Phonetics and Phonology in Europe 2019 satellite workshop:
Prominence between cognitive functions and linguistic structures (CoFLiS)

In the COFLIS workshop, we capitalise on several decades of research on prosodic prominence to unravel the key components of the notion of prominence. By exploring the contribution of the signal, of meaning and of linguistic structure to the definition of prominence, and by relating prominence to basic cognitive concepts such as chunking and attention, we aim to provide a renewed understanding of prominence. The workshop will feature four invited talks, covering the measurable, structural and functional components of prominence. Rather than share new experimental evidence, invited speakers will focus on the theoretical implications of their use of prominence in their research. Invited talks are complemented by 12 peer-reviewed poster presentations on prominence in phonetics and phonology.

Registration is free of charge and includes access to the event, lunch and coffee breaks. A limited number of non-presenting participants can register by sending an e-mail to *coflis.workshop@gmail.com* until Fri May 31st, 23:59 Niue time (UTC-11). Due to room space limitations, accepted registrations will be on first-come first-served basis.

Organising Committee: Francesco Cangemi, Stefan Baumann, Michelina Savino, Martine Grice
Contact: pape-coflis@uni-koeln.de - 0049 176 3172 6832 (FC) - 0039 080 5714707 (MS)

Workshop: 20.06.2019, 10:30 – 18:30
Address: Via Scipione Crisanzio 42, 70122 Bari BA
Department: Dipartimento di Scienze della Formazione, Psicologia, Comunicazione
Building: Palazzina Chiaia-Napolitano
Conference Room: "Aula don Tonino Bello" (ground floor)

Preliminary Programme:

10:30 – 11:00	Opening remarks
11:00 – 11:45	Fabio Tamburini
11:45 – 12:30	Mark Ellison
12:30 – 13:00	Poster slam
13:00 – 14:00	Lunch break
14:00 – 16:00	Poster session
16:00 – 16:30	Coffee break
16:30 – 17:15	Jason Bishop
17:15 – 18:00	Kai Vogeley
18:00 – **18:30**	Closing remarks

Invited lectures:
- **Jason Bishop** (City University of New York)
  Implicit prosody, explicit prosody, and individual differences in pragmatic skill
- **Mark Ellison** (Australian National University)
  Prosodic prominence and forward-modelling in production
- **Fabio Tamburini** (University of Bologna)
  Prosodic prominence from the speech processing perspective
- **Kai Vogeley** (University Hospital Cologne)
  Salience or prominence: are we focusing our attention based on physical features or communicative intentions?

Poster presentations:

1. Is phonetic prominence underlyingly multimodal?
   Gilbert Ambrazaitis, David House
2. Non-rhythmic acoustic correlates of prominence clash:
   A case study from Italian prepared and unprepared speech
   Francesco Burroni, Sam Tilsen
3. Different cognitive sources lead to different prominence in child language?
   Aoju Chen, Anna Sara Hexeberg Romøren
4. Prosodic prominence in pronoun resolution: A cross-linguistic comparison between Italian/Swed.
   Chiara Gargiulo, Mechtild Tronnier, Petra Bernardini
5. Prominence as a sensory constraint in multi-modal word processing
   Inga Grantyn
6. Exploring the phonetic & phonological contribution to prominence in sentence-initial *I think*
   Daniela Kolbe-Hanna, Clare Maas, Laura Panne
7. Underlying mechanisms in the perception of metrical prominence:
   The role of occurrence frequency of different pitch accent types
   Sophie Kutscheid, Katharina Zahner, Adrian Leemann, Bettina Braun
8. Focus-induced articulatory prominence on velum actions in nasal geminates
   Miran Oh, Dani Byrd
9. The development of lexical stress in young Italian children
   Francesco Olivucci, Mario Vayra, Cinzia Avesani, Claudio Zmarich
10. Perception of non-native prosody: the case of prosodic prominence
    Valentina Schettino, Francesco Cutugno, Antonio Origlia
11. Prosodic prominence versus frequency effects on the acquisition of CCV branching onsets in BP
    Andressa Toni
12. Prosodic distance: Using variation to weight measures of prominence contrast
    Laurence White, Claire Delle Luche, Caroline Floccia

Instructions for poster presenters:
For the poster session, please print your poster in A1 portrait format (60cm width x 85cm height), and make sure to hang it on the board with your poster number during lunch break (13:00 – 14:00).
For the poster slam, please prepare a 1 minute and 30 seconds presentation for your poster, and send us a single image (a .pdf slide in landscape format) by Sat June 8th, 23:59 Niue time (UTC-11).

The workshop site is at few minutes walking distance from the Bari Central Station, which is also the place where airport bus, shuttlebus and local train stop. Directions to Bari Central Station:

From **Bari Airport**, by Local Railway ("Ferrotramviaria"): The local railway station can be reached on foot directly from the Terminal. Tickets can be bought at vending machines or an information desk (5 €), journey lasts about 15 minutes.
From Bari Airport, by Shuttlebus ("Tempesta Autoservizi"): The shuttlebus stops right outside the Terminal. Tickets can be bought on the bus (4 €), journey lasts about 25 minutes.
From Bari Airport, Local Bus service (AMTAB), bus line no. 16 (Viale Ferrari Aeroporto - P.zza Moro): Tickets can be bought on the bus (1.50 €), journey lasts about 40 minutes.

From **Lecce**, by National Railway (Trenitalia): Ticket (Regional trains) costs around 10 €, journey lasts about 2 hours, trains run around every 1.5 hour.
From Lecce, by car: take the National Route no. 16 ("Strada Statale 16 "Adriatica"), when approaching Bari follow indications to Bari "Stazione Centrale". Journey lasts about 1 hour and a half. Car parking just in front of the workshop venue (underground parking GESTIPARK, Piazza Cesare Battisti, entrance from via Scipione Crisanzio) Fare: 1.60 € per hour.

Implicit prosody, explicit prosody, and individual differences in pragmatic skill

Jason Bishop
City University of New York

Contemporary models of sentence-level prosody assume that an utterance's prosodic form is shaped by a number of factors—factors related to meaning; factors related to phonological grammar; factors related to speaking contexts; and factors related to processing [e.g. 1,2]. Focusing primarily on factors related to processing, this talk will consider the role that individual differences play in how listeners/comprehenders make use of prosodic prominence when decoding spoken utterances and written text. My primary focus will be on the role of what has been called "pragmatic skill"—that is, the extent to which an individual approaches information in a context-sensitive way, and relies on pragmatic inference and "theory of mind" rather than strictly logical form. Pragmatic skill can be measured in healthy/neurotypical populations using various instruments [e.g. 3,4,5] and these measures have been shown to correlate with performance on both online and offline sentence comprehension tasks [6,7,8,9]. The basic hypothesis that I explore in this talk is that this correlation reflects not the parsing of syntactic structure or in semantic/pragmatic interpretation per se, but rather the contribution that prominence patterns make to these stages of sentence comprehension. This interpretation has important implications for models of prosody's role in language processing, as well as for the use of "dimensional" approaches to the study of language impairments in special populations [e.g. 10,11].

[1] Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: what is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B*, 369: 20130395.

[2] Shattuck-Hufnagel S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *J. of Psycholinguistic Research*, 25, 193–247.

[3] Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised : a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. of Child Psychology and Psychiatry*, *42*, 241–251.

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

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

[6] Nieuwland, M., Ditman, T., & Kuperberg, G. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *J. of Memory and Language*, *63*, 324–346.

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

[8] Jun, S.-A., & Bishop, J. (2015). Priming implicit prosody: Prosodic boundaries and individual differences. *Language and Speech*, *58*, 459–473.

[9] Bishop, J. (2017). Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience*, *32*, 236–253.

[10] Miller, J., Lynam, D., Widiger, T., Leukefeld, C. (2001). Personality disorders as extreme variants of common personality dimensions: Can the five-factor model adequately represent psychopathy?. *J. of Personality* 69, 253–276.

[11] Wright, A., Krueger, R., Hobbs, M., Markon, K., Eaton, N., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. *J. of Abnormal Psychology* 122, 281–294.

Prosodic Prominence and Language Forward-Modelling

T. Mark Ellison
Australian National University
Universität zu Köln

m.ellison@anu.edu.au

Abstract

This talk introduces a general account of *prosodic prominence* as a response to likelihood, and relates it to prediction-based models of language cognition.

Correlation between the likelihood of a form and its prosodic prominence has been explored in a number of previous studies such as the following. Aylett & Turk (2004) suggest that prominence and duration function to keep the information flow rate as consistent as possible: less likely items, being more informative, attract more prominence and are expressed more slowly. Baker & Bradlow (2009) argue that the *a priori* frequency of a word and whether or not it has been previously mentioned in the discourse - both impacting the likelihood of the word to occur - affect the articulatory prominence of its realisation. Kakouros & Räsänen (2014) find that unlikely F0 contours make good predictors of human prominence judgements.

Recent years have seen a growing emphasis on prediction as a fundamental principle underlying cognitive processing. Pickering & Garrod (2013) propose that speakers construct *forward models* of their productions before they produce them, while listeners covertly imitate the production of the speech they hear, then build forward models from those covert imitations. Crucially, the job of the forward models is to make predictions. These predictions are constantly matched to sensory input and discrepancies used to adjust the forward model.

