age-bound neurobiological factors are not as important as prior language learning experience and language exposure (Baker et al., 2008). Those of us involved in teaching undergraduates or working in lifelong learning, and focussing on pronunciation with adult learners can indeed bring positive results, even with short periods of instruction (Frost, 2021).

Over the last two decades, much of the focus in teaching and researching pronunciation has moved away from teaching towards native norms in favour of intelligibility (Levis, 2020), with increasing consideration given to questions of speaker identity (McCrocklin & Link, 2016). Central to pedagogical development is the question of how various prosodic and phonological features interact to hinder or facilitate communication. Following the experimental framework of Nagle et al. (2019), we conducted a study which aims to pinpoint the specific pronunciation features within French-accented academic discourse that most significantly impact both objective intelligibility and comprehensibility (or the listener's perceived ease of understanding).

In this study, naïve participants from a variety of linguistic backgrounds (N=18) provided real-time ratings of perceived ease of understanding while listening to foreign-accented speech. Prosodic features of certain stretches of the two excerpts of lectures given in English by French speakers were modified to try to ascertain whether these features had an effect on the participants. Using specialized dynamic software, listeners continuously adjusted a variable to reflect their processing effort. These sessions were screen-captured to serve as the basis for subsequent stimulated recall interviews, during which participants identified and explained specific linguistic features that impeded their comprehension.

By triangulating quantitative perception data with qualitative interview insights, this study reveals the highly individualized nature of intelligibility and identifies the specific pronunciation features, and their complex interactions with non-phonetic factors, that impede comprehension. These findings offer critical implications for EMI teacher training and pronunciation instruction more generally, and suggest that future research must account for the listener's subjectivity regarding French-accented English to fully understand the mechanisms of communication breakdown.

Key words: English-medium instruction (EMI); intelligibility; lifelong learning; perception tests; prosody; stimulated recall interviews.

REFERENCES

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child—Adult Differences in Second-Language Phonological Learning: The Role of Cross-Language Similarity. *Language and Speech*, 51(4), 317–342. https://doi.org/10.1177/0023830908099068

Frost, D. (2021). Prosodie, intelligibilité et compréhensibilité : L'évaluation de la prononciation lors d'un stage court. *Les Langues Modernes*, 3(2020), 76–90.

Jiménez Muñoz, A. (2020). Shortcomings in the Professional Training of EMI Lecturers: Skills-Based Frameworks as a Way Forward. In D. González-Álvarez & E. Rama-Martínez (Eds.), *Languages and the Internationalisation of Higher Education* (pp. 120–138). Cambridge Scholars Publishing.

Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical Approaches to Measuring the Intelligibility of Different Varieties of English in Predicting Listener Comprehension: Measuring Intelligibility in Varieties of English. *Language Learning*, 68(1), 115–146. https://doi.org/10.1111/lang.12270

Levis, J. (2020). Revisiting the Intelligibility and Nativeness Principles. *Journal of Second Language Pronunciation*, 6(3), 310–328. https://doi.org/10.1075/jslp.20050.lev

McCrocklin, S., & Link, S. (2016). Accent, Identity, and a Fear of Loss? ESL Students' Perspectives. *The Canadian Modern Language Review*, 72(1), 122–148. https://doi.org/10.3138/cmlr.2582

Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, 41(04), 647–672. https://doi.org/10.1017/S0272263119000044

Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and Instruction*, 52, 69–79. https://doi.org/10.1016/j.learninstruc.2017.04.007

Roussel, S., Tricot, A., & Sweller, J. (2022). The advantages of listening to academic content in a second language may be outweighed by disadvantages: A cognitive load theory approach. *British Journal of Educational Psychology*, 92(2), 627–644. https://doi.org/10.1111/bjep.12468

Stocker, L. (2017). The Impact of Foreign Accent on Credibility: An Analysis of Cognitive Statement Ratings in a Swiss Context. *Journal of Psycholinguistic Research*, 46(3), 617–628. https://doi.org/10.1007/s10936-016-9455-x

Intonation and voice quality in Icelandic wh-exclamatives

Nicole Dehé & Marieke Einfeldt

(Universität Konstanz, Germany)

Exclamatives (EXCL) are often considered one of four main sentence types (declaratives (DECL), interrogatives, imperatives, EXCL; e.g., Aikhenvald 2016). However, in many languages, EXCL have been shown to lack a syntactic form of their own. Instead, interrogative and declarative syntax is used to express exclamations (e.g., Rosengren 1997). In other words, to signal EXCL/exclamations, interlocutors rely on other formal means to signal illocution type, i.e., prosody (e.g., Batliner 1988). It has been shown that prosodic means, both tonal (pitch accents, boundary tones) and non-tonal (e.g., duration, voice quality) signal illocutionary force, e.g., the difference between declarative questions and declarative statements (Heuven & Haan 2002), and between information-seeking (ISQ) and rhetorical (RQ) questions (Dehé et al 2024 for an overview). Prosodic means have also been shown to disambiguate between EXCL and non-EXCL. In particular, compared to non-EXCL, EXCL use more prenuclear accents within one utterance as well as different accent types and different boundary tones (e.g., Repp 2020, Repp & Seeliger 2020 and Wochner 2022 for German, Sahkai et al. 2021 for Estonian, Sorianello 2012 for Italian). EXCL also make more frequent use of non-modal voice qualities (e.g., Wochner 2022 for German, Sahkai et al. 2021 for Estonian). In Icelandic, differences between DECL, polar questions and wh-questions are marked by the use of different pitch accent types (Árnason 2005, Dehé & Braun 2020), and ISQ differ in intonation, voice quality and duration from RQ (Dehé & Braun 2020, Dehé & Wochner 2024). We therefore predict to find prosodic differences between Icelandic EXCL and non-EXCL, as well.

Nothing is known about the prosody of Icelandic EXCL. Three types of EXCL have previously been discussed in the syntactic-semantic literature (wh-EXCL, XP-EXCL, M-EXCL; Jónsson 2010, 2017, Delsing 2010). In this presentation, we report results of our prosodic analysis of 8 wh-EXCL, produced by 17 speakers as part of a production experiment. They start with en hve ('but how'), e.g., (En hve)ω Lára málar fallega! 'How beautifully Lára paints!'. En hve functions as a degree element (see Rett's 2008 Degree restriction), placing the ability of painting at the high end of a scale (Zanuttini & Portner's 2003 Scalar implicature). String-identical non-EXCL sentences are not available for this exclamative type due to syntactic properties of Icelandic. We therefore compare our results with previous results for wh-questions (see Dehé & Braun 2020 for intonation, and Dehé & Wochner 2024 for voice quality), which are based on data produced by the same 17 speakers during the same experiment. The with-in-participant comparison allows for a reduction of noise introduced by individual differences.

Overall, we analysed 125 en hve-EXCL for intonation and voice quality. Our results confirm that Icelandic EXCL are indeed signaled and distinguished from other illocution types by prosody:

(i) Use of nuclear pitch accent types: H*/!H*/^H* is most frequent in wh-ISQ, L+H*/L+!H*/L+^H* is most frequent in wh-RQ, and H*/!H* and L+H*/L+^H* are of almost equal frequency in EXCL.

(ii) The prenuclear area: While L*+H is most typical in wh-RQ, and H* is typical in wh-ISQ, EXCL have mostly L+H* prenuclear pitch accents.

(iii) Boundary tones: The final boundary tone is L% by default across the board in Icelandic (Árnason 1998, 2005, 2011, Dehé & Braun 2020), and this is also what we find for

EXCL. Regarding initial boundary tones, wh-ISQ typically begin high (H* or %H; Árnason 2005, Dehé & Braun 2020). High beginnings are less frequent in en hve-EXCL.

(iv) Voice quality: ISQ, RQ and EXCL all show glottal voice quality especially towards the end of the utterance. However, both en hve-EXCL and wh-RQ exhibit more use of breathy voice quality than wh-ISQ, but compared to wh-RQ, en hve-EXCL use it more frequently also in utterance initial position.

Taken together, our results for Icelandic EXCL confirm that prosody signals illocution type, thus our results add to our knowledge about the syntax-prosody and prosody-discourse interfaces. Intonational cues as well as voice quality, are important parameters in this respect.

REFERENCES

Aikhenvald, A. Y. (2016). Sentence types. In J. Nuyts & J. van der Auwera (eds.): The Oxford Handbook of Modality and Mood, Oxford: Oxford University Press, 141–165.

Árnason, K. (2011). The phonology of Icelandic and Faroese. Oxford: Oxford University Press.

Árnason, K. (2005). Hljóð. Handbók um hljóðfræði og hljóðkerfisfræði. Íslensk Tunga, I. Bindi. Reykjavík: Almenna bókafélagið.

Árnason, K. (1998). Toward an analysis of Icelandic intonation. In Stefan Werner (ed.): Nordic Prosody. Proceedings of the VIIth conference, Joensuu 1996, Frankfurt a. M.: Peter Lang, 49–62.

Batliner, A. (1988). Der Exklamativ: Mehr als Aussage oder doch nur mehr oder weniger Aussage? In H. Altmann (ed.): Intonationsforschungen, Tübingen: Max Niemeyer Verlag, 243–271.

Dehé, N., Braun, B., Einfeldt, M., Wochner, D., Zahner-Ritter, K. (2024). The prosody of rhetorical questions: A cross-linguistic view. Linguistische Berichte, Sonderhefte 35, 103–148.

Dehé N., Braun, B. (2020). The intonation of information-seeking and rhetorical questions in Icelandic. Journal of Germanic Linguistics, 32(1), 1–42, doi: 10.1017/S1470542719000114.

Dehé N., Wochner, D. (2024). Voice quality and speaking rate in Icelandic rhetorical questions. Nordic Journal of Linguistics 47(1), 111-120, doi: 10.1017/S0332586522000014.

Delsing, L.-O. (2010). Exclamatives in Scandinavian. Studia Linguistica, 64(1), 16–36.

Heuven, V. J. v., & Haan, J. (2002). Temporal distribution of interrogativity cues in Dutch: A perceptual study. In C. Gussenhoven & N. Warner (eds.): Papers in Laboratory Phonology 7, Berlin, Germany: Mouton de Gruyter, 61–86.

Jónsson, J. G. (2010). Icelandic exclamatives and the structure of the CP layer. Studia Linguistica, 64(1), 37–54.

Jónsson, J. G. (2017). Discourse particles and hvað-exclamatives. In J. Bayer & V. Struckmeier (eds.): Discourse particles: Formal approaches to their syntax and semantics (Linguistische Arbeiten 564), Berlin: de Gruyter, 100–114.

Sahkai, H., Asu, E. L., Lippus, P. (2021). Prosodic characteristics of exclamatives and questions in Estonian. Proceedings of the 1st International Conference on Tone and Intonation (TAI), 41–45, doi: 10.21437/TAI.2021-9.

Sorianello, P. (2012). A prosodic account of Italian exclamative sentences: A gating test. Proceedings of Speech Prosody 2012, 298–301, doi: 10.21437/SpeechProsody.2012-76.

Rett, J. (2008). Degree modification in natural language. PhD dissertation, Rutgers University, New Brunswick.

Repp, S. (2020). The prosody of wh-exclamatives and wh-questions in German: Speech act differences, information structure, and sex of speaker. Language and Speech, 63(2), 306–361, doi: 10.1177/0023830919846147.

Repp, S., Seeliger, H. (2020). Prosodic prominence in polar questions and exclamatives. Frontiers in Communication, 5(53), doi:10.3389/fcomm.2020.00053.

Rosengren, I. (1997). Expressive sentence types — a contradiction in terms. The case of exclamation. In T. Swan & O. J. Westvik (eds.): Modality in Germanic languages: Historical and comparative perspectives, Berlin, New York: De Gruyter Mouton, 151–184, doi: 10.1515/9783110889932.151.

Wochner, D. (2022). Prosody meets pragmatics: A comparison of rhetorical questions, information-seeking questions, exclamatives, and assertions. PhD dissertation, University of Konstanz.

Zanuttini, R., Portner, P. (2003). Exclamative clauses: At the syntax-semantics interface. Language 79, 39–81.

19:30 Gala dinner

Day 2: Friday, May 22nd

08:30 Welcome coffee

09:00 Plenary 3

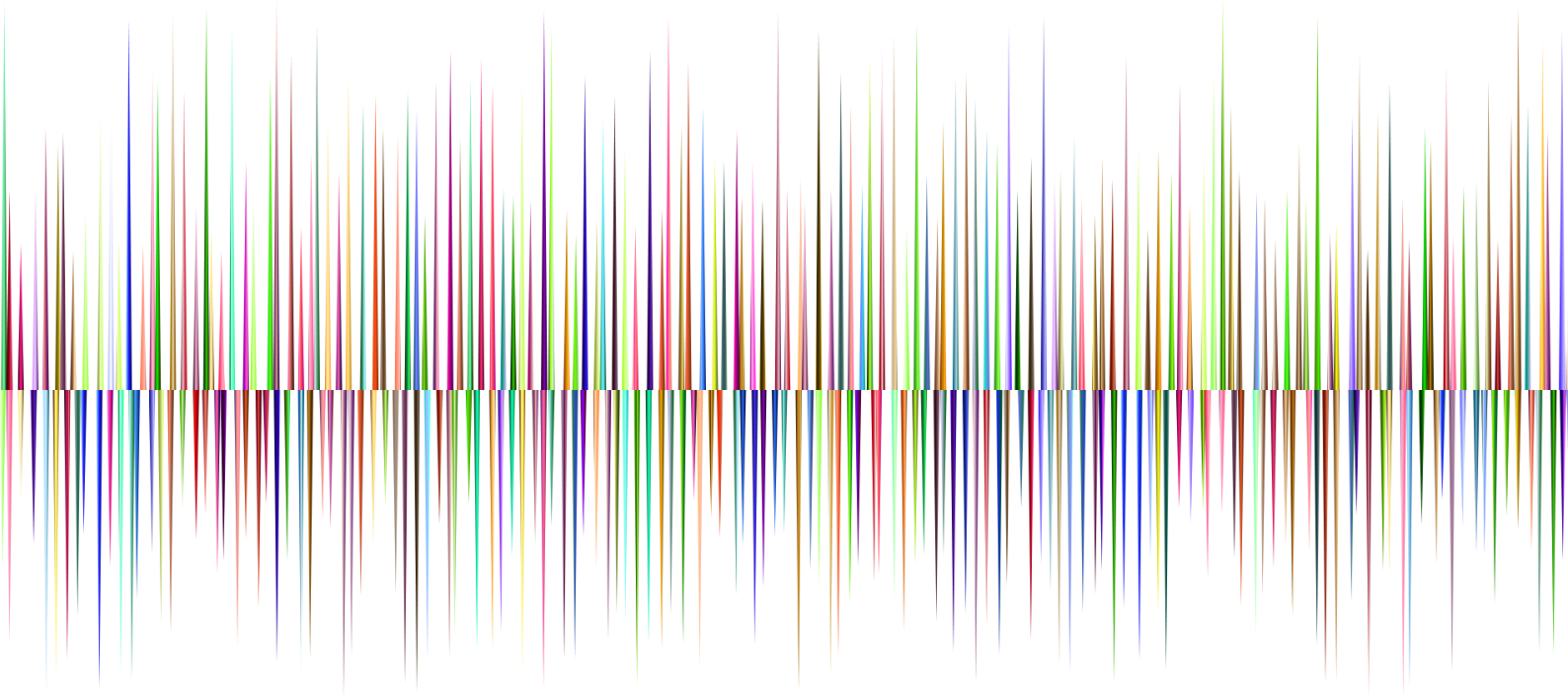
Jane Stuart Smith (University of Glasgow) *Insights from sociophonetic studies of Scottish English*

10:10 Parallel Sessions 5 (Rooms 1, 2, 3)

Room 1: Claire Pillot-Loiseau, Céline Horgues, Maxime Klingelschmitt & Sylwia Scheuer-Samson (Université Sorbonne Nouvelle, France) *Exploring local and global voice quality adjustments during English/French tandem interactions*

Room 2: Stella Ville (Université Grenoble Alpes, France) Cross-linguistic Interference in the Acquisition of Lexical Stress in L2 English

Room 3: Sampa Bestavasvili (Universität Freiburg, Germany) The Role of Voice Quality and Sex Perception in Age Estimation of Speakers with Dysphonia



Keynote speaker

Taking the rough with the smooth: Insights from sociophonetic studies of Scottish English voice quality

Jane Stuart Smith

(University of Glasgow)

Voice quality is a pervasive feature of speech which characterizes accents, dialects, individual speakers and even their emotional states, and likely contributes to segmental phonetics for accents (Abercrombie, 1967; Laver, 1991; Trudgill, 1974). Speaker voice quality is both immediately accessible to listeners, and yet at the same time, has proved remarkably tricky for phoneticians to pin down, leading to numerous different approaches (Kreiman, e.g. 2024).

In this talk, we begin by examining voice quality in the holistic sense of articulatory settings, using the specific case study of Scottish dialect of Glaswegian vernacular, which is infamous for its distinctive stereotype of harsh voice and protruded jaw, features also associated with aggression. We will start by considering evidence from socially-stratified study of Glaswegian voice quality using the auditory Vocal Profile Analysis (VPA) tool (Stuart-Smith, 1999, after e.g. Laver, 1991; San Segundo et al, 2017).

We will then consider both VPA and acoustic phonetic evidence which demonstrates how Glaswegian voice quality has changed over the 20th century, with implications for segmental features (Soskuthy and Stuart-Smith, 2021). We then take a step back to consider how Glaswegian voice quality fits into the wider picture of voice quality in Scottish English, by focusing at phonatory differences (Pearce, 2023). We will conclude by thinking about the developmental trajectory of Scottish voice quality, by looking at a large-scale child speech corpus study (Murali et al 2025).

REFERENCES

Abercrombie, D. (1967). Elements of General Phonetics. Edinburgh: Edinburgh University Press.

Kreiman, J. (2024). Information conveyed by voice quality. The Journal of the Acoustical Society of America, 155(2), 1264–1271.

Laver, J. (1991). The Gift of Speech: Papers in the Analysis of Speech and Voice. Edinburgh: Edinburgh University Press.

Murali, M., Cleland, J., Taylor, L., Young, D., Stuart-Smith, J., & Kuschmann, A. (2025). A study of voice quality and acoustic variability in sound prolongation performance in 5–12-year-old children. Folia Phoniatrica et Logopaedica, 78(1), 1–26.

Pearce, J. (2023). Creaky voice in three Scots varieties: Using F0 based identification to consider social and linguistic factors. In R. Skarnitzl & J. Volín (eds.), Proceedings of the 20th International Congress of Phonetic Sciences, Prague, Czechia (1816–1820).

Sóskuthy, M., & Stuart-Smith, J. (2020). Voice quality and coda /r/ in Glasgow English in the early 20th century. Language Variation and Change, 32(2), 133–157. (Note : l'article a été accepté/publié fin 2020/2021).

San Segundo, E., & Mompeán, J. A. (2017). A Simplified Vocal Profile Analysis Protocol for the Assessment of Voice Quality and Speaker Similarity. *Journal of Voice, 31*(5), 644-653.

Trudgill, P. (1974). The Social Differentiation of English in Norwich. Cambridge: Cambridge University Press.

Exploring local and global voice quality adjustments during English/French tandem interactions

Claire Pillot-Loiseau, Céline Horgues, Maxime Klingelschmitt & Sylwia Scheuer-Samson

(Université Sorbonne Nouvelle, France)

Tandem learning is an informal L2 learning context in which two speakers of two different languages form a colearning contract whereby each participant takes on both the role of the language 'expert' when speaking in their L1 but also that of the language 'novice' when speaking their foreign language (Brammerts & Calvert, 2003; Helming, 2002). Role reversal (or 'symmetrisation' of linguistic asymmetries in the terms of Vassallo & Telles, 2006) is induced when tandem participants switch languages, as reciprocity, along with learner autonomy, are key principles of tandem set-ups). Both role reversal (expert/novice) and language switch may promote pragmatic and linguistic adjustments in the speech of tandem participants. On the phonetic level in particular, when speakers speak another language than their L1 (Anderson-Hsieh et al., 1992; Gut, 2009) , or when they address a foreign-interlocutor (Long, 1983; Knoll et Scharrer, 2007; Piazza et al., 2025), they may adjust their articulation settings (Uther et al., 2007; Kangatharan et al., 2012), speech rate (Kühnert & Kocjančič Antolik, 2017), intonation (Wennestrom, 1994; Gut, 2003, Biersack et al., 2005; Delais-Roussarie et al., 2015; Embarki et al., 2023) but also their voice quality although this aspect has received considerably less attention (Pépiot, 2014; Brown & Sonderegger, 2025).

This paper investigates some of the voice quality adjustments occurring during English/French tandem interactions based on the video-recorded of 21 face-to-face English/French tandem dyads (SITAF corpus, Horgues & Scheuer, 2015), with a focus on the collaborative reading task in French and in English (North Wind and the Sun/La Bise et le Soleil) by female tandem partners at the start of their tandem partnership (session1, with readings in L2 only) and then at the end 3 months later (session2, with readings in L1 and L2). Drawing upon previous research on voice quality in this speech data (Pillot-Loiseau et al., 2019; Pillot-Loiseau et al., 2023), we aim to discuss the methodology of voice quality explorations to characterize speakers' voice quality and their ability to adjust it. Indeed, we will show how voice quality can be approached from a global perspective across the entire stretch of speech (with global measures such mean fundamental frequency (fo), spectral slope and CPPS (Cepstral Peak Prominence Smoothed, used by Pillot-Loiseau et al., 2023) or from a more local perspective (with local measures such as the frequency and duration of creaky voice occurrences, Pillot-Loiseau et al., 2019 or, when the laboratory experimental conditions allow it, physiological explorations such as electroglottography like in Benoist-Lucy & Pillot-Loiseau, 2013) enabling reliable measurements such as H1*-H2* (Garellek, 2015).

We will show how these two types of approaches come with assets, but also limitations. If global measures are useful to reflect an overall, holistic impression of speakers' voice timbre, they also flatten out meaningful variations in voice quality at utterance level, and are often discorrelated from verbal content, information structure or prosodic organization. If local measures only provide a fragmented characterization of voice quality, they may positively complement global measures, with a focus on certain aspects of voice quality. For example, we resort to calculating the proportion of creaky voice occurrences relative to the number of pronounced syllables (e.g. Pillot et al., 2019). Another refinement we propose consists in analyzing the distribution and scope (e.g. one segment, one syllable, one word, one prosodic unit) of such creaky occurrences, together with their prosodic position (stressed/unstressed syllable, alignment with right/left prosodic boundary, pre/post nucleus, coincidence with

hesitation phenomena). Other local phenomena like syllable-final glottal replacement or syllable-initial pre-vocalic glottal reinforcement may also play an important role in voice quality characterization.

By combining global and local approaches to voice quality analysis, our preliminary results indicate: 1) a language effect (French vs. English): locally, instances of creaky voice are more numerous and longer in English. Globally, fo is lower in English, the spectral slope is less negative, and CPPS is higher, indicating a more timbred voice in English, especially for English speakers; 2) an effect of language status (L1 vs. L2): locally, the duration of creaky voice occurrences is shorter in the participants' L2 than in their L1. Overall, fo is lower for French L2 but higher for English L2; CPPS is lower and the spectral slope is more negative in French or English L2 than in the L1, especially among English speakers; 3) an effect of tandem experience/familiarity (comparison between sessions 1 and 2): locally, the proportion of creaky voice occurrences in the L2 decreases from session 1 to session 2, especially among English speakers. The Native-English speakers in the SITAF corpus thus seem to show more flexibility than their French counterparts in changing their voice quality between their L1 and their L2, probably because they had more opportunities to interact in the target language at the time of the study. Overall, the spectral slope is less negative in the L2 among English speakers in session 2 compared to session 1; 4) a correlation between certain local and global indicators of voice quality: the greater the number of instances of creaky voice, the greater the total number of syllables and the lower the fo.

Linguistic factors (linguistic role of voice quality in L1 French and L1 English), but also contextual factors (the speaker's familiarity with the L2 and amount of L2 interactional experience) will be evoked to account for these results.

REFERENCES

Anderson-Hsieh, J., Johnson, J., Koehler, K. (1992). The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, 42(4), 529–555.

Benoist-Lucy, A., & Pillot-Loiseau, C. (2013). The Influence of language and speech task upon creaky voice use among six young American women learning French, *Interspeech 2013*, 2395-2399. International Speech Communication Association.

Biersack, S., Kempe, V. & Knapton, L. (2005). Fine-Tuning Speech Registers: a Comparison of the Prosodic Features of Child-Directed and Foreigner-Directed Speech, *Interspeech 2005*, 2401-2404.

Brammerts, H., & Calvert, M. (2003). Learning by communicating in tandem. In T. Lewis and L. Walker (eds) *Autonomous Language Learning in Tandem*. Sheffield: Academy Electronic Publications: 45-60.

Brown, J. & Sonderegger, M. (2025). A sociophonetic study of creaky voice across language, gender and age in Canadian English-French bilinguals, *Journal of Phonetics*, 112, 101431

Delais-Roussarie, E., Anvanzi, M., Herment, S. (2015). *Prosody and Language in Contact, L2 Acquisition, Attrition and Languages in Multilingual Situations*, Berlin: Springer-Verlag.

Embarki, M., Ziamar, K., Ho, L. W., Wei, D. (2023). Variation de fréquence fondamentale en L1 et L2 : cas des apprenants de français arabophones et malaisophones, *Langages*, 230(2), 79-98.

Garellek, M. (2015). Perception of glottalization and phrase-final creak. *The Journal of the Acoustical Society of America*, 137(2), 822-831.

Gut, U. (2003). Prosody in second language speech production: the role of the native language. *Fremdsprachen Lehren und Lernen*, 32, 133–152.

Gut, U. (2009). *Non-native speech. A corpus-based analysis of the phonetic and phonological properties of L2 English and L2 German*. Frankfurt: Peter Lang.

Kangatharan, J., Uther, M, Kuhn, L., Gobert, F (2012). A-M I S-P-EA-K-I-NG C- L- E- AR-L- Y E-N-OU-GH?: An investigation of the possible role of vowel hyperarticulation in speech communication, *Acoustics 2012*, 3943-3947.

Knoll, M. A., & Scharrer, L. (2007). Acoustic and affective comparisons of natural and imaginary infant-, foreigner- and adult directed speech. *Interspeech 2007*, 4. https://doi.org/10.21437/Interspeech.2007-29

Kühnert, B., & Kocjančič Antolík, T. (2017). Patterns of articulation rate variation in English/French tandem interactions. In J. Volin & R. Skarnitzl (Eds.), *Pronunciation of English by speakers of other languages*. Cambridge Scholars. https://halshs.archives-ouvertes.fr/halshs-01505928

Helmling, B. (2002). *L'apprentissage autonome des langues en tandem*. [Autono mous language learning in tandem]. Paris: Didier.

Horgues, C., & Scheuer, S. (2015). Why some things are better done in tandem. In *Investigating English pronunciation: Trends and directions* (pp. 47-82). London: Palgrave Macmillan UK.

Long, M. (1983). Linguistic and Conversational Adjustments to Non-Native-Speakers. *Studies in Second Language Acquisition* 5(2), 177-193.

Pépiot, E. (2014). Male and female speech: A study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *Proceedings of Speech Prosody, 2014*, 305–309. https://doi.org/10.21437/SpeechProsody.2014-49.

Piazza, G., Kalashnikova, M., Fernández-Merino, L., & Martin, C. D. (2025). Speakers' communicative intentions lead to acoustic adjustments in native and non-native directed speech. *Speech Communication*, 103250.

Pillot-Loiseau, C., Horgues, C., Scheuer, S., & Kamiyama, T. (2019). The evolution of creaky voice use in read speech by native-French and native-English speakers in tandem: A pilot study. *Anglophonia. French Journal of English Linguistics*, (27). https://journals.openedition.org/anglophonia/2005

Pillot-Loiseau, C., Harmegnies, B., Horgues, C., & Scheuer, S. (2023). Qualité vocale en lecture par des locutrices anglophones et francophones : comparaison acoustique avant et après 12 séances en tandem. *Langages*, 230(2), 59-78.

Uther, M., Knoll, A. & Burnham, D. (2007). Do you speak E.N.G.L.I.S.H? A comparison of foreigner and infant directed speech, *Speech Communication*, 49, 2-7.

Vassallo, M. L., & Telles, J. A. (2006). Foreign language learning in-tandem: Theoretical principles and research perspectives. *The ESPecialist*, 27(1), 83–118.

Wennerstrom, A. (1994). Intonational Meaning in English Discourse: A Study of Non-Native Speakers, *MAUSApplied Linguistics*, 15(4), Oxford: Oxford University Press, p. 399-420.

ACKNOWLEDGMENTS

This work has been supported by Partnerships Hubert Curien (PHC) TOURNESOL nos. 51884UJ and ANR-18-IDEX-0001. It contributes to the IdEx Université de Paris (ANR-18-IDEX-0001).

Cross-linguistic Interference in the Acquisition of Lexical Stress in L2 English: A Study with Hispanic Learners

Stella Ville

(Université Grenoble Alpes, France)

Research in language acquisition relies predominantly on university populations (Andringa & Godfroid, 2020), referred to as WEIRD (Henrich et al., 2010). By contrast, learners outside these profiles, here referred to as WISER[1] (Ville, 2025), remain underrepresented (Mathews-Aydinli, 2008). This paper contributes to this underexplored field by examining the effects of a prosody awareness programme in English on Hispanic learners (A1–A2), within an action research project conducted in 2023 at the Escuela Oficial de Idiomas of Palma de Mallorca. The intervention involved 82 learners (aged 16–67) over 12 weeks, following a multimodal and embodied approach (Baills et al., 2022; Chan, 2018). The study examines the effects of this approach on the development of prosodic awareness, its impact on oral performance, and its affective dimensions. The aim is to uncover the perceived barriers and benefits reported by both learners and teachers (Ville, 2025; forthcoming).

The corpus combines a lexical stress placement test, questionnaires and self-assessments, as well as delayed interviews with both learners and teachers. It also includes pre/post oral productions (reading aloud, repetition) analysed using PLSPP (Nakanishi & Coulange, 2024), an automatic tool for analysing lexical stress realisation. Following Mennen's (2006) distinction between phonological and phonetic influence, stress placement tasks target the phonological level (mental representations and organisation of the prosodic system), whereas acoustic analyses (vowel centralisation, dispersion) pertain to the phonetic level (the concrete acoustic realisation of prosodic categories, which may diverge across languages despite apparent equivalence).

An initial analysis (Ville & Rossato, 2026) of the phonological awareness test reveals differentiated effects according to the degree of cross-linguistic interference (Rasier & Hiligsmann, 2007): significant progress for incomparable words (e.g. tomorrow), persistent difficulty for misleading words (e.g. telephone), and a ceiling effect for similar words (e.g. plastic). Progress is more marked among bilingual (with varied L1 combinations) and A2-level learners. In oral production tasks, results show a bilingual advantage in phonological awareness, improvements in repetition across all groups, but no time effect in the reading, with a surprising advantage for monolingual learners. Progress in the perception and identification of stress patterns suggests a partial reorganisation of the phonological system, while persistent difficulties in production reflect phonetic interference from the L1, potentially requiring more sustained sensorimotor training. These findings suggest that the intervention first promotes a reconfiguration of prosodic representations, with phonetic adjustments appearing more limited. This dissociation confirms that prosodic acquisition involves distinct levels, liable to develop at different rates, and helps account for the gap between sensitivity to prosodic contrasts and production performance.

This paper extends the analysis by applying the lexical typology to the oral production data, to examine the relationship between phonological awareness and acoustic realisation, particularly in reading aloud. This approach aims to identify differentiated developmental trajectories according to learner profiles and to refine our understanding of prosodic acquisition mechanisms. Pedagogical implications concern the adaptation of training programmes to learner profiles, as well as the need to combine perceptual work and phonetic training within a multimodal and embodied framework.

[1] **WISER** learners are **W**idely varied (circumstances, cognitions, etc.), **I**nformally trained (limited formal education and language-learning experience), **S**trained-resource (time, energy, finances), **E**xperience-driven (motivated by tangible, practical outcomes) and **R**esistant to change (due to entrenched beliefs and past schooling experiences).

REFERENCES

Andringa, S., & Godfroid, A. (2020). Sampling Bias and the Problem of Generalizability in Applied Linguistics. *Annual Review of Applied Linguistics*, 40, 134–142. doi:10.1017/S0267190520000033

Bailles, F., Rohrer, P. L., & Prieto, P. (2022). Le geste et la voix pour enseigner la prononciation en langue. Visuospatial Gestures in Second Language Learning View project. *Mélanges Crapel*. https://www.researchgate.net/publication/359705064

Chan, M. J. (2018). Embodied Pronunciation Learning: Research and Practice. *The CATESOL Journal*, 30.1, 47–68.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? In *Behavioral and Brain Sciences* (Vol. 33, Numbers 2–3, pp. 61–83). Cambridge University Press. https://doi.org/10.1017/S0140525X0999152X

Mathews-Aydinli, J. (2008). Overlooked and understudied? a survey of current trends in research on adult english language learners. *Adult Education Quarterly*, 58(3), 198–213. https://doi.org/10.1177/0741713608314089

Mennen, I. (2006), "Phonetic and phonological influences in non-native intonation: an overview for language teachers", *Working Papers Queen Margaret University College* 9.

Nakanishi, N., & Coulange, S. (2024, July). Measuring speech rhythm through automated analysis of syllabic prominences. *Satellite Workshop of Speech Prosody*. https://hal.science/hal-04666098v1

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Cahiers de Linguistique Francaise*, (28), 41–66. http://hdl.handle.net/2078.1/86887

Ville, S. (2025). Language teachers' cognitions and the impact of approximation training observation. *PHONICA*, 21, 1–29. https://doi.org/10.1344/phonica2025.21.5

Ville, S. & Rossato, S. (2026), Sensibilisation à la prosodie et acquisition de l'accentuation en anglais, à paraitre.

The Role of Voice Quality and Sex Perception in Age Estimation of Speakers with Dysphonia

Sampa Bestavasvili

(Universität Freiburg, Germany)

Objective: To explore how voice quality affects our perception of age from individuals with voice disorders.

Background: Perceived low voice quality has been associated with a tendency of listeners to overestimate the age of healthy speakers (Goy et al. 2016). Sex is also correlated with decreased vocal quality, especially in healthy older females (Gorham-Rowan & Laures-Gore 2006). A study on dysphonic speakers by Harnsberger et al. (2010) showed that male speakers with dysphonia were judged older compared to male healthy speakers, yet their degree of their voice severity was not determined. This paper tested how different degrees of voice severity can affect vocal age perception in both male and female dysphonic speakers.

Method: Speech and voice samples were retrieved from 8 dysphonic speakers (older ~ 70 yrs: 2 female and 2 male | younger ~ 20 yrs: 2 female and 2 male) from the PVQD by Walden (2022). In each pair, speakers were best matched for age, sex, and voice disorder, but differed in the degree of voice severity; one had a lower degree and the other one higher. Professional assessment of the degree of severity was provided in the database and was based on the CAPE-V and GRBAS scales. 102 naïve participants then rated the speakers' age, sex, and voice quality.

Results: Listener perceived voice quality resembled the professional ratings. Biological male speakers with low voice quality were rated older (\beta = 26.00, \text{SE} = 7.58, $p < .05$), while perceived females with low quality were also rated older (\beta = 17.79, \text{SE} = 5.24, $p < .01$). A follow-up model revealed that biological males with low voice quality who were misperceived as female were the ones that were perceived older (\beta = 25.22, \text{SE} = 8.60, $p < .01$).

Conclusion: Sex misperception accompanied by low voice quality can impact how we perceive age in male persons with dysphonia. Clinicians can gain insight into the implications of voice quality in vocal aging and explain some of the inconsistencies observed in the literature.

REFERENCES

Gorham-Rowan, Mary M. & Jacqueline Laures-Gore. 2006. Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders* 39(3). 171–184. https://doi.org/10.1016/j.jcomdis.2005.11.005.

Goy, Huiwen, M. Kathleen Pichora-Fuller & Pascal Van Lieshout. 2016. Effects of age on speech and voice quality ratings. *The Journal of the Acoustical Society of America* 139(4). 1648–1659. https://doi.org/10.1121/1.4945094.

Harnsberger, James D., William S. Brown, Rahul Shrivastav & Howard Rothman. 2010. Noise and Tremor in the Perception of Vocal Aging in Males. *Journal of Voice* 24(5). 523–530. https://doi.org/10.1016/j.jvoice.2009.01.003.

Walden, Patrick R. 2022. Perceptual Voice Qualities Database (PVQD): Database Characteristics. *Journal of Voice* 36(6). 875.e15–875.e23. https://doi.org/10.1016/j.jvoice.2020.10.001.

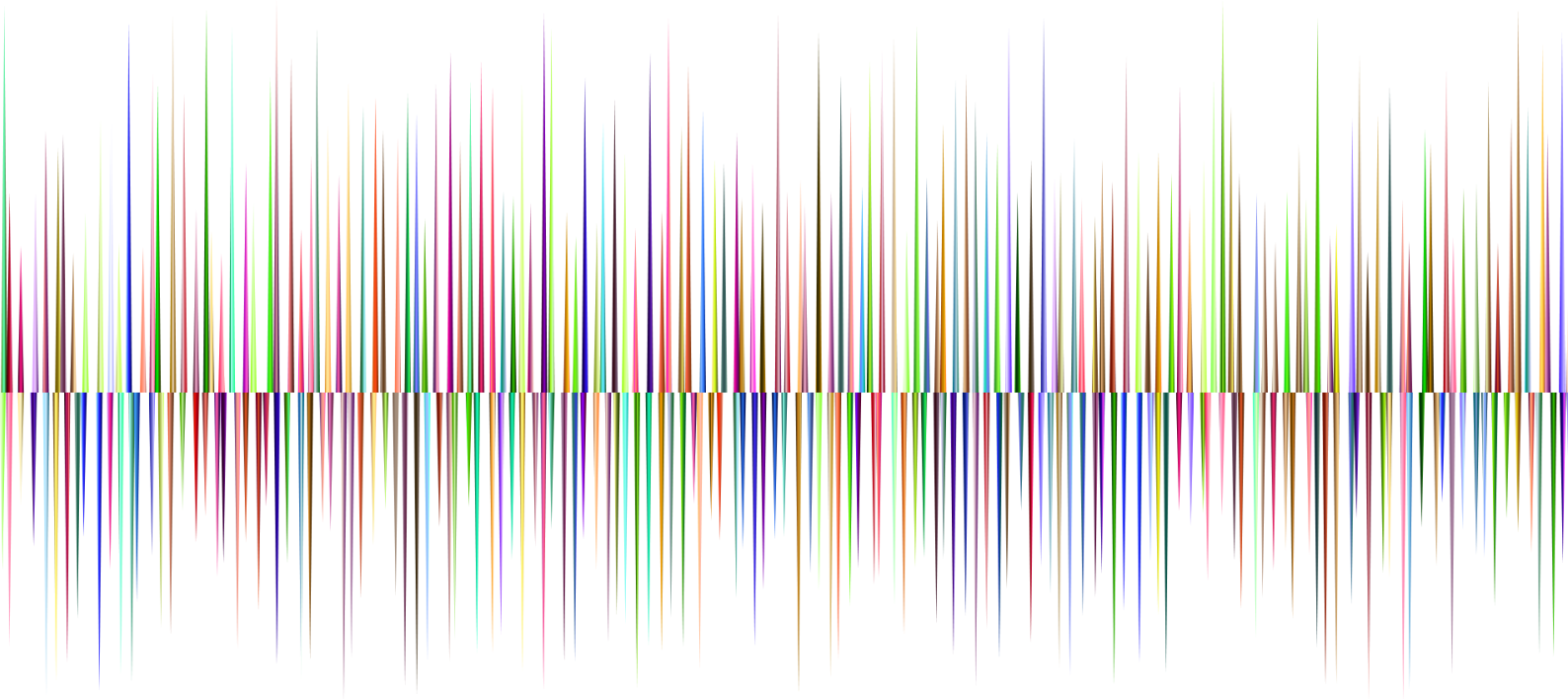
10:55 Coffee break

11:25 Parallel Sessions 6 (Rooms 1, 2, 3)

Room 1: **Pamela Mary Rogerson Revell & Martha Pennington** (University of Leicester / Birkbeck, University of London, United Kingdom) *The role of prosody and voice quality in L2 phonological acquisition*

Room 2: **Antoine Regis, Sophie Herment & Amandine Michelas** (Aix-Marseille Université, France) *How Do French Learners Perceive English Pitch Accent Contrasts?*

Room 3: **Sofia Sedunova** (Higher School of Economics, Russia) *Focus Prosody in Shughni Noun Phrases*



The role of prosody and voice quality in L2 phonological acquisition: a paradigm shift in pronunciation teaching and research

Pamela Mary Rogerson Revell & Martha Pennington

(University of Leicester / Birkbeck, University of London, United Kingdom)

Traditionally, segmental phonology has been seen as most important in second language phonological acquisition and pronunciation teaching, with prosody given less priority and voice quality rarely considered at all. Within the recent paradigm shift in L2 pronunciation research and teaching that has foregrounded the intelligibility principle (Levis, 2005), prosody is seen as equally important or more important than segmental aspects (Kazu & Kuvvetli, 2023; Ma et al., 2024; Pennington & Rogerson-Revell, 2019; Yenkimaleki & Van Heuven, 2021). Drawing on contemporary research and pedagogical frameworks, we trace this shift and examine its implications for pronunciation teaching and applied research.

Prosody and voice quality are distinct but deeply intertwined dimensions of speech. They share physiological mechanisms, most notably laryngeal activity, in which fundamental frequency (F0) underlies prosodic intonation while the manner of vocal fold vibration simultaneously shapes voice quality, and subglottal pressure influences both loudness and phonation type. Prosody and voice quality co-occur in expressing meaning and emotion; anger, for instance, may involve raised pitch alongside a tense or harsh voice quality, while sadness may combine lower pitch with breathiness. They can be difficult to disentangle functionally as well as acoustically, given that they share key acoustic parameters such as F0, spectral slope, and harmonics-to-noise ratio. Arguably, both prosody and voice quality are aspects of suprasegmental phonology significant to a speaker's overall communicative repertoire, style and sense of identity (Pennington & Rogerson-Revell, 2019).

Theoretically, the relationship between prosody and voice quality has been understood in different ways. Laver's (1980) influential framework treated voice quality as a long-term background setting upon which short-term prosodic modulations are superimposed; a distinction that has proven useful analytically even while acknowledging their interaction. More recent work in intonational phonology, including Gussenhoven (2004), has increasingly treated phonation type as an integral part of the prosodic system rather than a separate domain, further underscoring the need to consider these dimensions together.