This forward modelling account of interactive speech draws on earlier work by Wolpert (1997) and Davidson & Wolpert (2005) on active perception. The predictive account of active perception can, in turn, be cast as a special case of the *Free Energy Principle* proposed by Friston (2009).

This talk proposes a Bayesian model of prosodic prominence, and situates it in the context of forward-planning models, and more generally as a special case of the free energy model of cognition.

Keywords

prosodic prominence, predictability, Bayes, forward planning, free energy principle

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*(1), 31–56.

Baker, R. E., & Bradlow, A. R. (2009, December). Variability in Word Duration as a Function of Probability, Speech Style, and Prosody. *Language and Speech*, *52*(4), 391–413. doi: 10.1177/0023830909336575

Davidson, P. R., & Wolpert, D. M. (2005). Widespread access to predictive models in the motor system: a short review. *Journal of neural engineering*, *2*(3), S313.

Friston, K. (2009, July). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. doi: 10.1016/j.tics.2009.04.005

Kakouros, S., & Rasanen, O. J. (2014). Statistical unpredictability of F0 trajectories as a cue to sentence stress. In (Vol. 36).

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347. doi: 10.1017/S0140525X12001495

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in cognitive sciences*, *1*(6), 209–216.

Prosodic prominence from the speech processing perspective

Fabio Tamburini
University of Bologna

Like Kerberos, the three-headed dog that guards the gates of the Underworld, Prosodic prominence can be seen and studied from three different perspectives: (a) a perceptual/cognitive side, (b) a linguistic/functional view and (c) an acoustic/physical perspective.

In literature we can find several definitions of prominence taking a specific point of view, but, except for the perceptual side for which there is a general consensus to define prosodic prominence as the "perceptual salience of a linguistic unit relative to its environment", we are far from having such consensus on how it is measured subjectively and how it relates to objectively measurable acoustic events or linguistic structures such as lexical and sentence stress or prosodic focus.

In this talk, assuming a purely acoustic point of view, we will try to cast some light around the prominence phenomenon introducing the definitions and procedures that this perspective implies, with the specific goal to build automatic systems for the detection of prosodic prominence in the utterances from a multilingual point of view. Taking this perspective, we will discuss (a) which is the most appropriate kind of unit guiding the acoustic measures on the time axe, (b) the delicate dichotomy between discrete or continuous views on the prominence phenomenon and (c) the various ways of approaching the detection problem from the computational side.

Taking this narrowed approach is not completely in line with the recent effort of some scholars to unify the different views on prosodic prominence by looking at the entire "dog" instead of facing only with one of its three "heads", but we hope that, by clarifying some points on one of the three approaches, we will bring some contribution to the whole enterprise.

Salience or prominence: are we focusing our attention based on physical features or communicative intentions?

Kai Vogeley, Juliane Zimmermann

Department of Psychiatry, University Hospital Cologne, Germany & Institute of Neuroscience and Medicine – Cognitive Neurosciences (INM-3), Research Center Juelich, Germany

The concepts of prominence in the tradition of phonetic linguistics and salience in the tradition of visual psychology and neuroscience refer to strikingly similar properties, namely to the phenomenon of "standing out" in relation to a background (e.g. in terms of prosodic features [1]) in the domain of linguistics and to the phenomenon of "importance" or "noticeability" (e.g. colour, brightness [2]) in the domain of psychology and neuroscience. Clarifying similarities and differences of the concepts of prominence and salience is not only useful for the exchange between different research traditions and the stimulation of interdisciplinary research, but also for a better understanding of concepts within disciplines by integrating insights from neighbouring research domains. We propose that prominence and salience share essential properties and possibly also common processing mechanisms. Both concepts, salience and prominence, appear to include at least a stimulus component ("bottom-up"; e.g. physical features) and a perceiver component ("top-down"; e.g. perceiver´s expectation), both of which have been postulated for linguistic salience [3]. Linguistic salience, moreover, arises from a context component that can be defined along a situational, social, and linguistic dimension [4], which arguably also applies to prominence. This is in line with a recent proposal identifying three different perspectives on prosodic prominence, namely physical (physical correlates, e.g. change in fundamental frequency), functional (what prominence wants to convey, e.g. information status), and cognitive perspectives (how prominence is processed), the latter of the three depending on the former two [5]. In a similar fashion acoustic, structural and perceived prominence can be differentiated [6]. It appears that stimulus properties, the perceiver´s disposition and possibly the context are integrated via the general processing mechanism of attention attribution that serves a number of different functions for the perceiver (alerting, reorienting, availability, accessibility). This comparative approach allows us to refine and enrich the concepts of prominence and salience by integrating differentiations and insights from an "outside" perspective of another research tradition.

[1] Ladd, D. Robert (2008). *Intonational phonology*. 2nd edition. Cambridge: Cambridge University Press.

[2] Uddin, L. Q. (2016). What is Salience? In *Salience Network of the Human Brain*, Academic Press, 1-4. Retrieved from https://www.elsevier.com/books/salience-network-of-the-human-brain/uddin/978-0-12-804593-0

[3] Zarcone, A., van Schijndel, M., Vogels, J., & Demberg, V. (2016). Salience and Attention in Surprisal-Based Accounts of Language Processing. *Frontiers in Psychology*, 7, 844. https://doi.org/10.3389/fpsyg.2016.00844

[4] Schmid, H.-J., & Günther, F. (2016). Toward a Unified Socio-Cognitive Framework for Salience in Language. Frontiers in Psychology, 7. https://doi.org/10.3389/fpsyg.2016.01110

[5] Wagner, P., Origlia, A., Avezani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., … Vainio, M. (2015). Different parts of the same elephant: a roadmap to disentangle and connect different perspectives on prosodic prominence. *Proceedings of the 18th International Congress of Phonetic Sciences*. Retrieved from https://dial.uclouvain.be/pr/boreal/object/boreal:161827

[6] Bishop, J. (2016). Individual differences in top-down and bottom-up prominence perception (pp. 668–672). *Proceedings of the 9th International Conference on Speech Prosody*. https://doi.org/10.21437/SpeechProsody.2016-137

Is phonetic prominence underlyingly multimodal?

Gilbert Ambrazaitis[1] and David House[2]
[1]*Linnæus University, Växjö*, [2]*KTH (Royal Institute of Technology), Stockholm, Sweden*

Stating that spoken language is multimodal – involving an auditory and a visual mode – can today almost be considered trivial [1-6], and the same holds for the phonetic – or better: the physical – dimensions of prosodic prominence, involving both acoustic and kinetic parameters, in production and in perception [7-14]. The purpose of this contribution is to discuss the nature of this multimodality, based on previous and recent empirical evidence. For, despite a growing consensus on the multimodality of speech and prominence, we still lack a common understanding of the relation between the modes: Are kinetic/visual correlates of prominence optional? On the one hand, we still lack sufficient empirical evidence. On the other hand, we believe that we can, on the basis of the evidence we have today, already argue for a hypothesis of an underlying multimodality of prominence: That is, even if not always surfacing, we argue that it is reasonable to assume that in the execution of prominence, both kinetic and acoustic expressions are generally targeted.

To this end, we will review some previous and recent work on the co-occurrence of auditory and visual cues to prominence. For instance, it has been shown that visible articulatory movements such as jaw openings and lip articulations are subject to variation due to prominence, besides their primary function of encoding phonemes (e.g., [7-8]). In addition, a growing line of research focuses on the role of gestures in prominence production and perception. For instance, pitch accents are frequently accompanied by so-called beat gestures, typically produced by the hands, the head or the eyebrows (e.g., [9-11], [15-19]).

Furthermore, some attempts have been made to test whether the acoustic and the kinetic dimensions of prominence would "influence each other" in the sense that accompanying beat gestures would "affect" the phonetic realization of pitch accents [20-21]. Our own recent study [22] does in fact provide evidence for slightly larger pitch ranges found in connection with head movements, and even larger ones when also eyebrow movements are added.

These results, we argue, suggest a cumulative relation between acoustic and kinetic prominence expressions in speech production: the more of the one, the more of the other we tend to get. However, we will argue that the assumption that gesture "affects" pitch (or why not vice versa) only describes the relation (in an inappropriate manner), rather than explaining it. It is more reasonable to assume that a relatively large pitch range on the one hand, and a relatively strong involvement of gestures on the other hand, are two parallel responses to the communicative need to produce a high prominence level.

To make this explanation work out, we hypothesize that prominence is underlyingly multimodal, just as speech in general is. After all, spoken language is produced through bodily (incl. articulatory) movements, where some are both visible and audible (certain articulatory movements), some are audible only (many articulatory movements), while some are only visible (gestures). All three types of movements have been shown to take part in the production of prominence [7-19]. The simplest way to model the involvement of these different channels would be to assume that they all are controlled by the same force, i.e. the need to make a unit of speech more prominent. However, prominence need not always be multimodal at the surface, which is explainable by the fact that all prominence-related channels, both acoustic and kinetic, are multifunctional: Hence, the degree to which they are exploited for prominence in a given situation may depend on the balance between the communicative needs to signal prominence vs. other functions; it may also be language-specific and possibly individual.