Though historically situated outside mainstream phonology, the role of voice quality and L2 articulatory settings have long been recognised as important for effective L2 acquisition (Honikman, 1964; Laver, 1980; O'Connor, 1973), and have been gaining renewed attention in L2 pronunciation teaching and research (Ding et al., 2019; Henderson & Skarnitzl, 2022; Messum, 2017; Wilson & Gick, 2014). Recent experimental evidence confirms that proficient bilinguals maintain distinct articulatory settings per language (Wilson & Gick, 2014), lending empirical weight to pedagogical approaches that foreground voice quality in pronunciation instruction. Research also consistently demonstrates that suprasegmental instruction, targeting rhythm, stress, and intonation, improves both comprehensibility and fluency, with studies showing that prosodic cues also influence segmental accuracy and support vocabulary retention (Kazu & Kuvvetli, 2023; Ma et al., 2024; Yenkimaleki & Van Heuven, 2021).

In our presentation we argue for a greater focus on prosody and voice quality in L2 teaching by reviewing the relevant phonological and pedagogical theory and conceptual frameworks outlined above, supported by video demonstrations of prosody and articulatory settings in English and other languages. We then present some applications to L2 pedagogy

emphasizing top-down approaches with a focus on intelligibility and communicative effectiveness, supplemented by activities for raising awareness with multimodal input and applications of AI technology. By synthesising research strands that are often treated separately and relating these to pedagogic approaches, we aim to offer a unified account of the paradigm shift underway in L2 pronunciation. The findings are relevant to language teachers, curriculum designers, and researchers working at the intersection of phonetics, applied linguistics, and educational technology.

REFERENCES

Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., Chukharev-Hudilainen, E., Levis, J., & Gutierrez-Osuna, R. (2019). Golden speaker builder: An interactive tool for pronunciation training. *Speech Communication*, 115, 51–66. John Benjamins https://doi.org/10.1016/j.specom.2019.10.005

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Henderson, A. J., & Skarnitzl, R. (2022). "A better me": Using acoustically modified learner voices as models. *Language Learning & Technology*, 26(1), 1–21. ScholarSpace http://hdl.handle.net/10125/73462

Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *In honour of Daniel Jones* (pp. 73–84). Longman.

Kazu, İ. Y., & Kuvvetli, M. (2023). The influence of pronunciation education via artificial intelligence technology on vocabulary acquisition in learning English. *International Journal of Psychology and Educational Studies*, 10(2), 480–493. Ijpes https://doi.org/10.52380/ijpes.2023.10.2.1044

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377.

Ma, Q., Mei, L., & Qian, X. (2024). Exploring EFL students' pronunciation learning supported by corpus-based language pedagogy. *Computer Assisted Language Learning*. Taylor & Francis Online https://doi.org/10.1080/09588221.2024.2432965

Messum, P., & Young, R. (2017). Bringing the English articulatory setting into the classroom: (1) The tongue. *Speak Out!*, 57, 29–39. Sage Journals

O'Connor, J. D. (1973). *Phonetics*. Penguin.

Pennington, M. C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research: Contemporary perspectives*. Palgrave Macmillan.

Wilson, I., & Gick, B. (2014). Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research*, 57(2), 361–373.

Yenkimaleki, M., & Van Heuven, V. J. (2021). Effects of attention to segmental vs. suprasegmental features on the speech intelligibility and comprehensibility of the EFL

learners targeting the perception or production-focused practice. *System*, 100, 102557. Wiley

How Do French Learners Perceive English Pitch Accent Contrasts?

Antoine Regis, Sophie Herment & Amandine Michelas

(Aix-Marseille Université, France)

Pitch accents are central prosodic elements that encode information structure (e.g., given, new, or contrastive information) and discourse meaning (e.g., speaker uncertainty or reservations) through variations in fundamental frequency (F0). Within the Autosegmental-Metrical theory (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986), pitch accents consist of High (H) and Low (L) tones and may be monotonal (e.g., H*) or bitonal (e.g., L+H*, with the asterisk indicating the head tone, the tone most closely aligned with the stressed syllable). Standard American English has a rich pitch accent inventory, including H* (salient/new information), L* (given information), L+H* (contrastive information), L*+H (speaker uncertainty), and H+L* (speaker reservation; e.g., Pierrehumbert, 1980). By contrast, French has a very limited inventory, with H* being extremely frequent and a more variable L* typically occurring in utterance-final position (e.g., Delais-Roussarie et al., 2015). Importantly, unlike English, French pitch accents systematically affect the final syllable of words in phrase-final position and are never realized on word-initial or word-medial syllables. These differences between the English and French inventories of pitch accents are likely to have important consequences for speech perception. However, although numerous studies have focused on the perception of lexical stress contrasts by French learners (e.g., Dupoux et al., 2008; Schwab et al., 2020), relatively little is known about their perception of pitch accent contrasts in English. The present study addresses this gap.

To investigate this question, we tested sixty French late learners of English (who began formal instruction after age 10 and were still receiving instruction at the time of testing) and twenty native speakers of American English. Learners were divided into beginner, intermediate, and advanced groups based on their LexTALE scores (Lemhöfer et al., 2012). Participants completed an ABX discrimination task in which they heard trisyllabic first names, all stressed on the second syllable, produced by three American speakers, and indicated whether X matched A or B. Given that French learners are mostly exposed to a pitch accent in their L1 that closely resembles the English H*, we examined contrasts between H* and other pitch accents attested in English (L+H*; H*+L; L*+H; L*; Fig. 1). Performance on the contrasts involving the first three pitch accents mentioned (H*–L+H*; H*–H*+L*; H*–L*+H) was compared to the H*–L* contrast, which most closely resembles the only pitch accent contrast found in French.

Accuracy data (Fig. 2) were analyzed using Generalized Linear Mixed-effects Models (GLMMs) with a binomial distribution and a logit link function. The results revealed a significant main effect of proficiency ($\chi^2 = 56.63$, $p < .0001$; Beginner < Intermediate < Advanced < Native), showing that performance in discriminating pitch accents increases with L2 proficiency. In addition, a main effect of Pitch Accent Type ($\chi^2 = 91.08$, $p < .0001$) showed that participants performed similarly on the H*–L*+H contrast compared to the H*–L* contrast, since both contrasts involve two different head tones (H vs. L). By contrast, they had more difficulties discriminating contrasts sharing the same H head tone (H*–H*+L and H*–L+H*) compared to the H*–L* contrast. Crucially, a significant interaction between Pitch Accent Type and Proficiency level ($\chi^2 = 18.87$, $p < .05$) showed that beginners had greater difficulty discriminating the other contrasts compared to the H*-L* contrast, irrespective of their head-tone configuration. By contrast, intermediate and advanced learners performed equally well on the H*-L* contrast and on the other contrast involving two different head tones (i.e., H*-L*+H), like native American speakers.

These findings, in line with previous observations (Llanos et al., 2021), indicate that the head tone serves as a robust perceptual anchor for discriminating pitch accents. Importantly, our results also show that learners gradually learn to attend to the head-tone cue as their L2 proficiency increases, revealing a progressive fine-tuning of perceptual sensitivity. The study provides evidence that L1 pitch accent inventory constrains L2 acquisition.

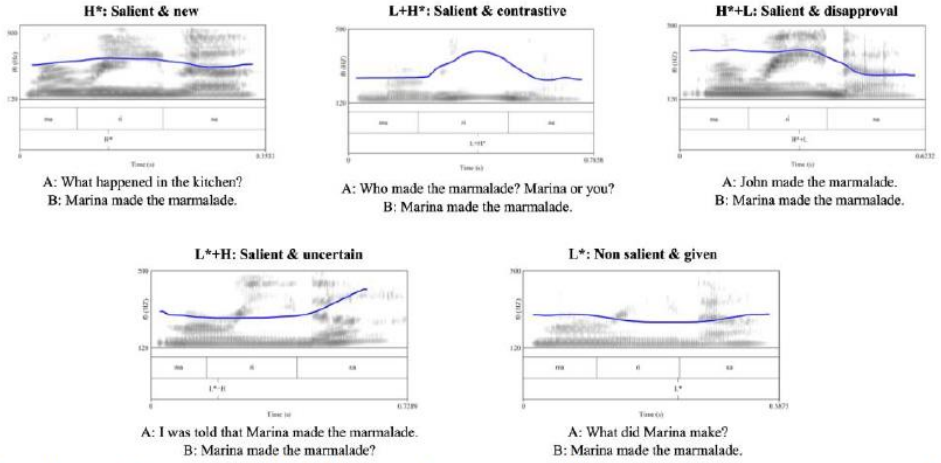


Fig. 1. Prosodic realization of the five pitch accents, illustrated here on the name *Marina*, as produced by one of the native American English speakers.

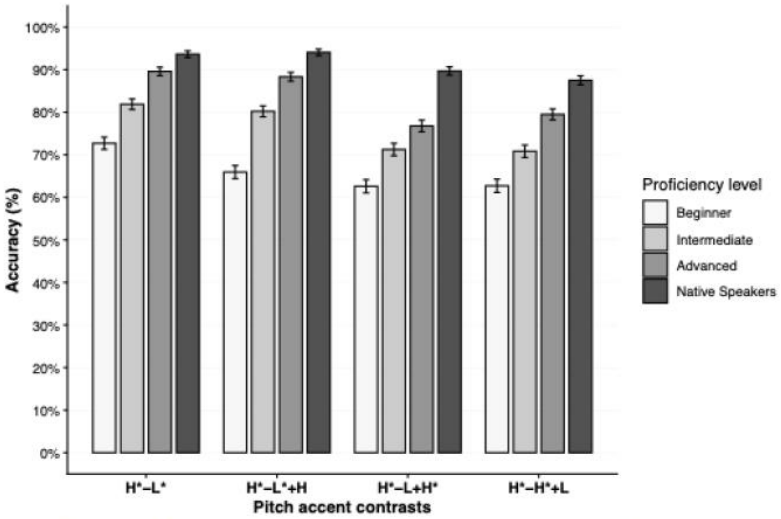


Fig. 2. Accuracy for the four pitch accent contrasts across proficiency groups. Error bars represent standard errors of the mean.

REFERENCES

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology,* 3, 255–309.

Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S.-A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R., & Yoo, H. Y. (2015).

Intonational phonology of French: Developing a ToBI system for French. *Intonation in Romance*, 63–100.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, 106(2), 682-706.

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.

Llanos, F., German, J. S., Gnanateja, G. N., & Chandrasekaran, B. (2021). The neural processing of pitch accents in continuous speech. *Neuropsychologia*, 158, 107883.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation). Massachusetts Institute of Technology.

Schwab, S., Giroud, N., Meyer, M., & Dellwo, V. (2020). Working memory and not acoustic sensitivity is related to stress processing ability in a foreign language: An ERP study. *Journal of Neurolinguistics*, 55, 100897.

Focus Prosody in Shughni Noun Phrases

Sofia Sedunova

(Higher School of Economics, Russia)

Shughni is an Iranian language spoken by ca. 100,000 people in the Pamir Mountains. Notably, a thorough examination of Shughni's intonational patterns is absent from previous studies. According to [1], declarative sentences are marked by a low phrasal tone, while questions are marked by a high boundary tone. On the other hand, [2] notes that declarative sentences have a slowly rising and then immediately falling intonation, while content questions begin with a high tone that falls during the utterance. The tone in yes/no questions (marked by the particle =ɔ) peaks at the verb and then drops at the question marker. However, limited research has been done on Information Structure in Shughni. As stated in [3], Shughni lacks special morphology for marking focus. Instead, focus is marked through a combination of prosodic and syntactic factors. All focused constituents receive sentence-level prosodic prominence, although no studies have provided acoustic data to support this claim.

This study investigates the realization of focus in Shughni noun phrases (NPs) using the Autosegmental-Metrical framework [4]. Data (112 recordings from 18 speakers: 2 male, 16 female; mean age = 41.7), recorded in Khorugh, Tajikistan, in 2024, were collected through a semi-spontaneous experimental approach involving a dialogue game that included noun phrases (e.g., Rūšt vorǰ 'Red horse') to examine how focus is realized when an adjective (non-final position) or a noun (final position) is narrowly focused in an NP, and when they are produced in a broad focus condition. Using Praat [5], all sentences were manually labeled for words, syllables, and vowels, and tones were labeled following the principles of intonational phonology. Statistical analysis was performed in Python.

Our results show that Shughni speakers use both phonetic and phonological markers to differentiate broad focus from narrow focus. Figure 1 shows the pitch contours of NPs with three types of focus. The slope on the adjective did not differ significantly between the three conditions, which can be attributed to the overlapping prenuclear rise that is phonetically similar to the pitch accent typical for narrow focus. The nuclear pitch accent in Shughni is aligned with the right edge of the Intonational Phrase (IP) [6]. Broad focus is typically expressed with an L* nuclear pitch accent on the right edge of the NP, realized as a low plateau. Narrow focus is realized with a falling H+L* nuclear pitch accent, characterized by a fall from a high pitch target on the pretonic syllable, which continues throughout the accented syllable and ends at its offset (Fig. 2). Deaccentuation is typically associated with post-focus, while accentuation is used for both focus and pre-focus. In comparison to the neutrally spoken or defocused counterpart, the focused constituent was more likely to be realized as a separate Intermediate Phrase (ip) in some contexts, marked by a high (H-) phrasal accent.

Considering that prosodic phrasing may influence acoustics, we also conducted statistical analysis of segments in monosyllabic words. Vowel mean duration and intensity are displayed in Figure 3. The duration of the vowel was significantly shorter in defocused words than in narrowly focused ones ($p < .001$). However, intensity was not significantly affected by focus type ($p = .027$).

Overall, the data support the claim that prosodic alignment is an adequate way to describe the prosodic realization of focus in Shughni. These results align with cross-linguistic contexts, where emphasized elements are typically realized with increased articulatory effort, such as greater duration, amplitude, and pitch excursion size [7].

REFERENCES

[1] D. I. Edelman and Sh. P. Yusufbekov, ‘Shugnanskij jazyk [Shughni]’, in *Jazyki mira: Iranskie jazyki III*, Moscow: Indrik, 1999, pp. 225–242.

[2] K. Olson, *Shughni Phonology Statement*. SIL International, 2017.

[3] C. Parker, *A Grammar of the Shughni Language*, Doctoral Dissertation, Montreal: McGill University, 2023.

[4] J. Pierrehumbert, *The phonology and phonetics of English intonation*. PhD dissertation, M. I. T., 1980.

[5] Boersma, Paul & Weenink, David (2025). Praat: doing phonetics by computer [Computer program]. Version 6.4.47, retrieved 7 November 2025 from https://praat.org

[6] S. Sedunova, Comparison of declarative and interrogative intonation in Shughni. *Book of Abstracts PaPE 2025. 6th Phonetics and Phonology in Europe*, 2025.

[7] C. Gussenhoven, *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press, 2004.

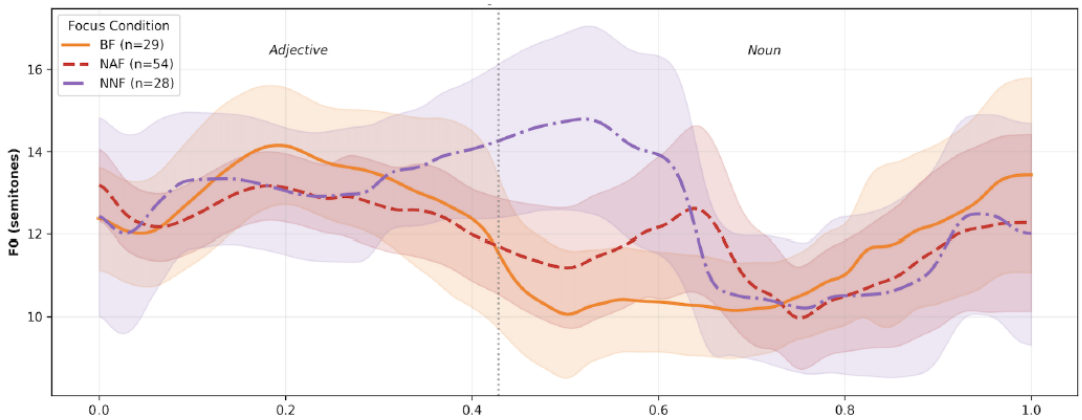


Figure 1: *Three time normalized mean-f0 contours (in semitones) for adjective (NAF), noun (NNF) and NP (BF) focus in the NP; semi-transparent shading represents 95% confidence intervals*

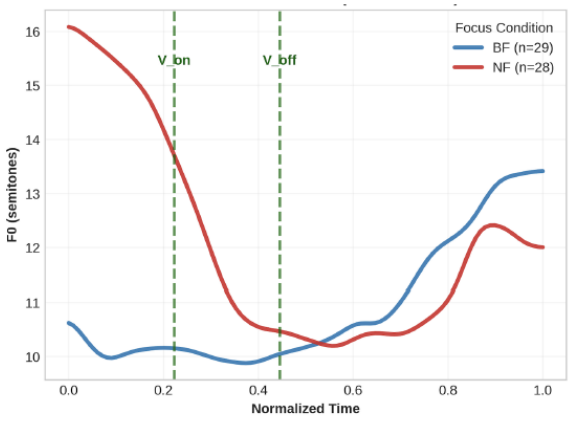
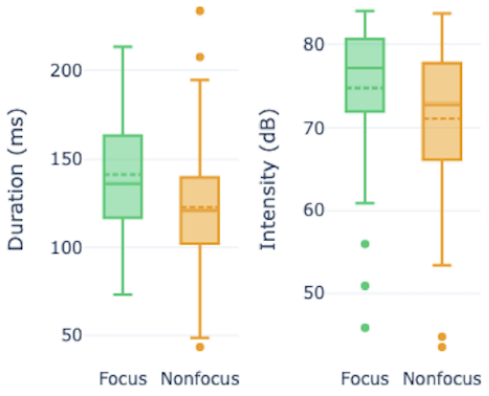


Figure 2: *Time normalized mean-f0 contours (in semitones) for broad (BF) and narrow (NF) focus in nouns; dashed lines mark stressed vowel onset (V_on) and offset (V_off)*

Figure 3: *Boxplots of vowel duration (ms) and intensity (dB) in focused and non-focused monosyllabic words*

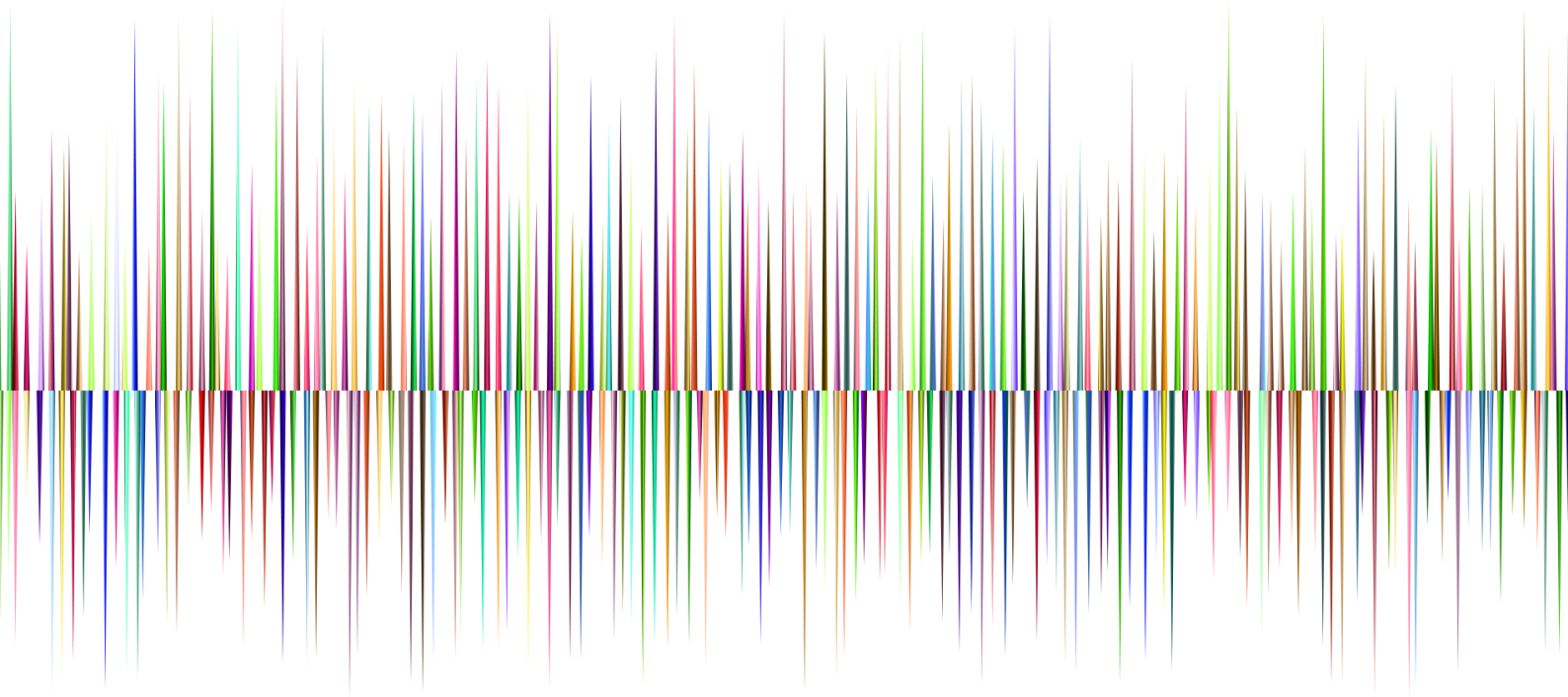


12:10 Parallel Sessions 7 (Rooms 1, 2, 3)

Room 1: **Isabella Reiter, Bettina Braun & Svenja Krieger** (Universität Konstanz, Germany) *Segmental and Suprasegmental Effects on L1 Italian Coda Productions in L2 English*

Room 2: **Marion Coadou-Toscano** (Aix-Marseille Université, France) Assessing Voice Quality Variations: Evolution in Time and New Perspectives

Room 3: **Farhat Jabeen & Jeremy Steffman** (Universität Bielefeld, Germany / University of Edinburgh, United Kingdom) *Toning down of voiced aspirates: A Bayesian analysis of connected speech in Punjabi*



Segmental and Suprasegmental Effects on L1 Italian Coda Productions in L2 English

Isabella Reiter, Bettina Braun & Svenja Krieger

(Universität Konstanz, Germany / University of Vienna, Austria)

We investigate the realization of English monosyllabic words by advanced Italian learners of English. While the English syllable coda can be complex and allows an occupied coda [1], Italian generally prefers open syllables, and does not allow obstruents in the syllable coda word-finally [2]-[4]. This difference poses challenges in second language (L2) acquisition and may generally lead to repair strategies during the learning process (e.g. consonant deletion or vowel epenthesis [5]). Prior research has shown that repairs are driven by the identity of the coda segment itself [6]. Recent findings revealed effects of intonation in English loanwords in Bari Italian: English names more often had epenthetic vowels in rising vs. in falling contours [7]. This was interpreted as "text-adjustment" to create morae to associate the rising tone [7]. In this production study, we test the segmental and intonational effects on the presence of vowel epenthesis in L2 acquisition using two intonation conditions (final falls/rises). The L2 setting may activate the L2 phonology more strongly than the loanwords tested in [7].

We tested 20 Italian learners of English from different regions (Northern, Central and Southern Italy) that had acquired English as an L2 after the age of 6 years. We chose 20 high-frequent target words that are expected to be known by learners, 9 with stops (5 voiced, 4 voiceless), 11 with fricatives (6 voiced, 5 voiceless). Ten further words containing open syllables or sonorant codas were added as fillers. These words were inserted into two different sentence types to elicit different nuclear contours: at the end of declarative statements (e.g. I am reading a …) to elicit a falling contour, and at the end of polar questions (e.g. Are you reading my new …?) to elicit a rise [8]. All factors were manipulated within-subjects, and intonation also within-items. Participants were tested online and were recorded using the audio software audacity via its loopback function (sample rate: 44100 Hz, 16Bit). The 800 audio recordings were manually annotated regarding boundary tones (rise, fall, plateau) and repair strategy, if applicable (see Table 1). Results show that this manipulation worked: Participants produced 50% falls (typically in declaratives) and 48% rises (typically in polar questions) and 2% plateau. Plateaus were not considered in the subsequent analyses. The most frequent repair strategies were vowel epenthesis, deletion, and devoicing, see Table 1. Other productions (e.g. the realization of /θ/ as [t]) were categorized as 'other'.

The occurrence of vowel epenthesis (yes/no), the repair strategy of interest, was analyzed using a mixed-effects logistic regression model with MANNER OF ARTICULATION, VOICING and INTONATION as fixed effects, and PARTICIPANTS and ITEMS as crossed random effects [9]. Random slopes were included if the model converged. Results showed a significant effect of INTONATION and an interaction between MANNER OF ARTICULATION and VOICING (all $p < 0.05$, see Fig. 1 for model predictions). Vowel epenthesis was more frequent in rises than falls ($ß = 0.9$, $SE = 0.3$, $p < 0.05$). There were few epenthetic vowels in fricatives with no effect of voicing ($p > 0.2$) and more in stops, particularly when voiced ($p < 0.01$).

The data clearly show that segmental and suprasegmental factors affect the frequency of vowel epenthesis in L2 speech to avoid syllables with a coda obstruent, particularly a voiced one. The data replicate previous results on segmental effects [6]. The effect of intonation was significant, but weaker than in [7]. This may be due to the different language modes, resulting in different degrees of language activation: [7] tested loanwords in the speakers' L1 Italian,

resulting in more instances of vowel epenthesis. Instead, we tested Italian learners of English in their L2, which presumably causes a higher degree of activation of their L2, resulting in a lower number of vowel epenthesis. We propose a straightforward way to model vowel epenthesis in learners using Optimality Theory [10].

intonation	manner	voicing	correct	epenthesis	deletion	devoicing	other
fall	fricative	unvoiced	97	**6**	0	0	3
fall	fricative	voiced	38	**6**	5	68	1
fall	plosive	unvoiced	70	**2**	1	0	4
fall	plosive	voiced	55	**17**	2	13	10
rise	fricative	unvoiced	65	**10**	3	0	4
rise	fricative	voiced	29	**12**	0	61	0
rise	plosive	unvoiced	78	**12**	5	0	3
rise	plosive	voiced	41	**29**	2	14	11

Table. 1: Frequency of realizations per segmental environment

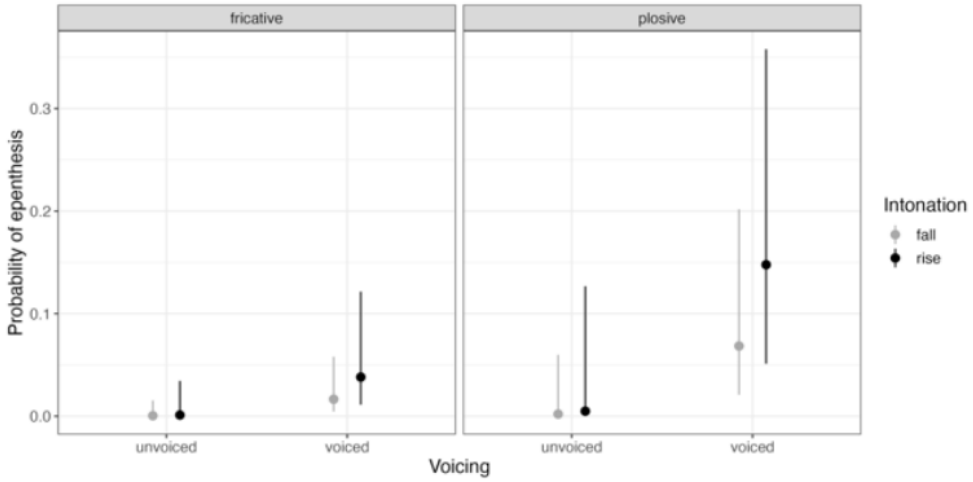


Fig. 1: Predicted probability of epenthesis per voicing, manner of articulation and intonational contour

REFERENCES

[1] Lorenz, Frank. (2025). *English Phonetics and Phonology: An Introduction*. Cambridge: Cambridge University Press.

[2] Costamagna, Lidia. (2010). L'apprendimento della fonologia dell'italiano da parte di studenti sinofoni: criticità e strategie [The acquisition of Italian Phonology by Chinese Students: Critical Issues and Strategies]. *Associazione Italiana Centri Linguistici Universitari (AICLU)*, Roma, 19, 49-67.

[3] Krämer, Martin. (2009). *The phonology of Italian*. Oxford: Oxford University Press.

[4] Marotta, Giovanna. (1995). Coda condition in Italian and underspecification theory. In Kjell Elenius & Peter Brandeurd (eds.) *XIIIth Int. Congr. Phon. Sciences (ICPhS)*, Stockholm, 95. 3(1), 378-381.

[5] Espinosa, Juan A. C. (2001). Sonority and Constraint Interaction: the Acquisition of Complex Onsets by Spanish Learners of English. *Anglogermanica online: Revista electrónica periódica de filología alemana e inglesa*, (1), 1-20.

[6] Grice, Martine, Michelina Savino, Alessandro Caffò & Timo B. Roettger. (2015). The tune drives the text: Schwa in consonant-final loan words in Italian. In *International Congress of Phonetic Sciences (ICPhS)*, Glasgow.

[7] Grice, Martine, Michelina Savino & Timo B. Roettger. (2018). Word final schwa is driven by intonation – The case of Bari Italian. *The Journal of the Acoustical Society of America*, 143(4), 2474-2486. https://doi.org/10.1121/1.5030923

[8] Pierrehumbert, Janet B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology).

[9] Bates, M. Douglas & Deepayan Sarkar. (2007). *lme4: Linear Mixed-Effects Models Using S4 Classes*. R Package Version 0.99875-6 ed

[10] Prince, Alan S. & Paul Smolensky. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, N.J., and Computer Science Department, University of Colorado, Boulder.

Assessing Voice Quality Variations: Evolution in Time and New Perspectives

Marion Coadou-Toscano

(Aix-Marseille Université, France)

Twenty years ago, studies on voice quality (VQ) were scarce, and even more so regarding accent variations. Indeed, while rhythm and intonation variations across some accents of the British Isles had already been described by numerous studies, it seemed that another component was necessary to describe the suprasegmental features of an accent. John Laver's fundamental work [1] was one of the first to define VQ in terms of supralaryngeal and laryngeal settings. Twenty years later, [1] still represents a landmark in the taxonomy of VQ description. While more recent models, such as John Esling's laryngeal articulator model [2], have since refined the physiological understanding of these mechanisms, the present study remains grounded in Laver's framework as it provides a thorough and sufficiently discriminative basis for comparing accent variation. Moreover, the debate on whether to include articulatory settings along with phonatory settings appears to be over, as this definition of VQ is often used in more recent studies, e.g. [3] or [4]. On the whole VQ has increasingly been studied by phoneticians in the past few years, as [5] shows.

For my research [6], it was necessary to find a way to assess the speakers of the IViE corpus [7]) in order to see if their voice quality varied across five accents of the United Kingdom, namely Belfast, Newcastle, Liverpool, Cardiff, and Cambridge. The Vocal Profile Analysis (VPA) [8] was originally aimed at assessing pathological voices, but its design also allowed for the measurement of non-pathological voices with accurate scalar degrees. However, as this protocol provided raw data, a statistical analysis was carried out to compare the results across the accents of the corpus. Thus, the discriminant analysis showed that the Belfast accent was the only one to truly stand out compared to the others, particularly regarding the male speakers. Not unexpectedly, not only does this protocol remain a useful tool, but it also inspired E. San Segundo and J.I. Mompean to create a simplified version of it [4].

Regarding objective measurement, several methods were available at the time, such as Inverse Filtering [9] or electroglottography, as seen in [10]. For technical reasons, we chose to analyze the Long-Term Average Spectrum (LTAS) [11]. The results of the Principal Component Analysis (PCA) of the LTAS also made it possible to isolate the Belfast speakers from the rest of the corpus. However, this type of acoustic analysis now seems to be less common. A primary reason for this may be that the LTAS is highly sensitive to external factors, ranging from recording conditions to idiosyncratic characteristics such as vocal tract anatomy. This implies that the differences observed between the Belfast accent and the others could be attributed to factors other than VQ. Voicesauce [12] appears to be a promising tool for measuring voice quality variations across accents of the British Isles, a task which, to our knowledge, has not yet been undertaken.

The use of VQ as a tool for teaching English as a Foreign Language (EFL) remains an area of research yet to be fully explored. Again, studies are scarce, although the primary objective of Honikman's ground-breaking work [13] was to improve the pronunciation of her EFL students. This was the basis for my interest in comparing the voice quality of French students to speakers from the IViE corpus. The results of the VPA analysis in [14] showed that the French speakers exhibited a higher degree of lip rounding, as [13] and [15] suggested; however, they also demonstrated a less nasalized voice and a lower degree of tension in the

vocal tract and larynx than the Cambridge speakers. To corroborate these findings, it would be useful to apply the VPA to a larger corpus, such as ANGLISH [16].

Ultimately, despite its fundamental importance, VQ continues to play a marginal role in EFL teaching. The reasons for this are manifold, ranging from a lack of training for teachers in phonology—and even more so in VQ—to the limited time dedicated to second language acquisition, particularly in France. These findings suggest that further investigation is needed into how VQ could be taught to trainee teachers and integrated into modern technology-based teaching tools.

REFERENCES

[1] Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

[2] Esling, J. H., Moisik, S. R., Benner, A. B., & Crevier-Buchman, L. (2019). *Voice Quality, the Laryngeal Articulator Model, and Phonetic Usages*. Cambridge: Cambridge University Press.

[3] Wilhelm, S. (2019). "Voice Quality in British English. Its Nature, Functions and Applications". *Anglophonia, French Journal of English Studies*, 27.

[4] San Segundo, E., & Mompean, J. A. (2017). "A Simplified Vocal Profile Analysis Protocol for the Assessment of Voice Quality and Speaker Similarity". *Journal of Voice*, 31(5), 644.e11-644.e27.

[5] Herment, S., & Fournier, P. (2019). "Voice Quality in English: an Introduction". *Anglophonia*, 27.

[6] Coadou, M. (2007). *Qualité de voix et accents régionaux en anglais britannique*. Unpublished PhD thesis. Aix-Marseille University.

[7] Grbe, E., Low, L., & Nolan, F. (1998). *English intonation in the British Isles, the IViE corpus*. University of Oxford. www.phon.ox.ac.uk/IViE.

[8] Laver, J., Wirz, S., Mackenzie, J., & Hiller, S. M. (1991). "A perceptual protocol for the analysis of vocal profiles". In Laver, J. (ed.): *The gift of speech: readings in the analysis of speech and voice*. Edinburgh: Edinburgh University Press, 265-280.

[9] Gobl, C., & Ní Chasaide, A. (2003). "The role of the voice quality in communicating emotions, mood and attitude". *Speech Communication*, 40, 189–212.

[10] Rossato, S., Audibert, N., & Aubergé, V. (2004). "Emotional voice measurement: a comparison of articulatory-EGG and acoustic-amplitude parameters". *Proceedings of Speech Prosody*, Nara, 91-95.

[11] Coadou, M., & Rougab, A. (2007). "Voice Quality and Variation in English". *ICPhS XVI*, Saarbrücken, 6-10 August 2007.

[12] Shue, Y. L., Keating, P., Vicenik, C., & Yu, K. (2011). "VoiceSauce: A program for voice analysis". *ICPhS XVII*, Hong Kong, 17-21 August 2011.

[13] Honikman, B. (1964). "Articulatory settings". In Abercrombie, D. et al. (eds): *In honour of Daniel Jones*. London: Longman, 73-84.

[14] Coadou, M., & Audibert, N. (2009). "Voice Quality and English as a Foreign Language: A Pilot Study". *Proceedings of the 3rd International Workshop on Advanced Voice Functions Assessment (AVFA 2009)*. Madrid, Spain.

[15] Esling, J. H., & Wong, R. F. (1983). "Voice Quality Settings and the Teaching of Pronunciation". *TESOL Quarterly*, 17(1), 89-95.

[16] Tortel, A. (2008). "ANGLISH : base de données comparatives L1 & L2 de l'anglais lu, répété et parlé". *Travaux interdisciplinaires du Laboratoire Parole et Langage*, 27, 111-122.

Toning down of voiced aspirates: A Bayesian analysis of connected speech in Punjabi

Farhat Jabeen & Jeremy Steffman

(Universität Bielefeld, Germany / University of Edinburgh, United Kingdom)

This study constitutes the first experimental investigation of lexical tones in Pakistani Punjabi and their relation to word initial obstruents. Previous research on Indian Punjabi offers conflicting accounts of the language's tonal inventory. Some studies propose a two-tone system (Mann et al., 1961), while others argue for a three-tone contrast (Gill and Gleason, 2013; Hussain et al., 2019). Nevertheless, there is broad consensus that low tones in Punjabi are linked to the historical loss of voiced aspiration (Gill and Gleason, 2013; Hussain et al., 2019). If tonal contrasts in Punjabi emerged solely from the neutralisation of voiced aspiration, one would predict a binary tonal system: Low tone for the lost voiced aspiration versus absence of tone. However, as most studies report a three-way tonal contrast, this indicates a more complex relationship between tonal realisation and consonantal features in Punjabi. Based on this, we investigate 1) the relation between word-initial obstruents and F0 contour in Pakistani Punjabi and 2) the influence of syllable count on tonal realisation.

We extracted nouns from a corpus of narrative speech (27.3 minutes) produced by 10 speakers of Pakistani Punjabi (6 male, 4 female). The narratives were divided into inter pausal units separated by at least 150ms. To minimize the role of prosodic phrasing, words preceding Intonational Phrase boundaries were excluded. The target words were monosyllabic (n=246) or bisyllabic (n=227) and began with an obstruent (bilabial, coronal, or velar). We measured F0 at 20 time-normalized samples across each word and used Functional Principal Component Analysis (FPCA) to extract PC scores (e.g., Arvaniti et al., 2024). We used Bayesian Mixed effects linear regression in a multivariate implementation.

Panel A of Figure 1 illustrates the mean F0 contour for each obstruent class. Monosyllables: F0 onset remains stable across obstruent types. The last quarter of the F0 trajectory for voiced aspirated obstruents (DH) is distinct from other classes. Bisyllables: Voiced unaspirated obstruents (D) exhibit a steep rising contour. Notably, this contour is produced only in this context. Panel B shows that PC1, accounting for 84% of the variance, corresponds broadly to F0 height, with lower scores corresponding to higher trajectories. PC2 accounted for 14% of variance and corresponds to slope, with lower scores representing falling trajectories.

Given credible main effects and interactions for syllable count and obstruents, we computed marginal means and contrasts for each obstruent class and syllable count (panel C) interaction. In Table 1, we report "probability of direction" (pd) denoting that a difference exists with a particular directionality (Makowski et al., 2019). We found no differences in PC1 scores based on obstruent class or syllable count. For PC2, DH exhibited falling F0 contour in monosyllables, contrasting with voiced unaspirated (D) and voiceless unaspirated (T) obstruents (probable effects). In bisyllabic words, D was produced with a rising contour, while DH carried flat F0 (certain effect). These results indicate the lack of a straight forward relation between tones and the lost voiced aspiration in Pakistani Punjabi. While the trajectories of T and TH remain stable, the F0 contour of D and DH is modulated by syllable count. Notably, the interplay between obstruent class and syllable count is limited, explaining only 14% of variance in the data (PC2).

To summarise, we found no default association between low tone and historically voiced aspirated obstruents in Pakistani Punjabi. Aspirated obstruents, regardless of voicing, patterned together in both the PCs. This suggests that Punjabi tones may be a byproduct of laryngeal features rather than being derived from the loss of voiced aspiration. Previous analyses of Punjabi tones have primarily examined words produced in isolation or in carrier phrases. This study is the first to analyse Punjabi tones in fluent speech. Consequently, our findings may better capture tonal representations in natural, connected Punjabi. Alternatively, these results may reflect dialectal differences between Indian and Pakistani Punjabi. In future, we plan to systematically investigate the influence of vowel quality and lexical stress on tones in Punjabi.

	Monosyllabic		**Bisyllabic**	
Contrasts	PC1	PC2	PC1	PC2
DH - D	89	**99**	70	**100**
DH - TH	62	74	70	60
DH - T	85	**97**	91	62

Table 1: Probability of direction for pairwise comparisons. pd < 95% = Uncertain effect, 95% = possible effect, 97% = Likely effect, 99% = Probable effect, 99.9% = Certain effect.

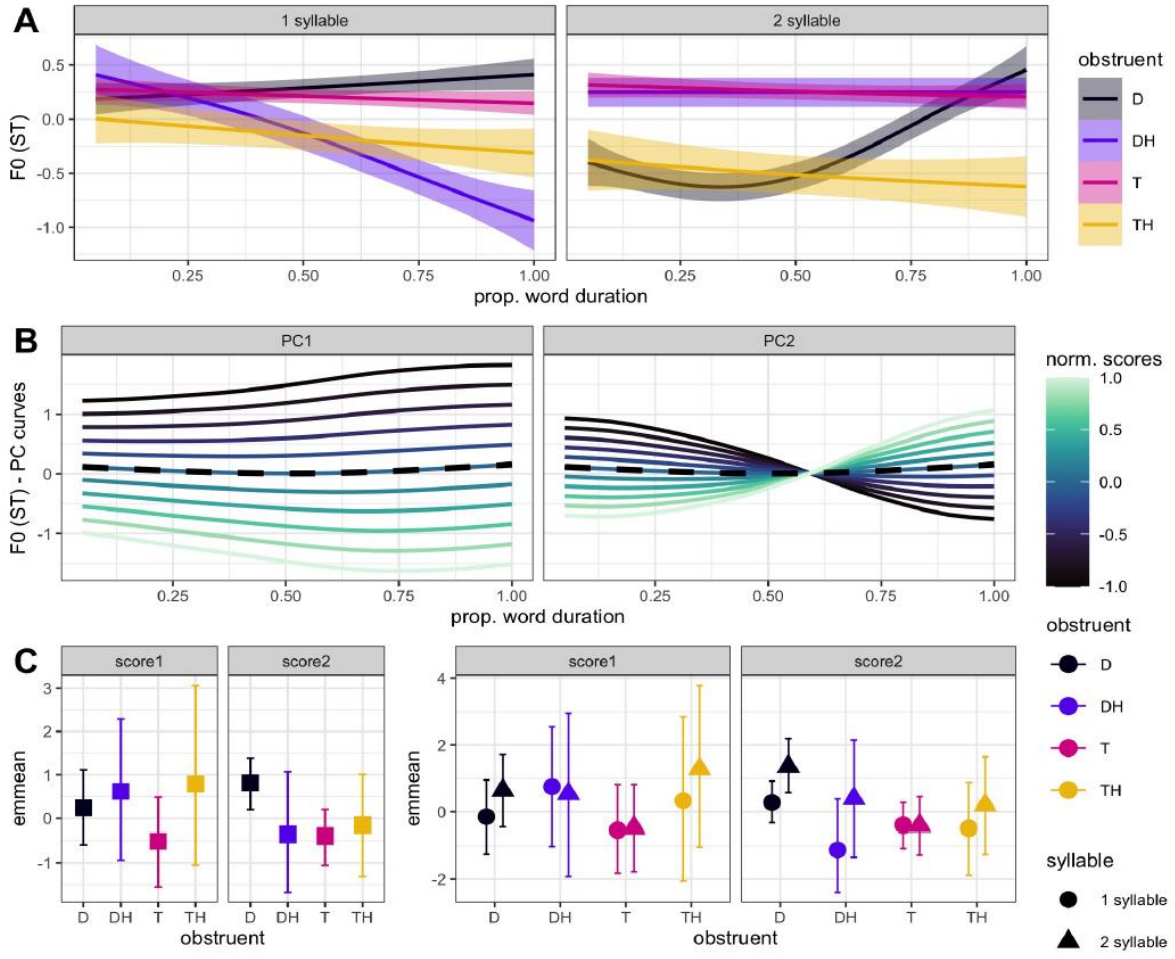


Figure 1: A: Mean F0 and 95% confidence intervals (fit with GAM). B: PC curves. Dashed line denotes mean F0 curve. C: PC scores based on obstruent (left) and obstruent and syllable (right). Points denote estimated median and whiskers show 95% credible intervals. D = voiced, T = voiceless, H = aspiration.