At the workshop, we will further discuss the hypothesis of an underlying multimodality in the implementation of prosodic prominence, and some implications of this hypothesis.

[1] McGurk, H., & MacDonald, J.1976. Hearing lips and seeing voices. *Nature* 264, 746-748.
[2] Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press.
[3] McNeill, D. 2005. Gesture and Thought. University of Chicago Press.
[4] Biau, E., & Soto-Faraco, S. 2013. Beat gestures modulate auditory integration in speech perception. *Brain Lang.* 124, 143–152.
[5] Wagner, P., Malisz, Z., & Kopp, S. 2014. Gesture and speech in interaction: An overview, *Speech Commun.* 57, 209-232.
[6] Graziano, M., & Gullberg, M. 2018. When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Front. Psychol.,* 9:879.
[7] Beskow, J., Granström, B., & House, D. 2006. Visual correlates to prominence in several expressive modes. In: *Proc. of Interspeech 2006*, (Pittsburg, PA), 1272–1275.
[8] Hermes, A., Becker, J., Mücke, D., Baumann, S., & Grice, M. 2008. Articulatory Gestures and Focus Marking in German. In: *Proc. of Speech Prosody 2008*, (Campinas, Brazil), 457–460.
[9] Swerts, M., & Krahmer, E. 2010. Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *J. Phonet.* 38, 197–206.
[10] Esteve-Gibert, N., & Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech Lang. Hear. Res.* 56 (3), 850–864.
[11] Ambrazaitis, G., & House, D. 2017. Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Commun.* 95, 100-113.
[12] Dohen, M., & Loevenbruck, H. 2009. Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang. Speech* 52, 177–206.
[13] Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. 2015. Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *J. Phonet.* 49 (1), 41–54.
[14] Scarborough, R., Keating, P., Mattys, S. L., Cho, T., Alwan, A., 2009. Optical phonetics and visual perception of lexical and phrasal stress in English. *Lang. Speech* 52, 135–175.
[15] Jannedy, S., & Mendoza-Denton, N. 2005. Structuring information through gesture and intonation. *Interdiscip. Stud. Inf. Struct.* 3, 199–244.
[16] Leonard, T., & Cummins, F. 2011. The temporal relation between beat gestures and speech. *Lang. Cogn. Processes* 26, 1457–1471.
[17] Loehr, D. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *J. Assoc. Lab. Phonol.* 3, 71–89.
[18] Rusiewicz, H.L. 2010. *The Role of Prosodic Stress and Speech Perturbation on the Temporal Synchronization of Speech and Deictic Gestures* (Doctoral dissertation). University of Pittsburgh.
[19] Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S. 2004. The timing of speech-accompanied gestures with respect to prosody. In: *Proc. of From Sound to Sense,* (MIT, Cambridge, MA), 97–102.
[20] Krahmer, E., & Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57, 396–414.
[21] Roustan, B., & Dohen, M. 2010. Co-Production of Contrastive Prosodic Focus and Manual Gestures: Temporal Coordination and Effects on the Acoustic and Articulatory correlates of Focus. In: *Proc. of Speech Prosody 2010*, (Chicago).
[22] Ambrazaitis, G., & House, D. 2019. Multimodality in prominence production and its sensitivity for lexical prosody. Paper at *3rd Phonetics and Phonology in Europe conference (PaPE 2019)*, Lecce, Italy.

Non-rhythmic acoustic correlates of prominence clash:
A case study from Italian prepared and unprepared speech

Francesco Burroni[1] and Sam Tilsen[1]
[1]*Cornell University*

A controversial area of the phonetics/phonology interface encompasses cases of so-called prominence clash. On one hand, in the phonological literature, it is widely assumed that clash induces rhythmic readjustments, either in the form of prominence shift, a.k.a the Rhythm Rule ([1],[2],[3]), or of prominence deletion ([4],[5]), summarized in Fig.1. On the other hand, phonetic studies of clash environments have failed to observe acoustic correlates of prominence shift ([6],[7],[8],[9]) and do not provide a coherent picture compatible with a deletion account ([10],[7] vs [9]). An extra layer of complication is added by the uncertainty surrounding the role played by extra-linguistic factors, acknowledged by some researchers ([1],[9]), but dismissed by others ([3]). We present experimental data from Italian, a language for which much of the phonological literature assumes rhythmic readjustments ([3],[4],[11]), even though such adjustments have never been probed phonetically. We show that in clash environments vowel and syllable durations – the main correlate of stress in this language ([12]) - do not show evidence of stress shift; indeed, our findings show that durations *increase* in a clash context, suggesting enhancement of prominence rather than shift or deletion.

Methods. 10 native speakers of Standard Italian balanced for gender and geographic provenance produced 36 unique noun phrases elicited from visual stimuli representing a three-word noun phrase consisting of (1) a numeral, (2) a target noun with final prominence (*caffè* 'coffee', *città* 'city', and *colibrì* 'hummingbird') and (3) a color term which had either initial stress (to create a clash) and or non-initial stress (to create a clash baseline), this is summarized in Fig.2. Each of the 36 unique combinations were repeated for 10 blocks, alternating between odd blocks, where speakers could mentally rehearse the utterance before producing it, and even blocks, where they had to utter the intended noun phrase as fast as possible. In this way "preparation", representing an extralinguistic factor, was manipulated. The data were automatically segmented by a forced-aligner trained on the basis of 36 randomly selected manually-segmented unique trials from each subject

Results. We found no statistically significant difference in the duration of initial vowels in the same words appearing in clash vs non clash environments, $F(1,2758)=1.78$ $p > .05$. On the other hand, we found that durations of the final vowel in the target words were significantly higher in clash vs non clash, both in terms of raw duration and of duration ratio of the final vowel to the initial vowel, $F(1,2746)=213.15$ $p < .001$ (Fig. 3). No appreciable effect of preparation on duration in clash environments was detected, $F(1,1364)=0.72$ $p > .05$. The same statically findings hold true, at a 0.01 significance level, if the entire syllable is measured rather than vowel alone. Preliminary analyses of RMS intensity and F0 in the target word also fail to show any evidence of stress shift ([6]).

Conclusions. On the basis of the experimental data presented, we conclude that current eurhythmic approaches to prominence clash are not empirically motivated: neither a prominence shift nor prominence deletion accounts for our observation of increased duration in the stressed syllable of the target word. In sum, current approaches to the effects of clash on prominence need to be re-examined and conclusions based solely on shift intuitions should be called into question .

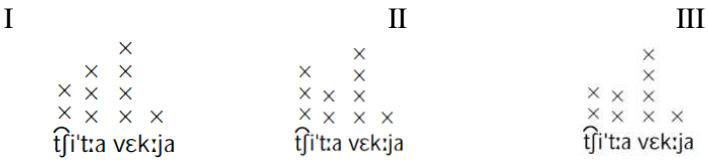


Fig. 1 Metrical grid representation of prominence clash (I) in the Italian NP *città vecchia* 'old town' undergoing prominence shift (II) or prominence deletion (III).

Numeral	Target Noun	Color Adjective
['due]	[t͡ʃitt'a]	['verdi] (C)
['nɔve]	[koli'bri]	['neri] (C)
['mille]	[ka'ffɛ]	[bor'do] (NC)
		[ma'rroni] (NC)

Fig. 2 Stimuli matrix, with initial vowel (where shifting prominence would be expected to dock) in red, first clash vowel in blue (where word-level prominence was originally expected) and phrasal level prominence (a.k.a. pitch accent) in green.

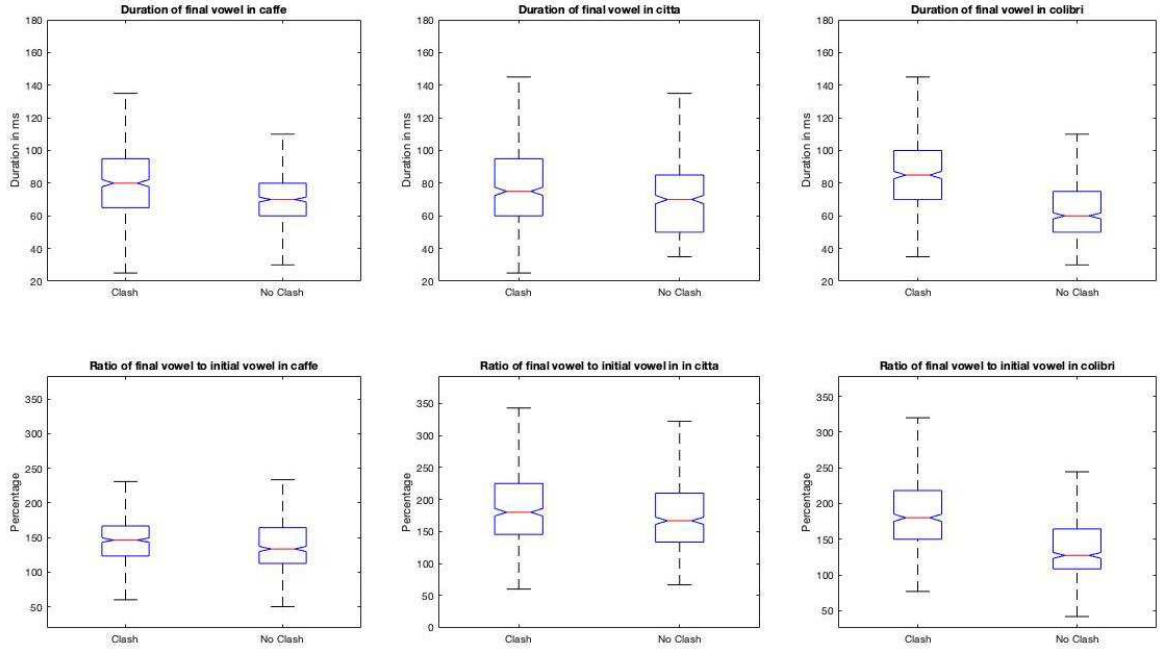


Fig. 3 Duration of final vowel (row 1) and duration ratio of final vowel/initial vowel (row 2) are significantly different for each target word ($p < 0.001$).