REFERENCES

Arvaniti, Amalia, Katsika, Argyro and Hu, Na. 2024. Variability, overlap, and cue trading in intonation. *Language* 100(2), 265–307.

Gill, Harjeet Singh and Gleason, Henry A. 2013. *A reference grammar of Punjabi*. Punjabi University, Patiala.

Hussain, Qandeel, Proctor, Michael, Harvey, Mark and Demuthi, Katherine. 2019. Punjabi (Lyallpuri variety). *Journal of International Phonetic Association* 50(2), 282–297.

Makowski, Dominique, Ben-Shachar, Mattan S. and Lüdecke, Daniel. 2019. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software* 4(40), 1541.

Mann, Gurinder Singh, Singh, Gurdit, Shah, Ami P., Scheffler, Gibb and Murphy, Anne. 1961. *An introduction to Punjabi: Grammar, conversation and literature*. Punjabi University, Patiala.

13:00 Lunch break

14:00 Poster Session

Ai Chen, Golshan Shakeebaee, Markus Bader & Frank Kügler (Goethe-Universität Frankfurt am Main, Germany) *Prosody of negation in different focus structures in German*

Maciej Karpiński & Bettina Braun (Adam Mickiewicz University, Poznań, Poland / Universität Konstanz, Germany) Exploring the effects of mutual visibility on the coordination of pitch and hand movements in task-oriented dialogues

Caterina Petrone, Ivan Ventocilla Loayza, Carine André, Christelle Zielinski, Cristel Portes & Roxane Bertrand *Prosody in other-repetitions fosters change in conversational trajectory*

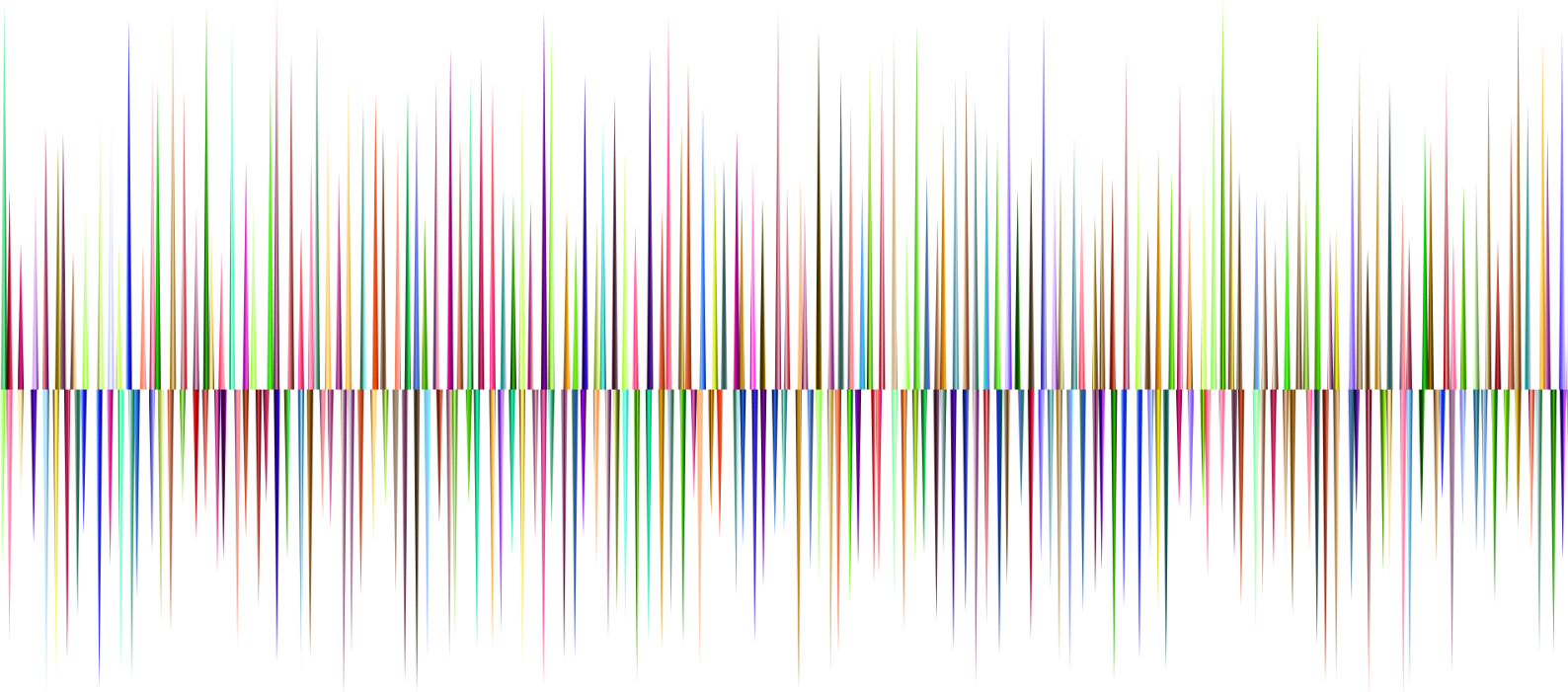
Artem Saloev & Nicolas Ballier (Université Paris Cité, France) *A Gender-Based Analysis of Mimi Prosodic Representations*

Sofia Sedunova & Yury Makarov (Higher School of Economics, Russia / University of Cambridge, UK / Vinogradov Russian Language Institute, RAS, Russia) *A first look at lexical stress in Shughni*

Šárka Šimáčková (Palacky University Olomouc, Czech Republic) Gesture and intonation in trainee teachers' L2 English: Alignment with prominence and responses to manual constraints

Yao Vera Yujia (University of Glasgow, United Kingdom) *The Realisation of Narrow Focus in Glaswegian English*

Ivan Yuen, Bistra Andreeva, Bernd Moebius & Mitko Sabev (Saarland University, Germany) How do semantic likelihood and information structure affect prosodic encoding in different tasks?



Prosody of negation in different focus structures in German

Ai Chen, Golshan Shakeebaee, Markus Bader & Frank Kügler

(Goethe-Universität Frankfurt am Main, Germany)

This study investigates the prosodic effects of negation in German. While previous research suggests that negation tends to carry high prominence, this claim remains a subject of debate (Yaeger-Dror, 1985, 1997). In Swedish, specific F_0 patterns have been observed more frequently in negative feedback compared to positive feedback (Tronnier & Allwood, 2004). Similarly, research in German has shown that negatively connoted statements exhibit a reduced global pitch range relative to positive ones (Reckling & Kügler, 2011), while studies on Catalan and English have identified specific intonational contours associated with negation (Hedberg & Sosa, 2003; Espinal & Prieto, 2011; Espinal et al., 2016).

Building on these findings, the present work examines whether the negation particle *nicht* ('not') influences overall pitch range and whether prosodic markers signal a forthcoming negation. We conducted three production experiments using context-based elicitation to trigger three distinct focus structures (focus on the noun phrase, the noun, or the adjective). Target sentences contained either the negation particle *nicht* or the affirmative contrastive particle *doch* ('after all'). Example is the following:

Context:

Marie war heimlich im Zimmer ihrer Schwester und hat dort (*Marie was secretly in her sister's room and*)

- **Exp 1:** ...eine gelbe Birne entdeckt. (*a yellow pear*)
- **Exp 2:** ...eine große Schale... (*a large bowl...*)
- **Exp 3:** ...einen Teller mit drei Birnen... (*a plate with three pears...*)
- **Q-Neg:** Was macht Marie...? (*What will Marie do...?*)
- **Q-Aff:** Marie wollte fasten... Was macht Marie...? (*Marie wanted to fast... What will she do...?*)

Target: Marie wird die gelbe Birne ihrer Schwester **nicht** [Neg] / **doch** [Aff] essen.

(*Marie will [not / after all] eat the yellow pear*.)

Participants ($N=15$) were presented with these contexts and read aloud 24 target sentences per experiment. Sentences were controlled for syllable count, with *nicht* or *doch* appearing in sentence-medial position. F_0 trajectories were analyzed using Generalized Additive Mixed Models (GAMMs), and pre-target word durations were compared using linear mixed models. Additionally, we analyzed the data from a perceptual perspective using the DIMA annotation framework (Kügler et al., 2022).

The results indicate an absence of prosodic markers anticipating *nicht* when compared to *doch*. Furthermore, no significant differences in F_0 values were found between the two particles, suggesting that *nicht* is prosodically similar to the inherently contrastive *doch*. Word durations likewise showed no significant variation between affirmative and negative sentences. While DIMA annotations revealed that tonal patterns for negative and affirmative sentences did not always align, no consistent tonal cue emerged as a predictor for the upcoming particle.

Finally, a comparison of prosodic contours across experiments confirmed that participants produced distinct realizations conditioned by the preceding focus structure.

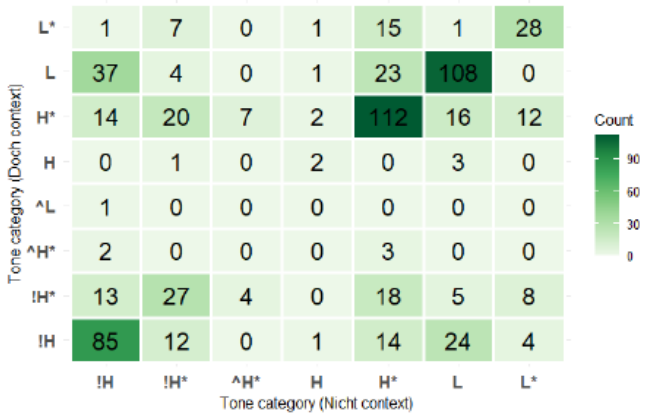


Figure 1: Comparison of tones

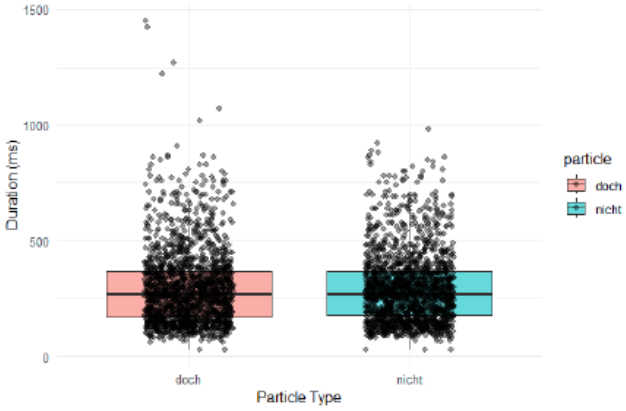


Figure 2: Duration of words

REFERENCES

Espinal, M. T., & Prieto, P. (2011). Intonational encoding of double negation in Catalan. *Journal of Pragmatics*, 43(9), 2392–2410.

Espinal, M. T., et al. (2016). Double Negation in Catalan and Spanish: Interaction Between Syntax and Prosody. In P. Larrivée & C. Lee (Eds.), *Negation and Polarity: Experimental Perspectives* (Language, Cognition, and Mind, 1), pp. 145–176. Cham: Springer International Publishing.

Hedberg, N., & Sosa, J. M. (2003). Pitch Contours in Negative Sentences. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, Barcelona, pp. 627–630.

Kügler, F., Baumann, S., & Röhr, C. T. (2022). Deutsche Intonation, Modellierung und Annotation (DIMA). *Transkription und Annotation von Gesprächen und Sprache: Multimodaler Interaktion. Konzepte, Probleme, Lösungen*, p. 23.

Reckling, F., & Kügler, F. (2011). Pitch range in positive and negative connoted statements of German. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, pp. 1670–1673.

Tronnier, M., & Allwood, J. (2004). Fundamental frequency in feedback words in Swedish. *18th International Congress on Acoustics*, pp. 2239–2242.

Yaeger-Dror, M. (1985). Intonational prominence on negatives in English. *Language and Speech*, 28(3), 197–230.

Yaeger-Dror, M. (1997). Contraction of negatives as evidence of variance in register-specific interactive rules. *Language Variation and Change*, 9(1), 1–36.

Exploring the effects of mutual visibility on the coordination of pitch and hand movements in task-oriented dialogues

Maciej Karpiński & Bettina Braun

(Adam Mickiewicz University, Poznań, Poland / Universität Konstanz, Germany)

Mutual visibility has been shown to modulate several aspects of communicative behaviour, including prosody [1] and gesticulation [2]. It remains unclear, however, whether both channels are equally affected. Speech prosody and co-speech gesture are considered to be closely interconnected [e.g., 3,4,5,6], yet little is known about how their interaction depends on the visual availability of the interlocutor. We investigated the effect of mutual visibility (yes vs. no) on the coupling between pitch and hand movements in task-oriented dialogue.

We explored the correlation between hand kinematics (position, speed and acceleration) and mean pitch. To that end, we used recordings of 20 pairs of speakers from the MultiCo corpus [8]. Native speakers of Polish performed a Tower task in pairs, facing each other at a distance of ca. 3m. The dialogue task was to build a tower together using imaginary blocks, which promoted the use of illustrators. Speech was recorded using head-on microphones. Motion capture data were collected using the iPi system, based on two depth-of-field cameras per speaker. In the no-visibility condition, an acoustically transparent blind was placed symmetrically between the speakers.

Pitch frequency was extracted in Praat [9] using filtered autocorrelation; coordinates of speakers' index fingers and heads were extracted; all were smoothed. Based on spatial coordinates and frame-rate information, uni- and 3-dimensional velocity and acceleration were calculated and log-normalised to yield normal distributions. To synchronise, f0 and motion data were averaged in 100 ms time bins. All measures were z-scored to account for biological speaker differences, and data with absolute z-scores exceeding 3 were excluded as outliers (4.7% of the data).

Principal component analysis of the motion data showed that 3-dimensional speed and acceleration loaded equally strongly on the first dimension (explaining more than 76% of the variance), the y- and z-coordinates on the second dimension (16% of the variance), and the x-coordinates of the right hand and the head height on the third one. We next calculated a linear mixed-effects regression model with these uncorrelated factors for the right hand: speed and accuracy, hand y and z-coordinates, and head height, all interacting with visibility. Speaker and IPU number were added as random effects. Non-significant interactions and main effects were eliminated (α=0.05).

Results showed significant interactions between visibility and the hand's height (y), frontness (z), and the head's height, see Fig. 1 (all $p<0.0005$). For the no-visibility condition, hand height was not correlated with mean pitch, but for the visibility condition, higher hand positions correlated with lower pitch. Further, in the visibility condition, extending the hand towards the interlocutor correlated with increased pitch, and vice versa in the no-visibility condition. Head height positively correlated with mean pitch, more strongly in the visibility condition.

Our results indicate that, even in the task that encourages the use of less interactive gestures for shape description, mean pitch is systematically related to hand movement. Moreover, this relation is modulated by mutual visibility. Although numerically small, the effects are statistically significant and point to an intricate, dynamic mechanism. The coupling

between pitch and hand movements is therefore not mechanistic but sensitive to interactional factors. In future work, we plan to expand our dataset, incorporate timing- and entrainment-related analyses, and examine representational and expressive aspects of pitch and body movement in relation to linguistic events.

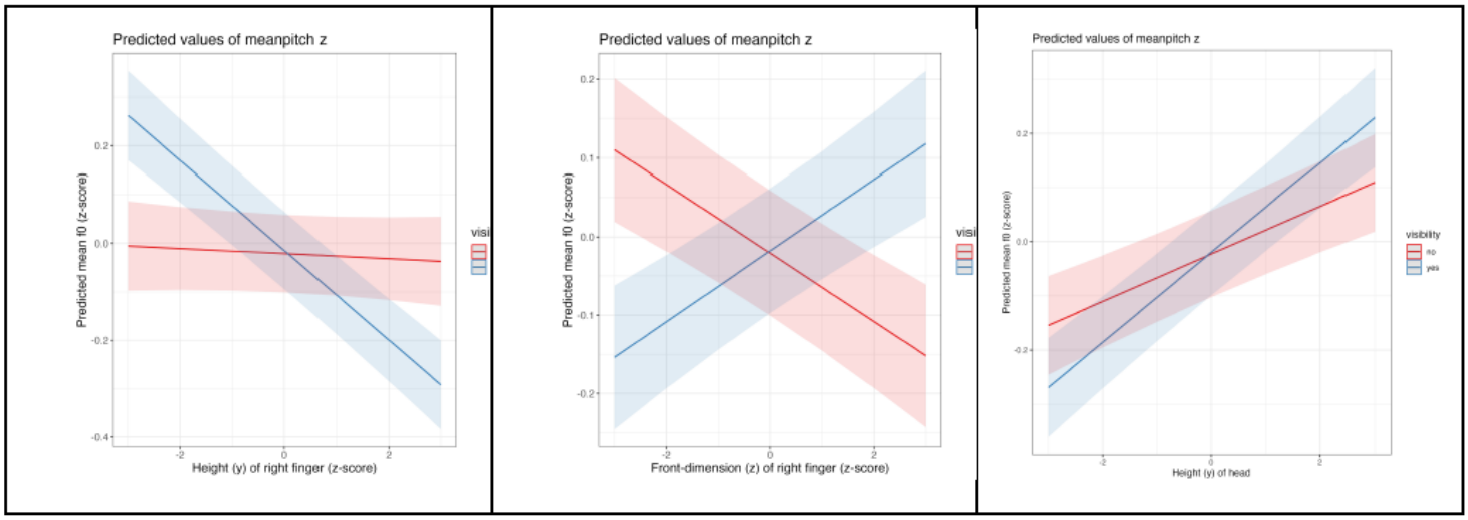


Figure 1: Interactions between visibility and hand and head position.

REFERENCES

[1] Wagner, P., Bryhadyr, N., Schröer, M. (2019) Pitch Accent Trajectories Across Different Conditions of Visibility and Information Structure - Evidence from Spontaneous Dyadic Interaction. *Proc. Interspeech 2019*, 3985-3989, https://doi.org/10.21437/Interspeech.2019-1619

[2] Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, 13 (1), 63–92. https://doi.org/10.1075/gest.13.1.03bav

[3] Bolinger, D. (1983). Intonation and gesture. *American Speech*, 58 (2), 156–174. https://doi.org/10.2307/455326

[4] McClave, E. Z. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research*, 27 (1), 69–89. https://doi.org/10.1023/A:1023274823974

[5] Gibbon, D. (2011). Modelling gesture as speech: A linguistic approach. *Poznań Studies in Contemporary Linguistics*, 47 (3), 470–487. https://doi.org/10.2478/psicl-2011-0026

[6] Brown, L., & Prieto, P. (2021). Gesture and prosody in multimodal communication. In S. Cacchiani & J. Norrick (Eds.), *The Cambridge handbook of sociopragmatics* (pp. 430–453). Cambridge University Press.

[7] Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3 (1), 71–89. https://doi.org/10.1515/lp-2012-0006

[8] Karpiński, M., Klessa, K., Jarmołowicz-Nowikow, E., Taborek, J., Sawicka-Stępińska, B., & Piosik, M. (2023). MultiCo [Data set]. *CLARIN-PL Digital Repository*. http://hdl.handle.net/11321/942

[9] Boersma, Paul & Weenink, David (2026). *Praat: doing phonetics by computer* [Computer program]. Retrieved 10 November 2025 from https://praat.org

Prosody in other-repetitions fosters change in conversational trajectory

Caterina Petrone, Ivan Ventocilla Loayza, Carine André, Christelle Zielinski, Cristel Portes & Roxane Bertrand

(Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France)

In French, declarative questions with a rising intonation contour mostly convey a neutral yes-no question or a positively biased confirmation request (Beyssade and Marandin 2006). However, when they are modified by phonetic markers of disbelief, they rather express the speaker's incredulity towards the conveyed proposition. In this paper, we investigate whether, in French, prosodic cues of disbelief may affect the direction of the conversation among two interlocutors (or conversational trajectory), by focusing on the prosody of the other repetition, i.e., the repetition of a content just proposed by the interlocutor.

Other repetition (OR) is a conversational practice mostly considered to foster alignment with the current conversational activity and affiliation with the preceding speaker's stance (Stivers 2008). However, a few authors showed that some uses of OR can, on the contrary, trigger some forms of disalignment and/or disaffiliation (Rossi 2020, Persson 2020). Since rising declaratives are known to mostly convey a neutral yes-no question or a positively biased confirmation request (Beyssade and Marandin 2006), they can be used with OR to indicate that the speaker temporarily pauses his alignment until confirmation of the issue at hand. In this case, the preferred conversational trajectory of the addressee should be to confirm his initial claim. Conversely, when phonetic marks of incredulity are added to the rising OR, the speaker rather conveys disaffiliation, and his addressee may sometimes use a backdown trajectory to restore affiliation.

In the present study, we designed an online perception experiment to test these hypotheses. Participants had to listen to an original turn, followed by an OR with a confirmation request prosody versus an incredulity prosody. They then had to choose between responses of the initial speakers indicating his preferred conversational trajectory: "Voilà, c'est ça." *Yes, that's it.* for the confirmation trajectory versus "Enfin, non…" *Well, no...* for the backdown trajectory. We hypothesized that the confirmation prosody on the OR will foster the confirmation trajectory while the incredulity prosody will foster the backdown trajectory. We measured the number of chosen response ("Voilà, c'est ça." versus "Enfin, non…") and reaction times depending on the prosody of the OR (confirmation prosody versus incredulity prosody).

One hundred and twenty native French speakers (sixty men, sixty women) participated in the experiment online through Labvanced (Finger et al. 2017). Compared to the confirmation prosody, incredulity prosody was characterized by significantly longer onset and rhyme duration for the third syllable ([lo] in Figure 1), higher H and later L2 alignment to that same syllable's onset.

The results of a generalized linear mixed model showed that:

1. regardless of prosody, listeners prefer to respond with "Voilà, c'est ça." rather than "Enfin, non…" ($z=3.16$, $p=0.0015$),
2. confirmatory prosody elicits more "Voilà, c'est ça." responses than incredulous prosody ($z=4.44$, $p<0.001$),
3. the effect of gender is not significant (marginal p: $p=0.07$).

Concerning reaction times:

1. listeners select “Voilà, c’t ça.” more quickly after a prosodic pattern of confirmation than after one of incredulity (t=8.09, p<0.001);
2. listeners select “Voilà, c’est ça.” more quickly after a prosody of confirmation than “Enfin, non…” after a prosody of incredulity (t=5.38, p<0.001),
3. there is no difference in RT between prosody of disbelief and prosody of confirmation when listeners choose “Enfin non…” (p=0.61).

We conclude that French listeners are sensitive to phonetic markers of disbelief produced with rising intonation contours and conveying incredulity. When this incredulity prosody is produced with other repetitions, it sometimes leads listeners to change the conversational trajectory to restore the interlocutors’ affiliation to a shared conversational stance.

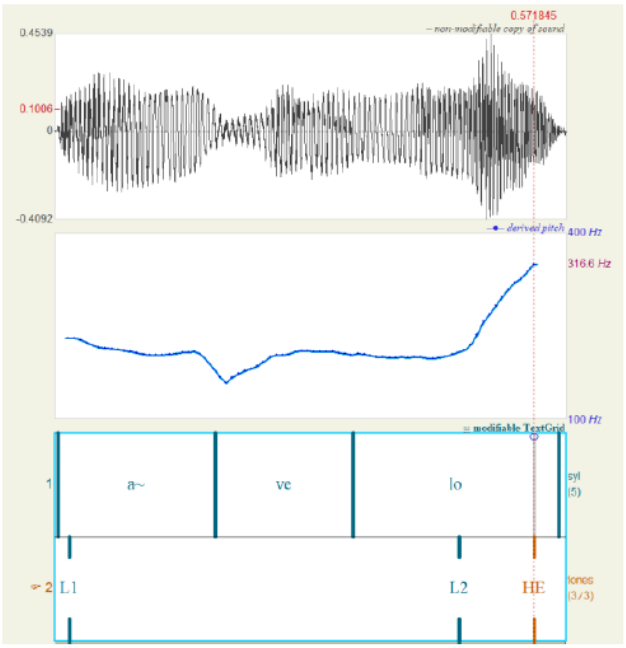


Figure 1. The rising contour on the other repetition with incredulity prosody. For confirmation prosody, the contour is phonologically similar but with shorter onset and rhyme duration for the third syllable [lo], lower H and earlier L2 alignment to that same syllable’s onset

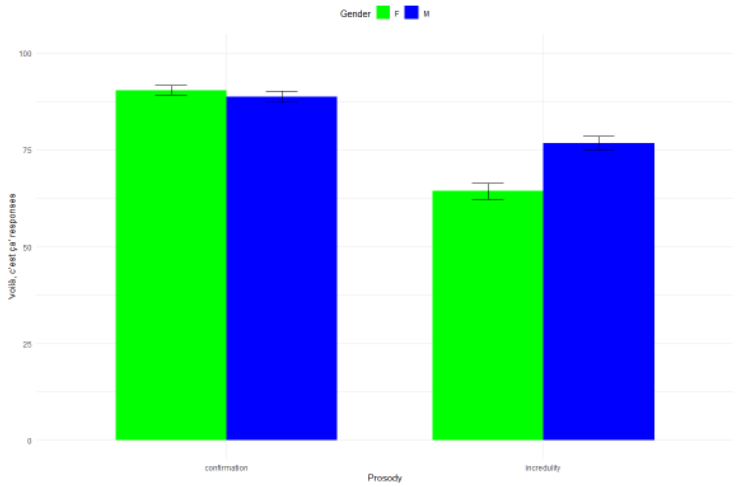


Figure 2. Boxplots representing the chosen responses depending on prosody (confirmation on the left and incredulity on the right) and on gender (women in green, men in blue). Only the prosodic difference is statistically significant.

REFERENCES

Beyssade, C., & Marandin, J. M. (2006). The speech act assignment problem revisited: Disentangling speaker's commitment from speaker's call on addressee. *Empirical issues in syntax and semantics*, 6(37-68).

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: a unified JavaScript framework for online studies. *In International Conference on Computational Social Science* (Cologne).

Persson, R. (2020). The prosody of other-repetition in French talk-in-interaction. In G. Rossi (Ed.), *Other-repetition in conversation across languages* (Special Issue). *Language in Society*, 49(4), 519–552. https://doi.org/10.1017/S0047404520000251

Rossi, G. (2020). The prosody of other-repetition in Italian: A system of tunes. *Language in Society*, 49(5), 653–684. https://doi.org/10.1017/S0047404520000627

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on language and social interaction*, 41(1), 31-57. https://doi.org/10.1080/08351810701691123

A Gender-Based Analysis of Mimi Prosodic Representations

Artem Saloev & Nicolas Ballier

(Université Paris Cité, ALTAE)

Foundation speech models increasingly rely on tokenization, such as Whisper (Radford et al., 2023). Recent models, especially Mimi (Défossez et al., 2024), a neural audio codec, encode acoustic and prosodic information. Mimi encodes speech with eight codebooks, where each codebook uses a set of 2048 discrete tokens that encode segmental and prosodic information. While these representations are assumed to be speaker-invariant abstractions, the type of correspondence between acoustic correlates and the Mimi token values in its different codebooks is still an open question, and there has been little empirical investigation into whether these prosodic cues (pitch, duration, glissando) are encoded equally well across different speakers.

We investigate how the prosodic information is encoded in Mimi and how encoding varies across gender. We present an initial probing study (Belinkov, 2022) examining how well Mimi's discrete representations capture prosodic distinctions using a controlled speaker group. We constructed a small, fully synthesized dataset using WinPitch (Martin, 2004), based on a single utterance ("He arrived on time"), in which the last syllable was systematically manipulated to vary prosodic parameters, including pitch height, duration, and glissando values. The corpus of 1000 wav files is divided into two synthetic speaker conditions generated using Microsoft SAPI voices (Microsoft, n.d.): Mark (male) and Zira (female), corresponding to lower- and higher-F_0 frequency profiles. This design allows a clean probe of whether Mimi's codebooks exhibit gender-dependent differences in prosodic predictability.

We applied a leave-one-out method using TiMBL's k-nearest neighbour classifier (Daelemans et al., 2004) to evaluate how well Mimi tokens predict the pitch values manipulated within male/female utterances. Confusion matrices enable inspection of pitch values that are reliably encoded from the point of view of their correlation with token values. TiMBL outputs Gain-ratio values, which account for the contribution of each feature to the classification task. This modeling choice reflects the discrete nature of token representations, as classification frameworks allow predictors to be treated as categorical variables, whereas regression would impose a continuous structure that may not align with the symbolic nature of the tokens. This pilot experiment reveals some speaker-dependent asymmetry in how Mimi's discrete representations encode prosodic distinctions. For the female voice type, the k-NN classifier achieves 15.5% accuracy. Gain-ratio analysis shows that numerous codebooks (especially codebook_1, 2, and 7) contribute strongly to classification, suggesting that Mimi provides more prosodic information for higher-pitched voices. In contrast, the male voice type suggests that token values are more decorrelated from pitch values, as their accuracy to predict them drops to 6.6%. Together, these results demonstrate that Mimi's discrete representations encode pitch contrasts more reliably for higher-pitched voices, revealing a systematic representational bias that emerges even in tightly controlled synthetic data.

REFERENCES

Belinkov, Y. (2022). Probing classifiers: Promises, shortcomings, and advances. *Computational Linguistics*, 48(1), 207–219.

Daelemans, W., Zavrel, J., Van Der Sloot, K., & Van den Bosch, A. (2004). Timbl: Tilburg memory-based learner. Tilburg University.

Défossez, A., Mazaré, L., Orsini, M., Royer, A., Pérez, P., Jégou, H., ... & Zeghidour, N. (2024). Moshi: A speech-text foundation model for real-time dialogue. *arXiv preprint arXiv:2410.00037*.

Martin, P (2004). WinPitch LTL II, a multimodal pronunciation software. In *InSTIL/ICALL 2004 Symposium on Computer Assisted Learning*.

Microsoft (n.d.). Appendix A: Supported languages and voices. Retrieved from https://support.microsoft.com/windows/appendix-a-supported-languages-and-voices-4486e345-7730-53da-fcfe-55cc64300f01

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pp. 28492–28518.

A first look at lexical stress in Shughni

Sofia Sedunova & Yury Makarov

(Higher School of Economics, Moscow, Russia / University of Cambridge, United Kingdom / Vinogradov Russian Language Institute, RAS, Russia)

Shughni, an Iranian (< Indo-European) language spoken by ca. 100,000 speakers in Tajikistan and Afghanistan, has been claimed to exhibit an increase in the fundamental frequency and loudness of stressed vowels [1]. Nevertheless, no studies have provided acoustic data in support of this statement, and the difficulty of researching this topic is not limited to the lack of empirical evidence. Since Shughni presumably has word-final stress placement (cf. other Iranian languages, e.g., Tajik or Persian) [1–3], it is difficult to find (near-)minimal pairs to study stressed and unstressed vowels in identical environments.

In our study, we address both issues using recent field data (786 recordings; 6 speakers: 3 male, 3 female; mean age = 41.7; recorded in Khorugh, Tajikistan, in 2025). To track acoustic changes in the same vowel (stressed vs. unstressed), we use:

1. words with identical vowels (e.g., *bırık* ‘thin; narrow’), assuming word-final stress;
2. word pairs containing the stressed derivational suffix *-ı*; e.g., in *safed* ‘white’ vs. *safedı* ‘white(ness)’, the same vowel can be assessed in stressed and unstressed positions.

The measured acoustic parameters include vowel duration, intensity, and mean F0, as well as F1 and F2 measured at the midpoint. The data were annotated in Praat and processed using Python. For statistical analysis, we employed a Linear Mixed Effects Model using `statsmodels` [4] with formula (1).

Param∼Condition+Age+Gender+(1|Speaker)+(1|PairID)

Under condition (1), lexical stress resulted in a main effect on duration across all vowel types, with stressed vowels being on average 32 milliseconds longer than unstressed vowels. Stress also significantly increased fundamental frequency by an average of 8.7 Hz and shifted F2 forward by 87 Hz. Conversely, stress resulted in a slight but significant decrease in average intensity by 1.9 dB. As for condition (2), the stress effect on duration and F0 remained highly significant, adding an average of 42 milliseconds and 14.7 Hz to the vowel, respectively. However, no significant effects were found for F1 and F2. At the same time, mean intensity increased on average by 2 dB. Additionally, Figure 1 illustrates changes associated with stressed and unstressed conditions affecting Shughni vowel quality.

Overall, the observed differences in duration and F0 between stressed and unstressed vowels align well with word stress typology [5] and partially confirm Olson’s claim. At the same time, the latter overlooks important changes in duration associated with the stressed position. The interaction of phonological vowel length and lexical stress needs further investigation to determine whether the vowel duration difference is neutralized in an unstressed position. The lack of significant changes in F1 and F2 confirms that phonetic quality reduction is not present in Shughni (at least in disyllabic words). Nevertheless, Figure 1 shows that stressed vowels — except for /ɛ/ and /ø/ — tend to be more central compared to unstressed ones; therefore, the effect of stress on vowel quality is limited to certain vowels.

As for intensity, the observed effects are highly significant but point in opposite directions. Under condition (1), the drop (-1.901 dB) might be associated with the presence of a phrase boundary, which is signalled by a reduction in intensity in phrase-final position [6]. Under condition (2), where the stressed stem vowel is not an utterance-final segment, the 'true' (and typologically expected) intensity effect emerges (+2.060 dB).

The obtained results appear consistent with the context of Iranian languages; intensity and duration are, for example, well-known correlates of stress in Persian (though the former is claimed to be dependent on phrasal accent) [7] and in Kurdish, where F0, mean intensity, and duration show higher values in stressed syllables [8].

Parameter	Condition	Estimate (β)	Std. Error	t-value	p-value
duration	(1) Identical vowels in one word	0.032	0.004	8.504	< 0.001 ***
duration	(2) Vowels in word pairs (with and without *-i*)	0.042	0.002	21.793	< 0.001 ***
F1_mid	(1) Identical vowels in one word	-6.482	12.070	-0.537	0.591
F1_mid	(2) Vowels in word pairs (with and without *-i*)	19.119	12.060	1.585	0.113
F2_mid	(1) Identical vowels in one word	87.067	34.883	2.496	0.013 *
F2_mid	(2) Vowels in word pairs (with and without *-i*)	-24.527	24.730	-0.992	0.321
F0_mean	(1) Identical vowels in one word	8.685	3.210	2.706	0.007 **
F0_mean	(2) Vowels in word pairs (with and without *-i*)	14.742	2.155	6.840	< 0.001 ***
Intensity_mean	(1) Identical vowels in one word	-1.901	0.465	-4.089	< 0.001 ***
Intensity_mean	(2) Vowels in word pairs (with and without *-i*)	2.060	0.286	7.202	< 0.001 ***

Table 1. *Main effects of lexical stress and stress shift across all vowels*

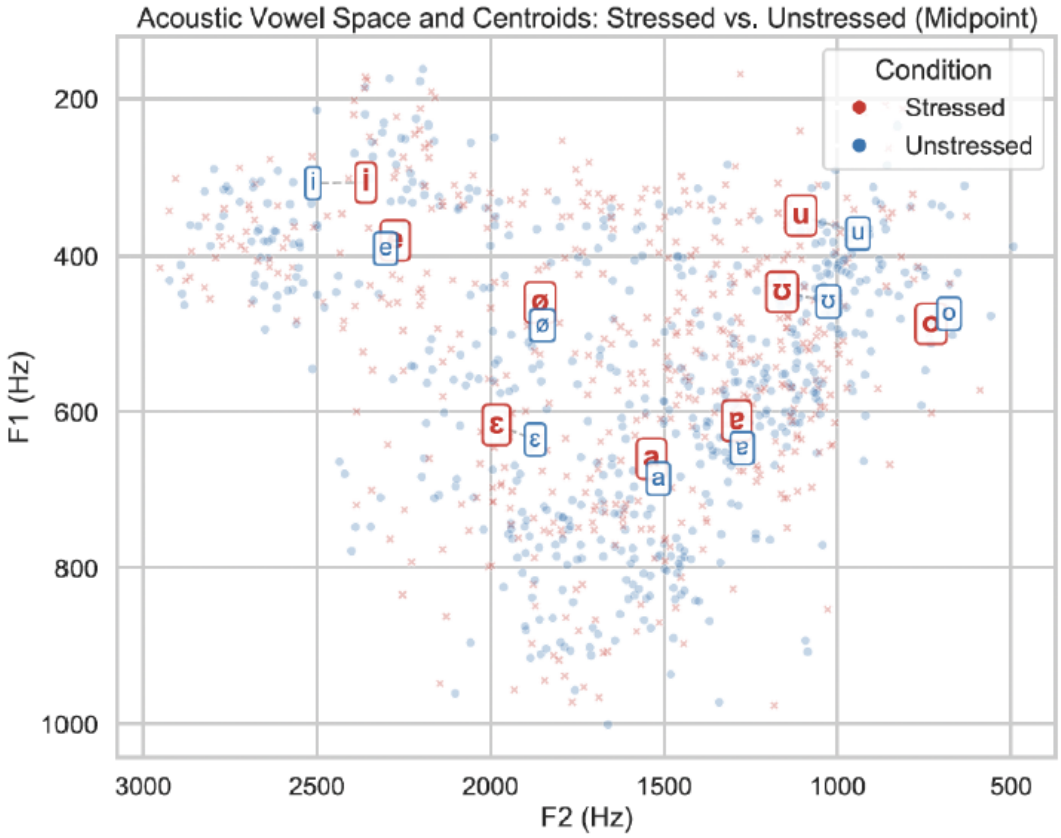


Fig. 1. *Shughni vowel space in stressed and unstressed conditions; /ɪ/ is not included*

REFERENCES

[1] K. Olson, *Shughni Phonology Statement*. SIL International, 2017.

[2] D. I. Edelman and Sh. P. Yusufbekov, ‘Shugnanskij jazyk [Shughni]’, in *Jazyki mira: Iranskie jazyki III*, Moscow: Indrik, 1999, pp. 225–242.

[3] D. I. Edelman and L. R. Dodykhudoeva, ‘Shughni’, in *The Iranian Languages*, G. Windfuhr, Ed., London: Routledge, 2009, pp. 787–824.

[4] S. Seabold and J. Perktold, ‘statsmodels: Econometric and statistical modeling with python’, in *9th Python in Science Conf.*, 2010.

[5] M. Gordon and T. Roettger, ‘Acoustic correlates of word stress: A cross-linguistic survey’, *Ling. Vanguard*, vol. 3, no. 1, 2017.

[6] L. A. Streeter, ‘Acoustic determinants of phrase boundary perception’, *The Journal of the Acoustical Society of America*, 64(6), 1582-1592, 1978.

[7] V. Sadeghi, ‘Acoustic correlates of lexical stress in Persian’, in *Proc. ICPhS XVII*, Hong Kong, 2011, pp. 1738–1741.

[8] A. Mohammadi, ‘Acoustic correlates of stress in Central Kurdish’, in *PaPE 2025 Book of Abstracts*, Palma, 20, 2025.

Gesture and intonation in trainee teachers' L2 English: Alignment with prominence and responses to manual constraints

Šárka Šimáčková

(Palacký University Olomouc, Czech Republic)

Speech prosody and gestures interact (Kendon 1980, Wagner et al. 2014). Non-referential beat gestures have been linked to both prosodic boundaries (Krivokapić 2014) and prominence (Prieto et al. 2018). In native English, gesture peaks align closely with pitch-accented syllables (Loehr 2012), and gestures and prosodic phrasing have been argued to emerge from shared planning (Shattuck-Hufnagel & Ren 2018). We ask whether comparable coordination is observed in fluent L2 English, examining (1) the temporal alignment of gesture strokes and pitch accents and (2) the relationship between hand movement and melodic dynamism.

We report a pilot testing of a larger project on student-directed teacher talk. Four Czech trainee teachers of English produced seven classroom instructions in a 2×2 design crossing audience age (younger vs. older) with hand-movement condition (free vs. fixed hands).

To examine the gesture-prominence alignment, we analysed the free-movement productions (56 instructions in total), annotating syllables and pitch accents in Praat (Boersma & Weenink 2024) and gesture apices in ELAN (Sloetjes 2017). For each apex, the temporal distance from the nearest prominent syllable onset and its F0 peak was calculated following Loehr (2012). The dataset contains 176 stroke apices, of which 137 were associated with prominent syllables, with no difference between the audience age conditions. Most apices occurred after the syllable onset and clustered near the pitch accent with values on both sides of zero, indicating that the gesture peak could precede or follow the tonal target. Overall, the findings suggest systematic gesture–prominence coordination in this type of fluent L2 speech, while also revealing individual variation: two speakers show tightly timed pre-accentual apices, whereas two exhibit broader distributions.