[1] Liberman, M and Prince, A. On Stress and Linguistic Rhythm. Linguistic Inquiry, 8(2):249–336, 1977.

[2] Hayes, B. The Phonology of Rhythmin in English. Linguistic Inquiry, 15(1):33–74, 1984.

[3] Nespor, M and Vogel, I. Clash avoidance in italian. Linguistic Inquiry, 10(3):467–482, 1979.

[4] Nespor, M and Vogel, I. On clashes and lapses. Phonology, 6(1):69–116, 1989.

[5] Gussenhoven, C. The English Rhythm Rule as an Accent Deletion Rule. Phonology, 8(1):1–35, 1991.

[6] Cooper, W. E. and Eady, S. J. Metrical phonology in speech production. *Journal of Memory and Language,* 25(3: 369-384, 1986.

[7] Vogel, I *et al.* The phonology and phonetics of the rhythm rule. In B. Connell and A. Arvaniti, editors, *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, pages 111–127. CUP, Cambridge, 1995.

[8] E. Grabe and Warren, P. Stress shift: do speakers do it or do listeners hear it. In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, pages 95–110. CUP, 1995.

[9] Tilsen, S. Utterance preparation and stress clash: Planning prosodic alternations. In *Speech Production and Perception: Planning and dynamics*, pages 119–156. Peter Lang Verlag, Frankfurt am Main, 2012.

[10] Horne, M. Empirical evidence for a deletion formulation of the rhythm rule in english. Linguistics, 28(5):959–982, 1990.

[11] Krämer, M. The phonology of Italian. Oxford University Press, Oxford, 2009.

[12] A. Eriksson, P. M. Bertinetto, M. Heldner, R. Nodari, and G. Lenoci. The acoustics of lexical stress in Italian as a function of stress level and speaking style. In *INTERSPEECH 2016*, San Francisco, USA, September 8–12, 2016, pages 1059–1063, 2016.

Different cognitive sources lead to different prominence in child language?

Aoju Chen[1] and Anna Sara Hexeberg Romoren[2]
[1]*Utrecht University,* [2]*Oslo Metropolitan University*
aoju.chen@uu.nl; romo@oslomet.no

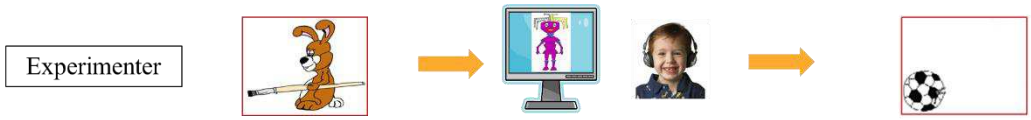
Past work from different theoretical perspectives has shown that both informational importance and predictability influence prosodic prominence in speech production. Watson and colleagues [1] have found that these different cognitive sources of prominence are associated with different aspects of the speech signal: higher informational importance is linked with higher pitch and higher predictability is linked with shorter word duration.

Children's use of prosodic prominence has mostly been studied in the tradition of informational importance, concerning notions such as referential newness, focus and contrast [see 2, 3 for reviews]. Existent work has shown that children from a West Germanic language can use accentuation to encode referential newness, focus and contrast by the age of four or five, although their choice of pitch accent type is not adultlike until the age of seven or eight. Accentuation is usually determined by trained annotators, who base their perception on multiple cues, including pitch, duration and intensity. Consequently, it is not clear whether children already show adult-like association between informational importance and pitch at the age of four or five. Also, it is not known how predictability influences prosodic prominence independent of informational importance in four- to five-year-olds' speech production.

To shed light on how informational importance and predictability influence prosodic prominence in children's production, we analysed a subset of the data reported in [4] and [5]. It consisted of object nouns of SVO sentences produced as responses to either who-questions or what-questions by Dutch-speaking four- to five year-olds in two versions of a picture-matching game. The what-questions rendered the object nouns focal; the who-questions the object nouns non-focal. In both versions of the picture-matching game, the child was asked to help the experimenter to find the matching picture for each of the experimenter's picture, which contained incomplete information, by answering the experimenter's questions about the pictures. The two versions of the game differed in the predictability of the focal information. In the robot-version, having no visual access to or prior knowledge of the focal information, the child received the answer to the experimenter's question from a robot, who spoke in abnormal prosody, and reproduced the answer in his own prosody (Figure 1). The focal information was thus not predictable to the child, different from the non-focal information. In the robot-free version [6], the child had visual access to and prior knowledge of both focal and non-focal information from the beginning of each trial, which led the focal information to be similarly predictable to non-focal information (Figure 2).

In both [4] and [5], the four- to five-year-olds (N = 12 in the former, N = 10 in the latter) accented the object nouns substantially more frequently in the focal condition than in the non-focal condition (92% vs. 66% in [4]; 94% vs. 52% in [5]). If, like adults, children associated informational prominence with pitch and predictability with duration, then the object nouns should be produced with higher pitch maximum or pitch span in the focal condition than in the non-focal condition to a similar degree in both studies (prediction 1). Further, the object nouns should be produced with increased duration in the focal condition than in the non-focal condition in the data obtained from the robot version of the game [4] but not in the data obtained from the robot-free version of the game [5] (prediction 2).

Initial results showed that the focal object nouns and their non-focal counterparts indeed differed in pitch maximum and span but not in word duration in [5]. In the presentation, we will report the full analysis on the four- to five-year-olds' data and follow-up analysis on older children's data (if the four- to five-year-olds are not adultlike), and discuss the implications of a cognitive approach to prosodic prominence for research on prosodic development.

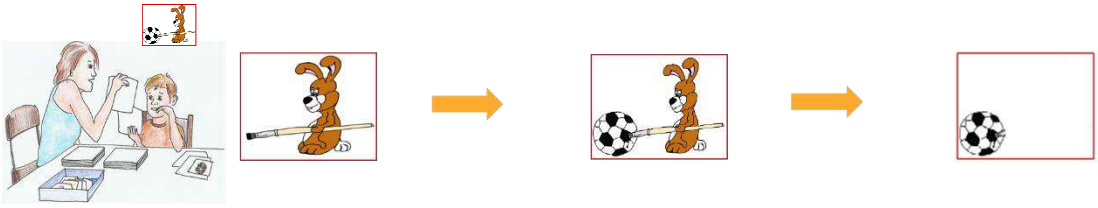


Experimenter (showing a picture to the child): Look! The rabbit! The brush is here too. It seems that the rabbit paints something. What does the rabbit paint?

Child clicks on the robot, who says 'The rabbit paints the football' in abnormal prosody, and reproduces the answer in his own prosody.

The experimenter looks for the matching picture in a box and hands both pictures to the child.

Figure 1. A step-by-step illustration of the robot-version of the picture matching game [4].



Experimenter (showing a picture to the child): Look! The rabbit! The brush is here too. It seems that the rabbit paints something. What does the rabbit paint?

Child (looking at his own picture): The rabbit paints the football.

The experimenter looks for the matching picture in a box and hands both pictures to the child.

Figure 2. A step-by-step illustration of the robot-version of the picture matching game [6].

[1] Watson, D. G. (2010). The Many Roads to Prominence: Understanding Emphasis in Conversation. *The Psychology of Learning and Motivation*, *52*, 163–183.

[2] Chen, A. (2015). Children's use of intonation in reference and the role of input. In L. Serratrice & S. Allen (eds.) *The acquisition of reference* (pp. 83-104). Amsterdam: John Benjamins.

[3] Chen, A. (2017). Get the focus right across languages: Acquisition of prosodic focus-marking in production. In P. Prieto & Esteve-Gibert, N. (eds) *Prosodic development in first language acquisition* (pp. 295-314). Amsterdam: John Benjamins.

[4] Chen, A. (2011). Tuning information structure: intonational realisation of topic and focus in child Dutch. *Journal of Child Language, 38, 1055-1083.*

[5] Romøren, A. S. H. (2016). Hunting highs and lows: The acquisition of prosodic focus marking in Swedish and Dutch. PhD Dissertation. LOT: Utrecht University.