To explore the link between hand movement and melodic dynamism, we measured mean absolute pitch slope (MAS, ST/s) in all 112 utterances. For each audience-age condition, the impact of gesture restriction was operationalized as the difference between matched productions with free versus fixed hands (DIFF=Slope$_{free}$−Slope$_{fixed}$), where larger positive values indicate a greater reduction of pitch movement for restrained hands. We expected restraining the teachers' hands to limit their ability to enhance pitch variation, particularly in speech directed to younger learners, typically associated with greater expressivity (Kuder 2020). Against our expectations, restraining the hands tended to reduce MAS in utterances to the older (mean DIFF=+1.73 ST/s) but not the younger audience (mean DIFF=−0.87). At the individual level, this pattern is observed especially for teachers T1 and T4 (see Fig. 1). T3 showed the opposite tendency, with higher slopes in the fixed-hand condition (negative DIFF), especially when addressing younger students. This suggests possible compensatory enhancement of vocal modulation when manual gesture was unavailable.

The analyses suggest that temporal alignment between gestures and prominence is present in L2 speech, while restricting the hands is associated with shifts in melodic dynamism, potentially pointing to adaptive trade-offs between modalities.

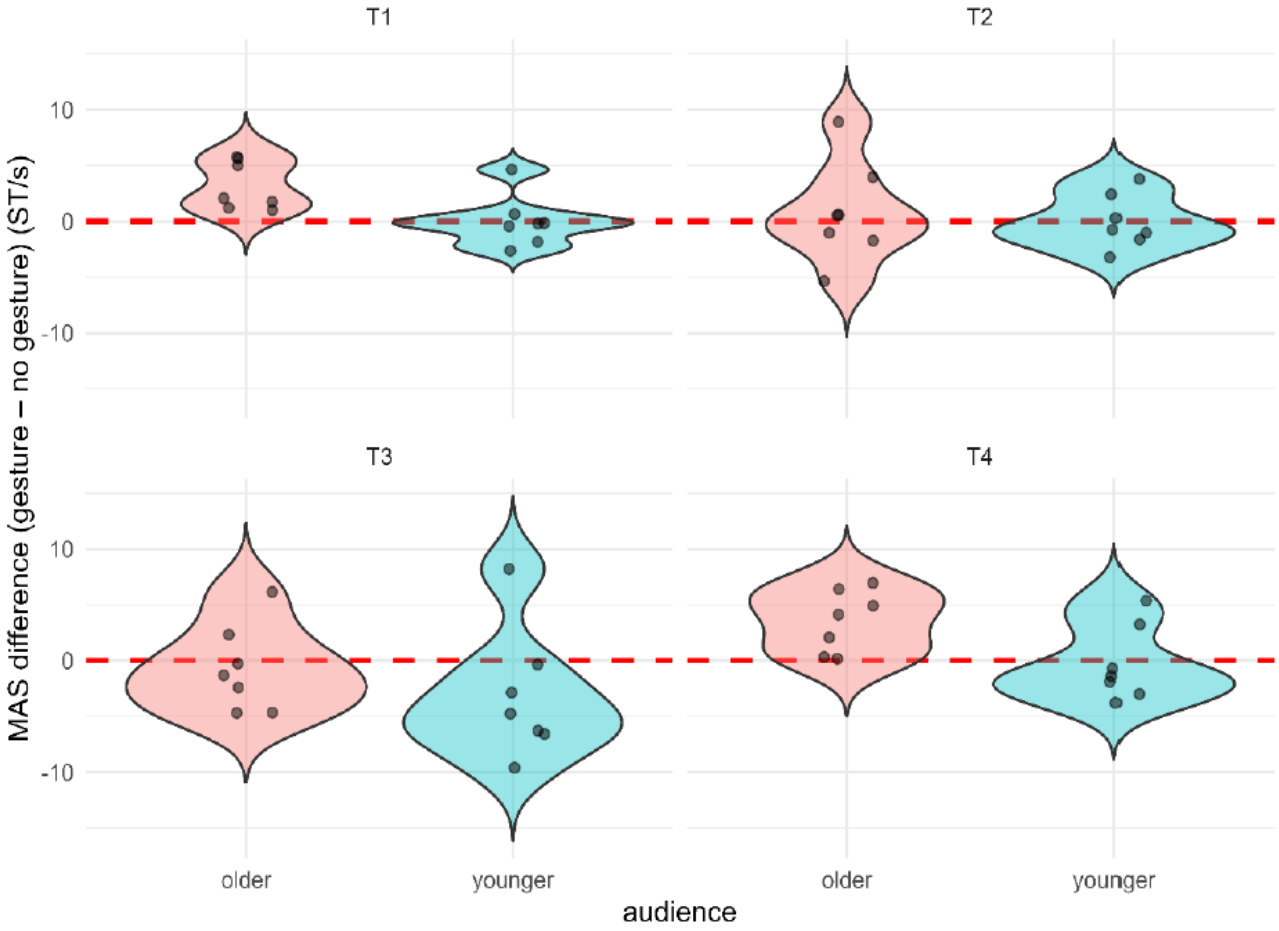


Figure 1. Gesture-related change in mean absolute slope (MAS difference = gesture − no gesture) by teacher and audience. Dots correspond to individual items and violins depict the distribution of values. The dashed red line indicates zero difference.

REFERENCES

Boersma, P. & Weenink, D. (2024). *PRAAT* [Computer program]. Version 6.4.20.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the... In *The relationship of verbal and nonverbal communication*, 25, 207.

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130397.

Kuder, E. (2020). *Second language teacher prosody*. Routledge.

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory phonology*, 3(1), 71-89.

Prieto Vives, P., Cravotta, A., Kushch, O., Rohrer, P. L., & Vilà-Giménez, I. (2018). Deconstructing beat gestures: a labelling proposal. *9th Int. Conf. on Speech Prosody*. Poznan.

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, 1514.

Sloetjes, H. (2017). *ELAN* [Computer program]. Version 6.7.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech communication*, 57, 209-232.

The Realisation of Narrow Focus in Glaswegian English

Yao Vera Yujia

(University of Glasgow, United Kingdom)

Linguistic prominence and focus are shaped by syntactic structure and communicative needs (Lambrecht, 1994). Different focus types—broad, narrow, and contrastive—are commonly realised through prosodic cues (Gussenhoven, 2005). For example, constituents under narrow focus often show higher pitch than under broad focus in languages such as Mandarin Chinese (Lee et al., 2015), American English (Xu & Xu, 2005), and German (Baumann et al., 2006). In addition, increased duration and intensity in the on-focus region have been reported across many languages, including American English (Liu & Xu, 2007; Xu & Xu, 2005), German (Baumann et al., 2006), Mandarin Chinese (Xu, 1999), Yoloxóchitl Mixtec (DiCanio et al., 2018), Hijazi Arabic (Alzaidi et al., 2019), and French (Lee et al., 2015).

In many languages, the post-focus region exhibits systematic phonetic reduction in duration, F_0, and intensity, a phenomenon known as post-focus compression (PFC). However, PFC is absent in some languages, such as Taiwan Mandarin (Xu et al., 2012) and Cantonese (Wu & Xu, 2010), underscoring the need for cross-linguistic investigation.

This study examines the prosodic realisation of focus in Glaswegian English, contributing to broader work on how focus is marked across languages and English varieties. All data collection has been completed, and we have carried out initial processing and preliminary interpretation of the dataset. These preliminary results suggest patterns that differ from much of the existing literature, which has largely focused on General American English and has often been treated as representative of English more generally.

In particular, our analyses indicate that Glaswegian English does not exhibit the same post-focus compression (PFC) pattern in F_0: unlike General American English, pitch does not show systematic post-focus lowering relative to broad focus. For initial and medial narrow focus, there is no consistent pattern in the pre-focus region. In the on-focus region, the contour shows a fall–rise compared to broad focus. Within the narrow-focus contour, the F_0 minimum is located within the focused word. In the post-focus region, F_0 was significantly higher under both initial and medial narrow focus than under broad focus: in each comparison (narrow − broad), the estimated difference curve remained positive, and its 95% confidence interval lay entirely above zero.

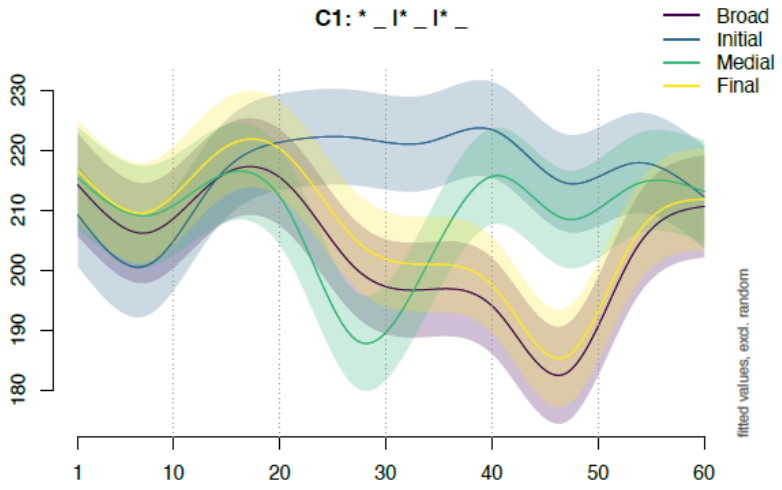


Figure 1F0 contours for narrow focus (initial/medial/final) and broad focus in six-syllable, initial-stress sentences (e.g., Lily married Molly; Marry ruined Lorries).

REFERENCES

Alzaidi, M. S., Xu, Y., & Xu, A. (2019). Prosodic encoding of focus in Hijazi Arabic. *Speech Communication*, 106, 127–149. https://doi.org/10.1016/j.specom.2018.12.006

Baumann, S., Grice, M., & Steindamm, S. (2006). Prosodic Marking of Focus Domains-Categorical or Gradient. *Prococeedings from Speech Prosody 2006*, 301–304.

DiCanio, C., Benn, J., & Castillo García, R. (2018). The phonetics of information structure in Yoloxóchitl Mixtec. *Journal of Phonetics*, 68, 50–68. https://doi.org/10.1016/j.wocn.2018.03.001

Gussenhoven, C. (2005). Transcription of Dutch Intonation (S.-A. Jun, Ed.; pp. 118–145). Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199249633.003.0005

Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620607

Lee, Y., Wang, B., Chen, S., Adda-Decker, M., Amelot, A., Nambu, S., & Liberman, M. (2015). A crosslinguistic study of prosodic focus. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4754–4758. https://doi.org/10.1109/ICASSP.2015.7178873

Liu, F., & Xu, Y. (2007). Question intonation as affected by word stress and focus in English. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1189–1192.

Wu, W. L., & Xu, Y. (2010). Prosodic focus in Hong Kong Cantonese without post-focus compression. *Speech Prosody 2010-Fifth International Conference*.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, 27(1), 55–105. https://doi.org/10.1006/jpho.1999.0086

Xu, Y., Chen, S., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review*, 29(1). https://doi.org/10.1515/tlr-2012-0006

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33(2), 159–197. https://doi.org/10.1016/j.wocn.2004.11.001

How do semantic likelihood and information structure affect prosodic encoding in different tasks?

Ivan Yuen, Bistra Andreeva, Bernd Möbius & Mitko Sabev

(Saarland University, Germany)

Predictability has been used to account for phonetic variations, often assuming a link between 'probability' of occurrence and signal properties of speech at multiple linguistic levels [1, 2, 5, 6, 9] including prosody [1, 5, 6]. However, other accounts turn to semantics and pragmatics, evoking linguistic notions such as information structure to directly account for prosodic variations [4]. It is uncommon to combine both viewpoints in examining the role of prosody in phonetic encoding of speech. One of the few came from [7] which manipulated meaning-based contextual probability and reported its influence on the fundamental frequency contour in different focus conditions in American English. A recent study of broadcast data in German also observed contributions of information status and surprisal on syllable duration [10]. However, [10] differed from [7]: the former is corpus-based and uses trigram surprisal (which is structure-related); whereas the latter is experiment-based and uses pragmatic contextual probability (which is discourse meaning-related). That is, data were elicited differently.

In light of these differences, the current study revisited the role of information theoretical and information structure properties in prosodic encoding of duration in German using two experimental tasks: reading aloud and sentence formulation. The two experimental tasks were chosen to differ systematically in their speech planning demands.

The likelihood of semantic association, as an index of semantic surprisal, was varied between two nouns in a sentence such as, "Sie hat eine Bibel (N1) in der Kirche (N2) gefunden" (*She has found a Bible in the church*) vs. "Sie hat eine Bibel (N1) in der Arztpraxis (N2) gefunden" (*She has found a Bible in the medical practice*). N1 and N2 words served as triggers in an online word-association experiment. The responses were checked against the N1-N2 test pairs for 'a match' (more likely) vs. 'no match' (less likely). The sentences were elicited via auditory prompt questions in three focus conditions: Broad, Narrow N1, Narrow N2. While the sentence stimuli were presented orthographically in the reading task, pictures of N1 and N2 were visually presented with an orthographically presented verb for the participant to formulate a sentence in the formulation task. We expect:

1. duration of N1 and N2 to change as a function of focus assignment,
2. a reduced durational difference between N1 and N2 in a semantically less likely pair, and
3. generalizability across experiments.

N1 and N2 duration were analyzed using `lmer` in R [3, 8], with FOCUS, SEMANTIC LIKELIHOOD and EXPERIMENTAL TASK as fixed effects. A preliminary analysis of 20, drawn from a total sample of 39 participants, showed the effect of FOCUS for N1 ($F=14.6$, $df=2$, $p=.0001*$) and N2 duration ($F=10.02$, $df=2$, $p=.001*$), in line with (i). The effect of SEMANTIC LIKELIHOOD reached statistical significance for N1 ($F=4.46$, $df=1$, $p=.04*$), but not N2 duration, in partial agreement with (ii). In addition, there was a significant interaction between FOCUS and EXPERIMENTAL TASK on N1 duration ($F=8.34$, $df=2$, $P=.002*$), contra (iii). The interaction arose for N1 duration because the Narrow N1 condition is longer than the Narrow N2 condition in the reading experiment; whereas the Broad focus condition is

longer than the Narrow N1 and Narrow N2 conditions in the sentence formulation experiment (Figure 1).

Post-hoc analysis further suggests that SEMANTIC LIKELIHOOD on N1 is driven by the Broad focus condition in the sentence formulation experiment. As for N2 duration, the Narrow N2 condition is longer than the Broad and Narrow N1 conditions in both experiments. Together, the results suggest that information structure robustly shapes prosodic encoding, while semantic likelihood exerts more limited and position-specific effects that depend on speech planning conditions.

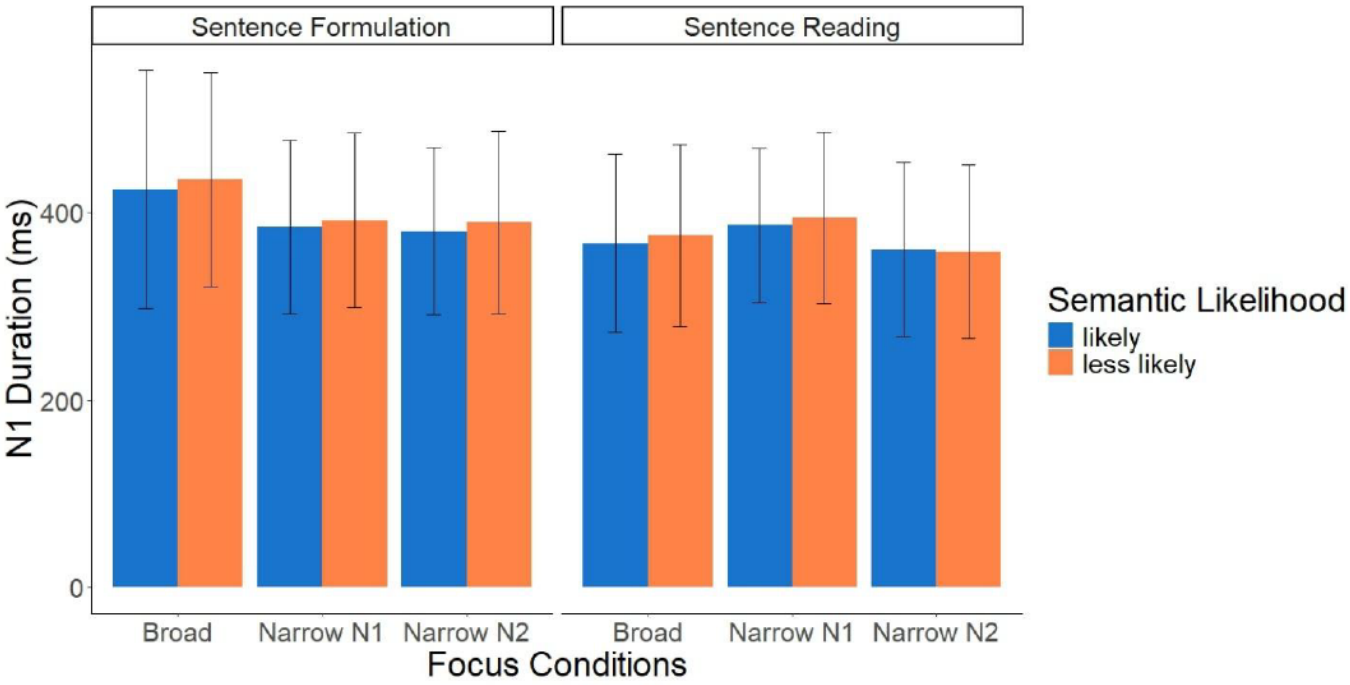


Figure 1. N1 duration as a function of focus conditions in two experimental tasks, with +/-1SD

REFERENCES

[1] M. Aylett and A. Turk, "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech," *Language and Speech*, 47: 31-56, 2004.

[2] R. Baker and A. Bradlow, "Variability in word duration as a function of probability, speech style and prosody," *Language and Speech*, 52(4): 391-413, 2009.

[3] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting Linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, pp. 1-48, 2015.

[4] S. Baumann, M. Grice and S. Steindamm, "Prosodic marking of focus domains – categorical or gradient?" *Proceedings of Speech Prosody*, Dresden, May, 2006.

[5] E. Brandt, B. Möbius and B. Andreeva, "Dynamic formant trajectories in German read speech: Impact of predictability and prominence," *Frontiers in Communication*, 6, 2021. doi: 10.3389/fcomm.2021.643528.

[6] Z. Malisz, E. Brandt, B. Möbius, M. O. Yoon and B. Andreeva, “Dimensions of segmental variability: Interactions of prosody and surprisal in six languages,” *Frontiers in Communication,* 2018. doi: 0.3389/comm2018.00025

[7] I. Ouyang and E. Kaiser, “Prosody marks different kinds of informativity: interactions between frequency, probability and focus,” *University of Pennsylvania Working Papers in Linguistics*, vol. 21, issue 1, article 24, 2015.

[8] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, 2023.

[9] S. Seyfarth, “Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation,” *Cognition*, 133:140-155, 2014.

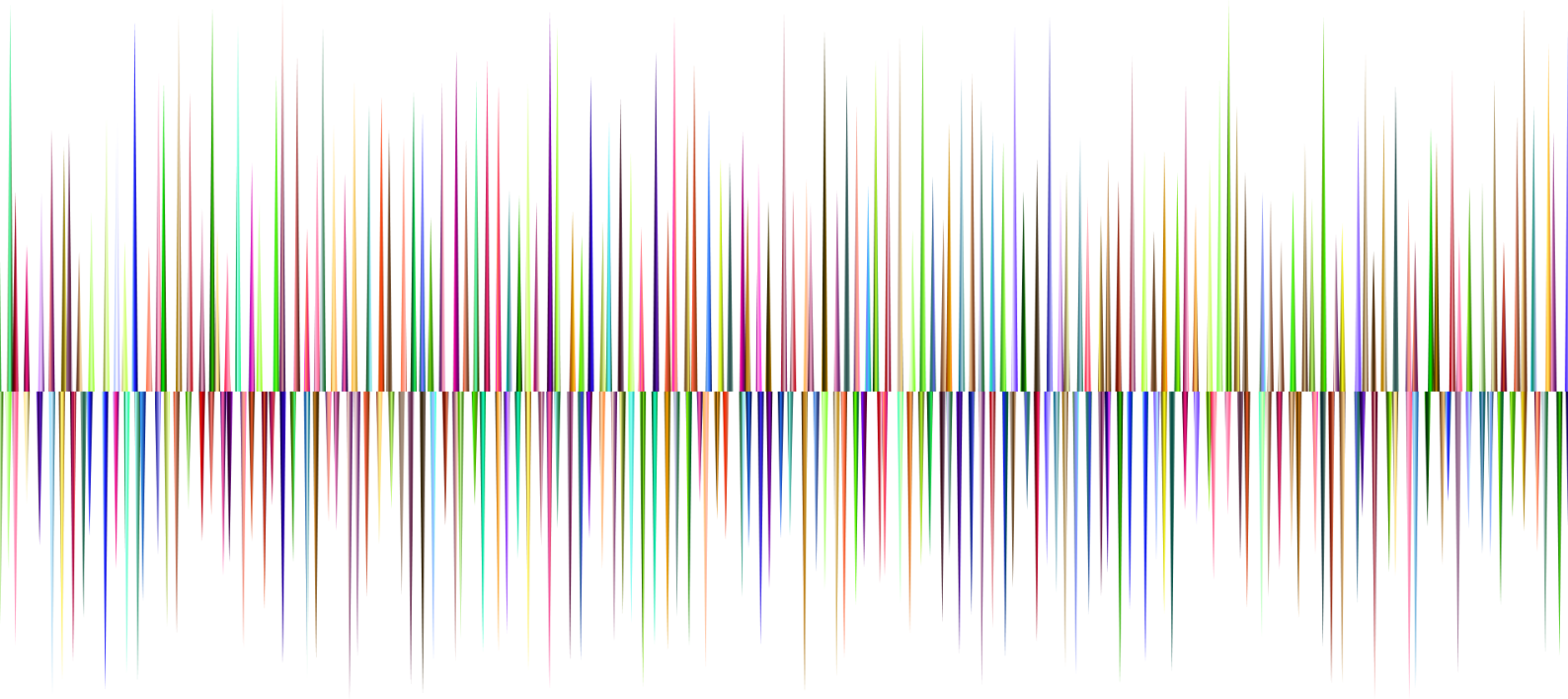
[10] I. Yuen, B. Andreeva, O. Ibrahim and B. Möbius, “Prosodic factors do not always suppress discourse or surprisal factors on word-final syllable duration in German polysyllabic words,” in R. Lemke, L. Schäfer, and I. Reich editors, *Information Structure and Information Theory*, pp. 215-234. Language Science Press, Berlin, 2024.