[6] Chen, A. (2014). Production-comprehension (A)Symmetry: individual differences in the acquisition of prosody focus-marking. In *Proceedings of the 7th International Conference on Speech Prosody*.

Prosodic prominence in pronoun resolution: a cross-linguistic comparison between Italian and Swedish

Chiara Gargiulo[1], Mechtild Tronnier[2] & Petra Bernardini[3]
Lund University

[1] chiara.gargiulo@rom.lu.se; [2] mechtild.tronnier@ling.lu.se; [3] petra.bernardini@rom.lu.se

Abstract

In the present study, we explore the cross-linguistic differences between Italian and Swedish in the use of prosodic cues to disambiguate pronoun coreference. A general definition of prominence is adopted to compare the two languages, i.e., "a linguistic entity is prosodically prominent when it stands out from its environment by virtue of its prosodic characteristics" (Terken & Hermes, 2000). We focus on three acoustic measures of the pronoun, i.e., relative length, average F0 range and relative intensity. A production task was completed by 28 Italian speakers and 28 Swedish speakers in their native language: they vocally disambiguated overt pronouns (e.g., (1b) and (2b)) to either the subject or object referent of a preceding sentence (e.g., (1a) and (2a)).

(1) a. Riccardo/ Diego lavorava in una clinica privata
'Riccardo/ Diego worked in a private clinic.'
b. «Riccardo ha conosciuto Diego quando lui lavorava in una clinica privata.»
'Riccardo met Diego when he was working in a private clinic.'

(2) a. Gustav/ Martin arbetade på en privatklinik.
'Gustav/ Martin worked in a private clinic.'
b. «Gustav lärde känna Martin när han arbetade på en privatklinik.»
'Gustav met Martin when he was working in a private clinic.'

Our main prediction is that Italian speakers use prosody differently from Swedish speakers, reflecting the diverse underlying properties of pronoun resolution in the two languages. In fact, Italian overt pronouns usually co-refer to object antecedents and null pronouns to subject antecedents ("PAS" - Carminati, 2002), while Swedish overt pronouns leave a state of ambiguity with respect to antecedent assignment. To confirm this difference, the participants completed also a control interpretation task. Concerning the production task, we hypothesize that the Italian speakers produce more prominent pronouns in (1b) when the preceding sentence (e.g., (1a)) contains a subject referent (i.e., the unpredictable referent) instead of an object referent (i.e., the predictable referent). On the other hand, we hypothesize that the Swedish speakers avoid to produce more prominent pronouns in (2b) in any of the two conditions, i.e., subject and object referent (e.g., (2a)), reflecting the intrinsic ambiguity of pronoun resolution in this language.

The results show that the Italian speakers produced more prominent pronouns (in terms of relative length, average F0 range and relative intensity) with subject instead of object referents (Figure 1). In contrast, the Swedish speakers unexpectedly produced more prominent pronouns (in terms of relative length and average F0 range) with object instead of subject referents (Figure 1). Interestingly, the relative intensity was not significant in the Swedish group (see Baumann & Roth, 2014). In the control interpretation task, the Italian group assigned overt pronouns to objects (and null pronouns to subjects), confirming the PAS, while the Swedish speakers unexpectedly selected more subject than object referents. These findings suggest that prosody affects the coreference of overt pronouns in opposite ways in Italian and Swedish, and that this reflects opposite underlying patterns in pronoun resolution (Figure 2).

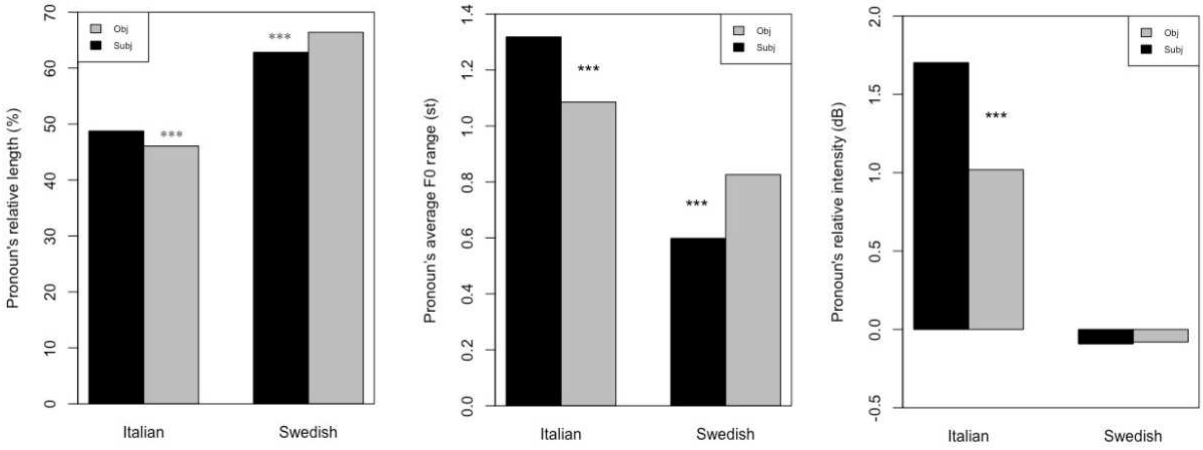


Figure 1. Mean of relative length, average F0 range and relative intensity of the pronoun for subject and object referents, in Italian and Swedish.

Object of analysis	Results				Predictions				Level of analysis	Experimental test
	Italian		Swedish		Italian		Swedish			
	sub	obj	sub	obj	sub	obj	sub	obj		
Inter-clausal pause & pronoun's prominence	>	<	<	>	>	<	=	=	Prosody	Production task
Pronoun's coreference	–	+	+	–	–	+	=	=	Syntax-pragmatics	Control interpretation task

Figure 2. Summary of our results on overt pronoun resolution, compared to our predictions[4].

References

[1] Baumann, S., & Roth, A. 2014. Prominence and coreference – On the perceptual relevance of F0 movement, duration and intensity. *Proceedings of Speech Prosody*. 227–231. doi: 10.18653/v1/W17-4610.

[2] Carminati, M. N. 2002. *The processing of Italian subject pronouns*. Amherst, MA: University of Massachusetts at Amherst dissertation.

[3] Terken, J., & Hermes, D. 2000. The Perception of Prosodic Prominence. In Horne, M. (Ed.), *Prosody: Theory and Experiment. Text, Speech and Language Technology*, 14, Springer, Dordrecht. doi: https://doi.org/10.1007/978-94-015-9413-4_5.

[4] Concerning the prosodic level of analysis, the symbol ">" stands for "more prominent", the symbol "<" stands for "less prominent" while "=" stands for "equally prominent". Concerning the syntactic-pragmatic level of analysis, the symbol "+" stands for "predictable", the symbol "–" for "unpredictable" and the symbol "=" for "equally predictable".

Prominence as a sensory constraint in multi-modal word processing

Inga Grantyn
Humboldt University Berlin, Faculty of Language, Literature and Humanities,
Department of Romance Literatures and Linguistics,
Unter den Linden 6, 10099 Berlin, Germany

The assumed arbitrary relationship between a linguistic sign and its encoded concept raises the question of how this hypothesis could be verified (or falsified) by empirical methods (Ohala 1994, Blasi 2016). Roman Jakobson, one of the first phoneticians who empirically explored possible non-arbitrary sound-concept relations, demonstrated that some acoustic forms conserve highly iconic features whereas others loosen their tight associations between sound and meaning during diachronic evolution. One of his well known examples is the contrast pair *day-night,* in Latin [ˈdi.es]-[ˈnoks] (simplified IPA transcription) which, according to many recipients displays congruent phonetic features, i.e. 'dies' sounds bright and 'nox' - dark. Interestingly, when it came to sound change, these sensorially adequate Latin items diverted along different lines: in Russian the words [ˈdjenʲ] and [ˈnoʧʲ] retained the initial associations with bright and dark, whereas the sound change in French resulted in sensorially mismatching forms: [ˈʒuʁ] (*day*) is perceived as "dark" and [ˈnɥi] (*night*) - as "bright" (Jakobson 1979). This effect is usually attributed to the vowel quality as reflected by the formant structure, disregarding syllable prominence as an additional criterion. The question arises, whether the Classic Latin word for *day*, historically reconstructed as [ˈdi.es], would be perceived as dark instead of bright, by putting the lexical stress on the second syllable instead on the first. Accordingly, the French word for *night* [ˈnɥi] could be perceived as dark instead of bright, if the approximant-vowel sequence was pronounced as a hiatus, [ˈnu.i], accentuating the first syllable. In both cases it seems that language-specific constraints (related to stress-timed vs. syllable-timed accent patterns and morphology) override universal sensory constraints. The present empirical study addresses the following three questions:

i) Are test persons (TPs) capable of evaluating the sensory adequacy of acoustic forms (words from 6 Indo-European and Non-Indo-European languages) with respect to the encoded concept and, if so, do the TPs agree in their evaluation?

ii) What is the role of prominence (Kohler 2008, Niebuhr 2007) as a criterion for qualifying spoken words as being adequate or not?

iii) What are the implications of prominence for cognitive linguistics, e.g. cohort-based word recognition models (Aitchison 2003, Marslen-Wilson 1987)?

The quantitative results show that in a sample of 110 TPs with different L1, a significant part was able to evaluate spoken words on a scale from 0-10 (Sensory Adequacy Values, SAVs) and to agree on a top ten and worst ten item list. Female TPs exhibited a superior discrimination ability, i.e. there was a gender-related significant difference in the distribution of SAVs. Finally, in the context of sensory adequacy evaluation of the presented items, prominence proved to be a suitable qualitative factor influencing the ranking of the spoken words according to their perceived/associated sensory characteristics. Together, the results shed new light on the cohort-based models of spoken word recognition, suggesting that prominence may have a facilitatory or inhibitory effect on phoneme-based lexicon access and last but not least, on a hearer's processing performance.

	bright 'day'	**indifferent**	**dark** 'night'
match	['di.es] ['di.a]		['noks] ['no.če] *['nu.i]
indifferent		*[di.'a] *[di.'es] *[no.'če] *[nu.'i]	
mismatch	['nɥi]		['ʒuʁ]

Figure 1. *Item distribution according to dark/bright sensory match*

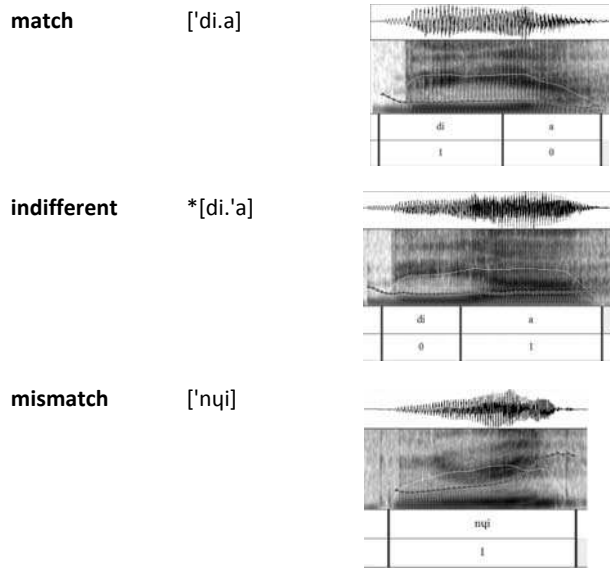


Figure 2. *Test items (PRAAT-Analysis)*

[1] Ohala, J.J. 1994. The frequency code underlies the sound symbolic use of voice pitch. In Hinton L., Nichols, J. & Ohala J.J. (eds.), *Sound symbolism.* Cambridge: Cambridge University Press, 325-347.

[2] Blasi, D.E. et al. 2016. Sound-meaning association biases evidenced across thousands of languages. *PNAS*, 113, 10818-10823.

[3] Jakobson, R. 1979. Linguistik und Poetik. *Poetik. Ausgewählte Aufsätze. 1921-1971*, Frankfurt am Main: Suhrkamp, 83–121.

[4] Kohler, K.J. 2008. The Perception of Prominence Patterns. *Phonetica*, 65, 257-269.

[5] Niehbuhr, O. 2007. *Perzeption und kognitive Verarbeitung der Sprechmelodie.* Berlin: Walter de Gruyter.

[6] Aitchison, J. 2003. *Words in the Mind. An Introduction to the Mental Lexicon.* Malden: Blackwell Publishing.

[7] Marslen-Wilson, W.D. 1987. Functional parallelism in spoken word-recognition. *Cognition,* 25, 71-102.

Exploring the phonetic & phonological contribution to prominence in sentence-initial *I think*

Daniela Kolbe-Hanna, Clare Maas and Laura Panne

Universität Trier

We examine the interplay of phonetic and phonological factors in creating prominence in *I think* based on data from a controlled experiment. Previous studies of *I think* [1, 4, 5, 7, 12] have pointed out the importance of prominence to the function of *I think*, but studies on prominence in English have provided conflicting results on the exact contribution of the individual parameters of duration, loudness and pitch to the production and perception of prominence [2, 6, 9, 11]. This study investigates which features in particular account for the production of prominent syllables in sentence-initial *I think* and whether a particular function correlates with a specific phonetic or phonological feature. Based on Aijmer [1], Kärkääinen [4], Simon-Vandenbergen [7] and Kolbe-Hanna [10], we focus on differences in discourse-pragmatic functions related to the question of which element in *I think* the speaker intends to draw attention to. Hence, *I think* with prominent *I* (***I** think*) functions as a booster or creates a contrast to an opposing stance – sometimes assumed, sometimes expressed openly in previous discourse. *I think* with prominent *think* (*I **THINK***), however, functions as an epistemic hedge to express uncertainty (*I [just] think, I don't know*), or simply as a frame for the following statement ([4]). We examine whether boosting or contrastive prominent ***I*** is produced differently than hedging ***THINK*** (cf. [8] and which phonetic and phonological features are involved in creating this prominence. We also aim to determine which phonetic features involved in the production of prominence exist when *i think* is non-prominent and whether certain phonetic or phonological patterns correlate with *I think* being integrated into or separated from larger intonation units.

In the experiment, 23 native speakers of British or American English read out five sentence stimuli in three iterations, one stimulus in two iterations and 43 filler sentences. Four of the six stimuli with *I think* were chosen or adapted to elicit prominence of either *I* – as in (1)– or *think* – as in (2). Two sentences do not contain any cues to assign more prominence to either of the two words – as in (3) and were designed to elicit speakers' "default" pronunciation of sentence-initial *I think.*

(1) *Most people don't like this flower, but I think it's beautiful*

(2) *I think my uncle was born in 1955, but I'm not sure.*

(3) *I think I've never seen fifteen people sleeping in one bed.*

Using PRAAT, we conducted acoustic phonetic measurements: (proportional increase of) duration of the voiced sounds, intensity (based on the root mean squared average) and pitch (fundamental frequency f_0). The findings from this acoustic analysis are backed up by results from a study on the perception of prominence by untrained native speakers.

Preliminary findings indicate that different factors contribute to creating prominence in different words. For instance, prominent *I* is always produced with longer duration than non-prominent *I,* whilst that prominence in *think* consistently occurs with a rise in pitch. This may be because *I* can be more easily lengthened than the phonologically short /ɪ/ in *think*, but may also be connected to the boosting / hedging function of the chunk under investigation (see [8]).

[1] Aijmer, K. 1997. *I think* – an English modal particle. In T. Swan and O.J. Westvik, (Eds), *Modality in Germanic languages*.. Berlin: Mouton de Gruyter. 1–47.
[2] Bolinger, D.L. 1958. A Theory of Pitch Accent in English. *WORD*. 14(2–39), 109–149.
[3] Boye, K. and Harder, P. 2007. Complement-taking predicates: Usage and linguistic structure. *Studies in Language*. 31(3), 569–606.
[4] Dehé, N. and Wichmann, A. 2010. Sentence-initial *I think (that)* and *I believe (that)*. *Studies in Language*. 34(1), 36–74.
[5] Dehé, N. and Wichmann, A. 2010. The multifunctionality of epistemic parentheticals in discourse: Prosodic cues to the semantic-pragmatic boundary. *Functions of Language*. 17(1), 1–28.
[6] Gussenhoven, C. 2010. *The phonology of tone and intonation*. Cambridge: Cambridge Univ. Press.
[7] Kärkkäinen, E. 2003. *Epistemic Stance in English Conversation*. Amsterdam: John Benjamins.
[8] Katz, J. and Selkirk, E. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language*. 87(4), 771–816.
[9] Kochanski, G. et al. 2005. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*. 118(2), 1038–1054.
[10] Kolbe-Hanna, D. 2016. Multifunctional *I think* in SPICE Ireland and VOICE Canada. Talk at *Discourse-Pragmatic Variation and Change* 3 (Ottawa).
[11] Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
[12] Simon-Vandenbergen, A.-M. 2000. The functions of I think in political discourse. *International Journal of Applied Linguistics*. 10(1), 41–63.
[13] Uhrig, P. 2018. NewsScape and the Distributed Little Red Hen Lab – A Digital Infrastructure for the Large-scale Analysis of TV broadcasts. A.-J. Zwierlein et al. (Eds.), *Anglistentag 2018 in Regensburg: Proceedings. Proceedings of the Conference of the German Association of University Teachers of English*. Trier : Wissenschaftlicher Verlag Trier.

Underlying mechanisms in the perception of metrical prominence – The role of occurrence frequency of different pitch accent types

Sophie Kutscheid[1], Katharina Zahner[1], Adrian Leemann[2] & Bettina Braun[1]
[1]*University of Konstanz,* [2]*University of Bern*

Intonation clearly affects the perception of phrasal prominence (e.g., [1, 2]). This project investigates how the intonational realization of words affects the perception of word stress (position of the stressed syllable) and which role experience (frequency of occurrence of different pitch accent types) plays therein. Recent studies have shown that the position of the pitch peak, resulting from different pitch accent types, affects the perception of the position of metrical stress [3-6], i.e., high-pitched syllables are more likely to be perceived as stressed. [4] further showed that – in a visual world eye-tracking paradigm – the reliance on high pitch as a cue to stress is reduced when participants are exposed to recordings that only contained accents in which the stressed syllable is *not* high-pitched (H+L*, L*+H, L*) before. This suggests that one of the mechanisms in stress processing is the learned association between high pitch and metrical stress based on the frequent co-occurrence of high pitch and stress [7]. Here, we put this hypothesis to further test by examining whether offline stress identification is also affected by the frequency of occurrence of different pitch accent types in the immediate input.