15:10 Parallel Sessions 8

Jun Wang (Université Grenoble Alpes, France) Perceptual vs. Acoustic Correlates of Prosodic Prominence in French

Sophie Fetter & Bettina Braun (Universität Konstanz, Germany) The role of prosodic cues for the interpretation of rhetorical questions

Varvara Petrova (Lomonosov Moscow State University, Russia) Word-level prosody of Digor Ossetic in a cross-dialectal perspective



Perceptual vs. Acoustic Correlates of Prosodic Prominence in French: A Study of Chinese Learners' Performance

Jun Wang

(Université Grenoble Alpes, France)

This study investigates how Chinese learners of French perceive prosodic prominence, comparing their perceptual judgments with acoustic measurements. The research addresses a key methodological question: How do auditory and instrumental evaluations compare in measuring prominence?

Prosodic prominence in French is primarily marked at the right edge of prosodic units (Jun & Fougeron, 2002; Delattre, 1966), with acoustic correlates including F0, duration, and intensity (Di Cristo, 1998; Vaissière, 2002; Martin, 2015). The accentual phrase, fundamental to French prosodic structure, typically displays final lengthening and F0 rise on the phrase-final syllable (Jun & Fougeron, 2000; Post, 2000). Non-native speakers, particularly those from tonal language backgrounds, may perceive prominence differently due to L1 transfer effects (Rasier & Hiligsmann, 2007; Gut, 2009).

We designed a three-level perception test targeting 34 Chinese undergraduate learners of French (ages 19-22, intermediate proficiency), varying systematically in prosodic complexity: 9 groupes rythmiques (2-4 syllables), 10 phrases courtes (5-9 syllables), and 5 phrases longues (11-20 syllables), yielding 187 syllables across 24 native-speaker recordings. Participants rated each syllable's prominence as strong (forte), weak (faible), or unsure (pas certain). Background questionnaires collected data on participants' phonetic training (primarily segmental), oral practice frequency, length of study, and musical ability. Syllable segmentation was performed using the Montreal Forced Aligner (McAuliffe et al., 2017) with French acoustic models. Acoustic analysis was conducted in Praat (Boersma & Weenink, 2024), extracting F0, duration, and intensity for all syllables. Within-utterance Z-score normalization allowed calculation of composite prominence scores, with threshold-based classification calibrated to maximize identification of phrase-final positions, providing an objective acoustic standard.

Preliminary acoustic-perceptual comparison reveals substantial divergence between learner judgments and acoustic measurements. Overall agreement rate is 25%, with learners systematically overestimating prominence : they identified 54% more syllables as strong (n=2.258) than acoustic analysis did (n=1.462). These findings strongly indicate L1 transfer effects, with learners likely over-relying on F0 cues while under-weighting duration, a pattern consistent with Mandarin Chinese's tonal system prioritizing pitch information (Jongman, Wang, Moore, & Sereno, 2010; Chen, 2006).

Complete analysis will: (1) calculate overall accuracy rates and correlation coefficients between perceptual judgments and acoustic measurements, (2) identify systematic error patterns through confusion matrix analysis, examining whether learners tend to over- or under-estimate prominence, (3) use multiple regression to test whether F0 disproportionately predicts learner judgments compared to duration and intensity, testing the L1 transfer hypothesis, and (4) investigate how learner characteristics (phonetic training, oral practice frequency, length of French study, musical ability) influence perceptual accuracy, and (5) analyze whether accuracy decreases with increasing utterance length.

This study contributes methodologically by demonstrating how integrated perceptual-acoustic approaches reveal non-native prominence perception mechanisms, underscoring

prominence's complexity as an acoustic construct. Results illuminate L1 transfer effects in suprasegmental perception, particularly regarding the differential weighting of acoustic cues by tonal versus non-tonal language speakers.

REFERENCES

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.2.06) [Computer software]. https://www.praat.org

Chen, Y. (2006). Durational adjustments under contrastive focus in Mandarin Chinese. *Journal of Phonetics*, 34(2), 176–201. https://doi.org/10.1016/j.wocn.2005.05.002

Delattre, P. (1966). Les dix intonations de base du français. *The French Review*, 40(1), 1–14. https://doi.org/10.2307/385815

Di Cristo, A. (1998). Intonation in French. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 195–218). Cambridge University Press.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In E. Bates, L. H. Tan, & O. J. L. Tzeng (Eds.), *Handbook of Chinese psycholinguistics* (pp. 209–217). Cambridge University Press.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology* (pp. 209–242). Kluwer Academic Publishers.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, 14(1), 147–172. https://doi.org/10.1515/prbs.2002.005

Martin, P. (2015). *The structure of spoken language: Intonation in Romance*. Cambridge University Press.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of Interspeech 2017* (pp. 498–502). https://doi.org/10.21437/Interspeech.2017-1386

Post, B. (2000). *Tonal and phrasal structures in French intonation*. Holland Academic Graphics.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2: Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, 28, 41–66.

Vaissière, J. (2002). Cross-linguistic prosodic transcription: French vs. English. In N. B.

Volskaya, N. D. Svetozarova, & P. A. Skrelin (Eds.), *Problems and methods of experimental phonetics: In honour of the 70th anniversary of Professor L. V. Bondarko* (pp. 147–164). St. Petersburg State University Press.

The role of prosodic cues for the interpretation of rhetorical questions: Evidence from an indirect lexical task

Sophie Fetter & Bettina Braun

(Universität Konstanz, Germany)

The prosodic realization of utterances affects their pragmatic interpretation (1; 2; 3). Establishing form-meaning relationships relies on controlled production data (e.g. 4; 5) or perception experiments that elicit participants' interpretation (e.g. 6; 7). Ideally, results from production and perception converge. Mismatches may suggest that the tasks are not ideally-suited or that the two modalities have different constraints. For the production and perception of rhetorical questions (which serve to make a point, henceforth RQs, cf. 8) compared to information-seeking questions (which serve to gather information, ISQs), (6) found mismatches between production and perception. For instance, in production, there were few prenuclear accents, but in perception, prenuclear L*+H led to more RQ interpretations, and prenuclear H* to more ISQ interpretations. Since the role of prenuclear accents has been contentious (9; 10; 11; 12), these mismatches deserve further scrutiny. We suggest a more indirect, lexical task (13), which may be easier for participants.

To this end, we used four minimal pairs (e.g. leiden ' to suffer' – leiten 'to lead') and altered VOT to 40ms to create an ambiguous version, using a praat script (14). These words were embedded at the end of wh-interrogatives, see (1).

(1) Wer will denn lei[d/t]en?

Who wants PRT lei[d/t]en

'Who wants to suffer ("RQ-word") / lead ("ISQ-word")?'

The wh-interrogatives were constructed such that one member of the pair was more compatible with an RQ-reading (i.e. nobody wants to suffer, henceforth "RQ-word"), the other with an ISQ-reading (i.e. inquiring who wants to take up a leading position, henceforth "ISQ-word"). The interrogatives were prosodically produced as RQ or ISQ, each modeled after prior production or perception findings, cf. (6; 15) (cf. Fig. 1). Specifically, RQs had a steep rising accent with both the low and high tonal targets in the stressed syllable (LH)*, ISQs an H* accent on the target word, both followed by a fall (L-%); the production model had no prenuclear accent, the perception model did (L*+H for RQ, H* for ISQ). The 16 experimental items (4 items x 2 illocution types x 2 models) were interleaved with 8 filler items (half RQ, half ISQ-prosody, one for production, one for perception model), resulting in 4 pseudo-randomized lists (separating same items with different prosodic realizations). The experiment was controlled with presentation (Neurobiological Systems). On each trial, participants saw two words on screen (position of ISQ and RQ word counterbalanced across lists), heard a wh-interrogative and clicked the left or right button on a button box. The expected response side was matched within lists (16 times each). Responses and reaction times (RT, relative to end of target word) were recorded.

Results from 30 native German participants showed a main effect of illocution type ($p < 0.05$), but no effects of model ($p > 0.15$) and no interaction ($p > 0.5$), see Fig. 2. As predicted, there were more clicks to the "ISQ-word" (e.g. leiten in (1)) when the prosody signaled an ISQ (59%) than when it signaled an RQ (46%). RT showed faster responses for ISQs overall ($p <$

0.05) and an interaction between response ("RQ-word" or "ISQ-word") and prosodically cued illocution type ($p < 0.05$): RQ-responses took longer with mismatching ISQ prosody.

The current procedure allowed us to test pragmatic interpretation more indirectly (10). The prosodic context influenced the lexical decision (akin to Ganong-style lexical effects on phoneme categorization 16), but the effects were weak. Post-hoc analyses showed that two items, which contained a further stop, led to responses at ceiling or floor, suggesting variation across items. The RT analysis showed generally faster responses for ISQ interpretations (cf. also 6), but the prosodic realization affected RTs for clicks on the "RQ-word". In future studies we plan to include more items to increase statistical power (also with spectral contrasts rather than temporal contrasts, cf. 17), use pictures to reduce effects of orthography and to include more fillers to better disguise the experimental items. Increasing the number of items is expected to allow us to test whether the models taken from prior production and perception findings are really equally effective.

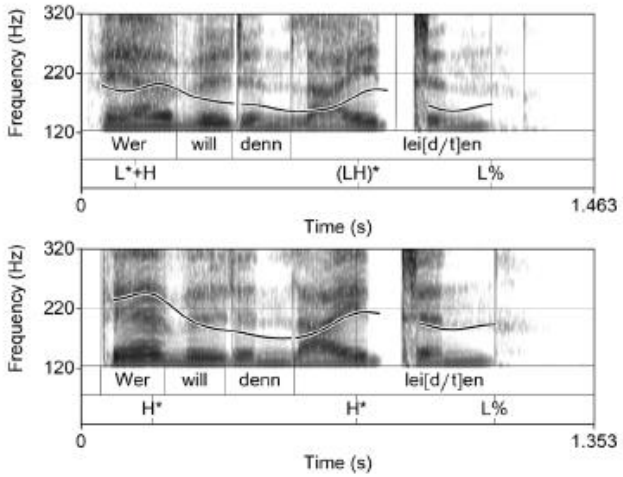


Figure 1: Prosodic RQ (top) and ISQ (bottom).

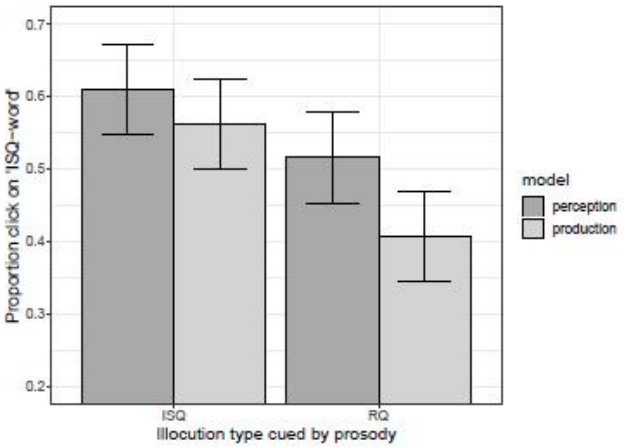


Figure 2: Proportion of click to "ISQ-word". Whiskers show the 95% confidence interval.

REFERENCES

[1] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, 2008.

[2] G. Elordieta and P. Prieto, *Prosody and Meaning*. Walter de Gruyter, 2012.

[3] D. Dahan, "Prosody and language comprehension," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 6, no. 5, pp. 441–452, 2015.

[4] B. Braun, N. Deh´e, J. Neitsch, D. Wochner, and K. Zahner, "The prosody of rhetorical and information-seeking questions in german," *Language and Speech*, vol. 62, no. 4, pp. 779–807, 2019. PMID: 30563430.

[5] M. del Mar Vanrell, I. Feldhausen, and L. Astruc, "The discourse completion task in Romance prosody research: Status quo and outlook," *Methods in Prosody: A Romance Language Perspecive*, vol. 191, 2018.

[6] B. Braun, N. Deh´e, M. Einfeldt, A. James, E. Kazak, E. Sevastjanova, and K. Zahner-Ritter, "A multi-cue study to the interpretation of German information-seeking and rhetorical questions," *Laboratory Phonology*, vol. 16, pp. 1–40, 12 2025.

[7] C. Gussenhoven, "Discreteness and gradience in intonational contrasts," *Language and Speech*, vol. 42, no. 2-3, pp. 283–305, 1999.

[8] M. Biezma and K. Rawlins, "Rhetorical questions: Severing asking from questioning," *Proceedings from Semantics and Linguistic Theory*, vol. 27, p. 302, 2017.

[9] C. Petrone and O. Niebuhr, "On the intonation of German intonation questions: The role of the prenuclear region," *Language and Speech*, vol. 57, no. 1, pp. 108–146, 2014.

[10] B. Braun and M. Biezma, "Prenuclear L*+ H activates alternatives for the accented word," *Frontiers in psychology*, vol. 10, p. 1993, 2019.

[11] D. B¨uring, "Intonation, semantics and information structure," *The Oxford handbook of linguistic interfaces*, pp. 445–474, 2007.

[12] S. Baumann, J. Mertens, and J. Kalbertodt, "How 'ornamental' are German prenuclear accents," in *Proceedings of Prosody and Meaning confernece PaM17*, 2017.

[13] L. C. Dilley and J. D. McAuley, "Distal prosodic context affects word segmentation and lexical processing," *Journal of Memory and Language*, vol. 59, no. 3, pp. 294–311, 2008.

[14] M. Win. https://github.com/ListenLab/VOT/blob/master/Make_VOT_Continuum_current_version.txt, 2022. vers. 33 from 07.12.2022, downloaded on 08.11.2025.

[15] K. Zahner-Ritter, M. Einfeldt, D.Wochner, A. James, N. Deh´e, and B. Braun, "Three kinds of rising-falling contours in German wh-questions: Evidence from form and function," *Frontiers in Communication*, vol. 7, p. 838955, 2022.

[16] W. Ganong, "Phonetic categorization in auditory word perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 6, pp. 110–125, 02 1980.

[17] F. Eisner and J. M. McQueen, "The specificity of perceptual learning in speech processing," *Perception & Psychophysics*, vol. 67, no. 2, pp. 224–238, 2005.

Word-level prosody of Digor Ossetic in a cross-dialectal perspective

Varvara Petrova

(Lomonosov Moscow State University, Russia)

Ossetic is an Indo-Iranian language spoken by approximately 340,000 people in Russia (primarily in the Republic of North Ossetia—Alania), as well as in Georgia and Turkey [1]. Two main dialects of Ossetic, Iron and Digor, vary significantly in terms of their lexical and phrasal prosodic marking. The stress in Iron falls on the first syllable if it contains a “strong” (long) vowel, and on the second syllable if both the first and the second syllables have “weak” (short) vowels as their nuclei [2].

Unlike Iron, Digor prosody has become a subject of disagreement among the researchers. While Isaev reported Digor primary stress to fall on the last strong vowel of the word [3], Thordarson’s notion of similarity between Iron and Digor systems with the exception of the fact that "the accent may be retracted to a syllable still farther back if the vowels of the preceding syllables are weak" implied that the accent is placed on the first strong vowel [4]. An acoustic study of Digor words containing the same vowels in isolated and utterance-final position showed vowel duration to increase gradually towards the end of the word [5]. This study aims at adding nuance to the existing knowledge of Digor prosodic patterns by considering more types of word structures and the influence of isolated vs utterance-medial environments. The audio data were collected in Vladikavkaz (North Ossetia) in 2024 and 2025 from 5 female speakers of Digor. The questionnaire consisted of two parts: bisyllabic words with all possible combinations of vowels, and trisyllabic words containing only mid vowels that represent all eight types of word structure in terms of phonological status of vowels (for example, tsardɜgas “alive” belongs to type SWS (strong-weak-strong)). Each stimulus was elicited three times in an isolated environment and three times in a carrier phrase dzaʁa [] ɜrttɜ χatti “Say [stimulus] three times”. The data were segmented in Praat [6] and statistically analyzed using MWU test. Data visualisation was performed using seaborn library for Python [7].

The discussion below will focus on non-isolated utterances of trisyllabic words; all other results will be discussed in the talk. The deciding factor in vowel duration is its phonological “strength”: strong vowels are consistently longer than weak ones regardless of their position. When vowels of the same phonological status are considered, the aforementioned lengthening towards the end of the word is visible in both strong (Figure 1a, $p < 0.01$) and weak vowels (Figure 1b, $p < 0.01$). In the sentences containing words with three strong vowels utterance-medially, this tendency is demonstrated by the second and the third ($p < 0.001$), but not the first and the second vowel ($p = 0.14$; Figure 1c). Given that the same vowels phonemes in Digor were reported to have shorter realisations before strong vowels than before weak ones [8], it can be assumed that the position before a strong vowel causes a durational decrease in the preceding vowel, but the position before an already shortened strong vowel does not have the same effect. Such quantitative marking of the first syllable appears to be completely neutralised by the utterance-final — but not word-final — increase in duration, which might explain the drastic differences in previous descriptions of Digor stress: although vowel duration generally tends to increase in the direction of the right edge of the word, a minority of words with non-last strong vowels (followed by either a weak or a shortened strong vowel) has what could be perceived as word-initial stress.

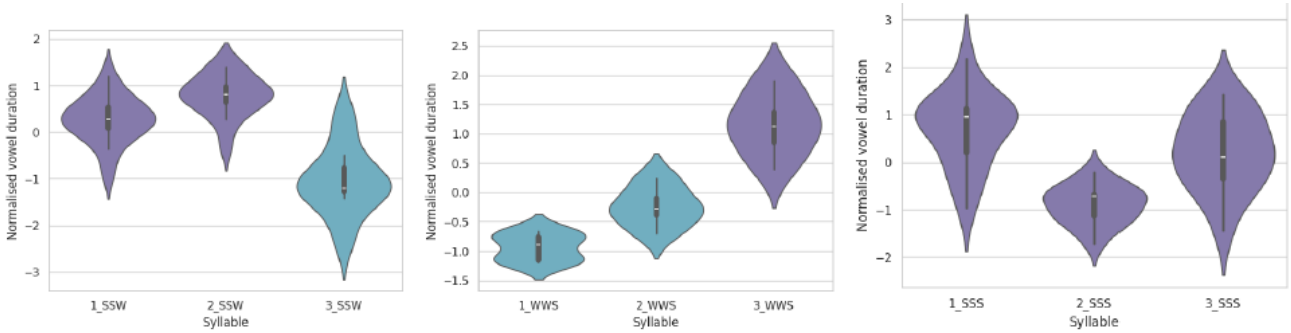


Figure 1. Z-normalised duration of vowels in the stimuli of types a) SSW, b) WWS, and c) SSS.

REFERENCES

[1] Rosstat. (2020). *The All-Russian census of 2020*. Retrieved from https://rosstat.gov.ru/vpn/2020/

[2] Abaev, V. I. (1959). *Grammatičeskij očerk osetinskogo jazyka* [A grammatical sketch of Ossetic] (p. 169). Ordžonikidze: Severo-Osetinskoe Knižnoe Izdatel'stvo.

[3] Isaev, M. I. (1966). *Digorskij dialekt osetinskogo jazyka* [The Digor dialect of the Ossetic language] (p. 226). Moscow: Akademia Nauk SSSR.

[4] Thordarson, Fridrik. (1989). Ossetic. In Schmitt (Ed.), 456–479. Wiesbaden: Reichert. Note: If Schmitt (1989) is a standalone edited volume, list as: Schmitt, J. (Ed.). (1989). Ossetic (pp. 456–479). Wiesbaden: Reichert.

[5] Petrova, V. (2024). Toward a description of Digor lexical prosody. In A. Botinis (Ed.), *ExLing 2024: Proceedings 15th International Conference of Experimental Linguistics* (pp. 101–104). Paris, France: ExLing Society. Retrieved from https://www.academia.edu/128896028/Toward_a_description_of_Digor_lexical_prosody

[6] Boersma, P., & Weenink, D. (2026). *Praat: Doing phonetics by computer* [Computer program] (Version 6.4.62). Retrieved March 13, 2026, from https://praat.org

[7] Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. https://doi.org/10.21105/joss.03021

[8] Sokolova, V. S. (1953). *Očerki po fonetike iranskikh jazykov* [Phonetical sketches of Iranian languages]. Moscow, Leningrad: Akademia Nauk SSSR, 1953.

16:40 Closing remarks and acknowledgments