To this end, we conducted a stress identification task with trisyllabic German nouns ([4]), 36 experimental WSW trials and 72 filler trials (half SWW, half WWS) and prefixed a 3-minute exposure phase with either solely low-pitched or high-pitched stressed syllables. The materials in the exposure phase were selected to be compatible with both low-pitched (H+L*, L*+H, L*) and high-pitched accents (L+H*) and contained 120 accented syllables in total. To engage participants, they rated the speaker's voice for different characteristics (e.g., *intelligent* [4]). In the subsequent test phase, participants judged the position of the stressed syllable on a three-buttoned response box. Stimuli for test were spoken by the same speaker as for exposure and presented in three PSOLA-resynthesized intonation conditions (early-peak, medial-peak, late-peak, Latin-Square Distribution) in isolation, see Fig. 1. Medial-peak contours were resynthesized half from early- and half from late-peak original recordings (f0 peak moved to the second syllable); early- and late-peak contours were resynthesized from originally recorded medial-peak contours (f0 peak moved to the first or third syllable, respectively). Participants were tested in a stress identification task, 18 following low-exposure (11 female, M_{age} = 24.0 years, SD = 4.1) and 18 following high-exposure (12 female, M_{age} = 23.3 years, SD = 3.8).

Correctness of the stress judgements for experimental words (WSW) were statistically analyzed using logistic mixed effects regression models [8]. Irrespective of the type of exposure, results showed a main effect of *intonation condition* (p<0.001), i.e., more correct responses in the medial peak condition as compared to the early- and late-peak condition. *Exposure-type* had no effect (p=0.66). Importantly, there was an interaction between *intonation condition* and *exposure-type* (p=0.03), with a larger effect of *intonation condition* on correctness after the high-exposure, see Fig. 2. Hence, the effect of *intonation condition* on stress judgements was modulated by the frequency of low- or high-pitched stressed syllables in the immediate input. An analysis of the error patterns further revealed a strong response bias towards the syllable carrying the f0 peak in the high-exposure, while this bias was reduced after the low-exposure phase, see Fig. 3.

Overall, our results showed that listeners' use of high pitch as a cue to stress was reduced in the low-exposure condition, similar to the findings from online speech comprehension [4]. Based on these results, we argue that the distribution frequency of different pitch accent types modulates the reliance on high f0 as a stress cue. We are currently investigating this mechanism further by testing Bern Swiss German listeners, who are predominantly exposed to low-pitched stressed syllables [9] (long-term exposure).

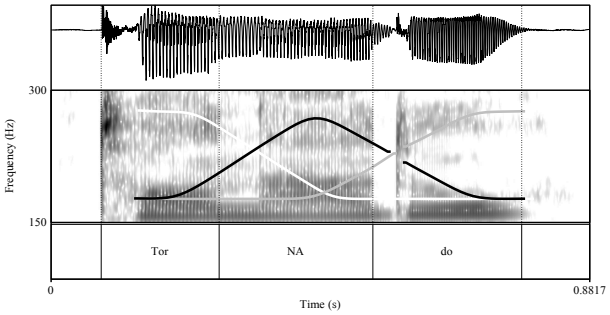


Figure 1. *Sound pressure wave, spectrogram and f0 contours for early- (white), medial- (black) and late-peak (grey) contours (PSOLA-resynthesized) in one experimental trial.*

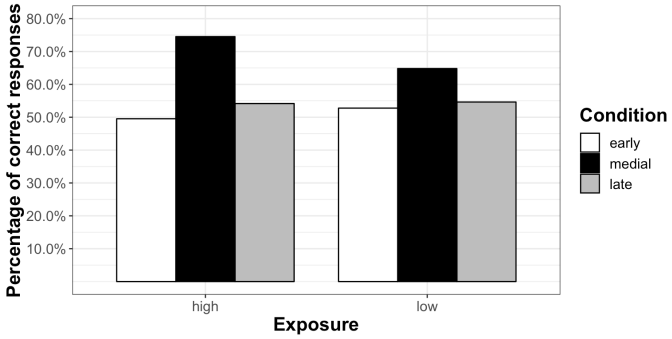


Figure 2. *Average correctness rates in the three intonation conditions (early-peak, medial-peak and late-peak condition), split by type of exposure (high vs. low).*

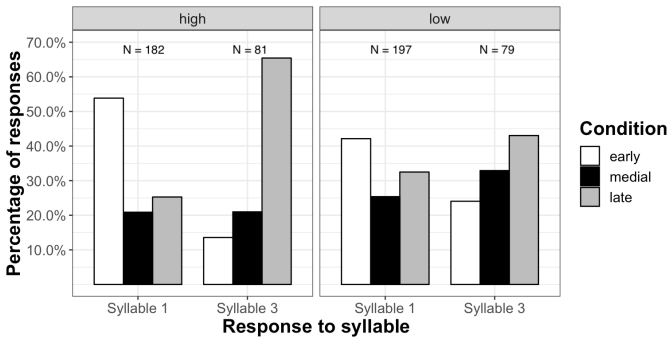


Figure 3. *Percentage of erroneous responses to syllable 1 and 3 in WSW targets (see N for absolute numbers, % refers to each N), in the three intonation conditions, by exposure-type.*

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

[2] Baumann, S., & Röhr, C. 2015. The perceptual prominence of pitch accent types in German. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*, (Glasgow, UK).

[3] Egger, S., & Braun, B. 2015. The role of pitch accents on the identification of metrical prominence. Poster presented at the *International Conference on Prominence in Language,* (Cologne, Germany).

[4] Zahner, K., Kutscheid, S., & Braun, B. 2019. Alignment of f0 peak in different pitch accent types affects perception of metrical prominence. *Journal of Phonetics*, 74, 75-95.

[5] Schwab, S., & Dellwo, V. 2017. Intonation and talker variability in the discrimination of Spanish lexical stress contrasts by Spanish, German and French listeners. *JASA,*142, 2419-2429.

[6] Dilley, L. C., & Heffner, C. C. 2013. The role of f0 alignment in distinguishing intonation categories: Evidence from American English. *Journal of Speech Sciences,* 3, 3-67.

[7] Peters, B., Kohler, K., & Wesener, T. 2005. Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache - Statistische Verteilung und sprachliche Funktion. In Kohler, K., Kleber, F., and Peters, B. (Eds.), *Prosodic structures in German spontaneous speech*. Kiel: IPDS, 185-201.

[8] Baayen, R. H., Davidson, D. J., & Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language,* 59, 390-412.

[9] Leemann, A. 2012. *Swiss German intonation patterns*. Amsterdam: John Benjamins Publishing.

Focus-induced articulatory prominence on velum actions in nasal geminates

Miran Oh and Dani Byrd
Department of Linguistics, University of Southern California, Los Angeles, CA USA

A close examination of articulation under focus production is a valuable route to deciphering how prosodic prominence is achieved. The seminal work of [1] first illuminated the dynamical mechanisms underlying the articulation of prominence, and numerous other studies have since examined the oral articulation of both consonants and vowels under focus in order to paint a picture of how speakers control the spatiotemporal properties of articulatory actions realizing prosodic salience. However significant lacunae remain in our knowledge when the segments under consideration are more complex. We endeavor to flesh out this understanding for multi-gesture structures having non-oral gestural components and complex internal temporal organization by examining the articulation of prominence for nasal geminates.

Geminates have a longer duration than singletons, but they do not consistently have a larger displacement or tighter constriction than singletons [2, 3], though such hyperarticulation can be notoriously difficult to observe for stop consonants given that once closure occurs only a small amount of compression can further result from hyperarticulation. Additionally for geminate segments comprising multiple gestures, the target of any lengthening associated with prominence is unclear. Therefore nasal geminates—with their (non-constriction) velum lowering component and inherent length due to gestural concatenation—are a valuable gestural structure for investigating the spatiotemporal articulatory implementation of focus.

This study examines nasal juncture geminates in Korean using real-time MRI (rtMRI) data. Target consonants are alveolar oral/nasal stops occurring as singletons and as a geminate across an Accentual Phrase boundary: singleton onset /V#**n**/, singleton coda /**n**#p/, and geminate /**n**#**n**/; with utterances occurring with boundary-initial focus either present or absent. RtMRI data were acquired from a single native speaker using the speech production protocol in [4, 5]. A centroid tracking analysis [6] and a region-of-interest image sequence analysis [7] were performed to provide kinematic trajectories of Tongue Tip (TT) constriction formation and Velum (VEL) lowering and raising gestures. For each gesture, we examine duration (plateau, oral constriction, & velum lowering), magnitude (TT constriction degree & Velum vertical displacement), and intergestural timing lag (between TT & VEL).

Findings show that singletons and geminates are best distinguished by TT and VEL gestural plateau duration (Fig. 1-2: left) and that these durations greatly lengthen under focus in geminates but not in singletons. The focus effect of TT constriction duration lengthening (Fig. 1-2: center) is similar across singletons and geminates, and VEL lowering duration increases under focus in singleton coda and geminates but not in the singleton onset. For TT constriction degree (Fig. 1: right) singleton codas have intrinsically lesser constriction degree than onsets and geminates and show some increase in constriction degree under focus. VEL lowering magnitude (Fig. 2: right) is larger in codas and geminates than in onsets, and onsets tend to reduce lowering and codas increase lowering under focus, while no focus effect on VEL lowering is seen for geminates. Lastly, intergestural timing between the TT and VEL lowering onsets (Fig. 3: left) is stable across segments and across focus conditions. However, TT onset to VEL raising onset lag (Fig. 3: right), which can be an index of nasality, increases notably under focus, particularly for geminates.

In conclusion, geminates and singleton nasals are distinguished by their constriction plateau duration as well as timing between TT and VEL raising onsets. These same features that most saliently distinguish the singletons and geminates become lengthened substantially under focus, with the velum remaining in its lowered position longer for geminates under focus than for singletons, suggesting the possibility of a degree of subtle degemination of the juncture geminates at a boundary under focus. [Supported by NIH DC03172 & DC007124]

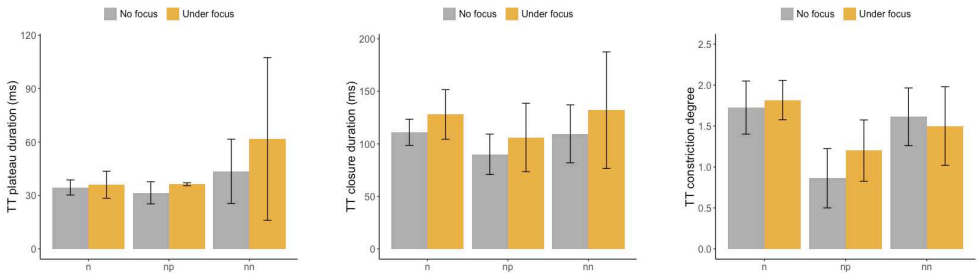


Fig. 1. *TT plateau duration (left), constriction duration (center), and constriction degree (right) for /#n, n#, n#n/ under focus (yellow) and no focus (grey).*

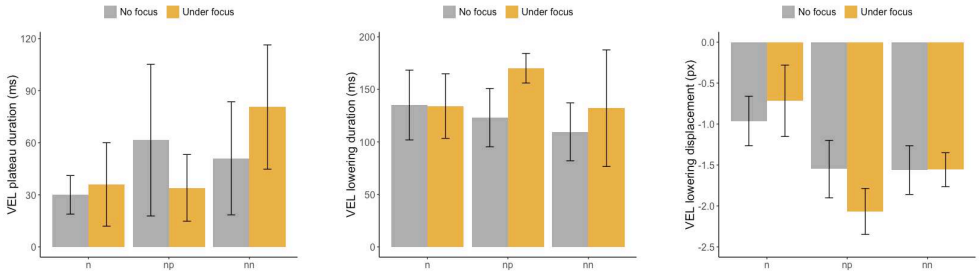


Fig. 2. *VEL plateau duration (left), lowering duration (center), and lowering magnitude (right) for /#n, n#, n#n/.*

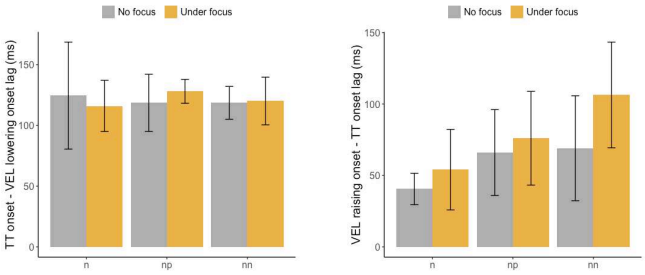


Fig. 3. *TT onset - VEL lowering onset lag (left) and VEL raising onset - TT onset lag (right).*

[1] Beckman, M., Edwards, J., & Fletcher, J. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In *Papers in Laboratory Phonology III*, edited by G. Docherty and D. R. Ladd (Cambridge U.P., Cambridge), pp. 68–86.

[2] Byrd, D. 1995. Articulatory characteristics of single and blended lingual gestures. In *Proc. of the XIIIth Int. Congr. of Phon. Sci.* (Stockholm, Sweden), 438–441.

[3] Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. 2009. Timing effects of syllable structure and stress on nasals: a real-time MRI examination. *J. of Phon.*, *37*(1), 97–110.

[4] Narayanan, S., Nayak, K. S., Lee S., Sethy, A, & Byrd, D. 2004. An approach to real-time magnetic resonance imaging for speech production. *J. Acous. Soc. of Am.,* 115(4), 1771–1776.

[5] Lingala, S., Zhu, Y., Kim, Y. C., Toutios, A., Narayanan, S., & Nayak, K. S. 2017. A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magnetic Resonance in Medicine*. 77(1), 112–125.

[6] Oh, M., & Lee, Y. 2018. ACT: An Automatic Centroid Tracking tool for analyzing vocal tract actions in real-time magnetic resonance imaging speech production data, *J. Acous. Soc. of Am.,* 114(4), EL290–EL296.

[7] Lammert, A., Ramanarayanan, V., Proctor, M., & Narayanan, S. 2013. Vocal tract cross-distance estimation from real-time MRI using region-of-interest analysis. In *INTERSPEECH* (Lyon, France), 959–962.

The development of lexical stress in young Italian children

Francesco Olivucci, Mario Vayra[1], Cinzia Avesani[2], Claudio Zmarich[2]
[1]University of Bologna, [2]CNR-ISTC (Padova)

1. Introduction. Our study aims to identify the developmental trajectory of lexical stress by analyzing its acoustic correlates in young Italian children. Stress is a structural property of a word that specifies which syllable in the word is the strongest ([1], a.o.). More than 50 years or research on stress accent languages provided cumulative evidence on the acoustic and articulatory correlates of stress in adults, all showing how stressed syllables (S) are endowed with a given set of acoustic/articulatory properties in a larger degree than unstressed (U) ones (e.g. [3], for a review). It is commonly accepted that U syllables pave the way for S ones to stand out of the linguistic context ([11]), implicitly assuming that by starting from an U syllable and by adding more prominence to it, the basic condition by which lexical contrasts can be realized is provided. In this study we question whether the path to the development of lexical stress does indeed proceed from the production of U syllables to S ones and, if so, when and how a clear differentiation occurs. **2. Background**. Previous studies on young English children's productions have investigated lexical stress development focusing mainly on vowel duration, intensity and F0 as acoustic parameters (e.g [6], [8], [9]). As for Italian, very few studies have investigated how children produce S and U syllables in terms of these parameters over the late pre-schoolers' years of life ([2], [7]). Moreover, all previous studies focused on groups of children, did not try to delineate developmental trajectories for each single child, nor started the analyses from the stage of the first vocabulary. **3. Method.** The productions of three children (BS, CN and LM), from North-East Italy (Trieste), recorded every 3 months from age 21 to 42 months were analyzed. Our targets are vowels in CV syllables (where C = stop, V = /a/, /i/) in trisyllabic and disyllabic words (about 20 S + 20 U syllables per stage). All of the recordings were collected and IPA transcribed. For each S and U vowel, we calculated duration, peak intensity, F1 and F2 at vowel midpoint, formant trajectories and spectral emphasis (see, for prominence in Italian adults, [8], [2]). As a control group we recorded and analyzed two adult females of the same geographical area. For each subject and each acoustic measure, we run a Linear Mixed Model with Stress and Age as Fixed effects and Modality (repetition vs spontaneous productions) and Prosody (single words vs words with nuclear accent vs words with pre-nuclear accent) as Random Effects. For formant trajectories we applied GAMM models. **4. Results**. Overall, the most statistically significant parameter is duration, followed by intensity (according to [3]), F1 (for [a]), F2 (for [i]) and spectral balance (but not spectral tilt). Children start differentiating S from U vowels following different developmental paths. In terms of duration S and U vowels ([a] and [i]) are statistically different since the first age (21 months) for two out of three children (BS and CN), while for LM this difference is significant only at 30 months. Also, the more the children grow up, the more S vowels are longer than U ones. For one child the development starts from a stage where S and U vowels are comparable in terms of duration (see for example data for 21 months for LM in Figure 1) to a stage (42 months) where S vowel are longer than U ones. 5. **Discussion**. While the fact that children acquire lexical stress in Italian by producing longer duration for S vowels is expected, what is unexpected is the path through which the difference between S and U vowels is achieved: across stages, the duration of S vowels is pretty much stable, while the duration of U vowels is decreasing (see figure 2). We speculate that children's production at the very beginning of their development present a "flat" prosody, where no element is prominent. At some point (21 months for someone, 30 months for others), one element begins to stand out from the context but not because the child starts putting more effort in it; rather, because (s)he learns how to reduce the prominence for U vowels to provide to right context for S vowels to stand out.

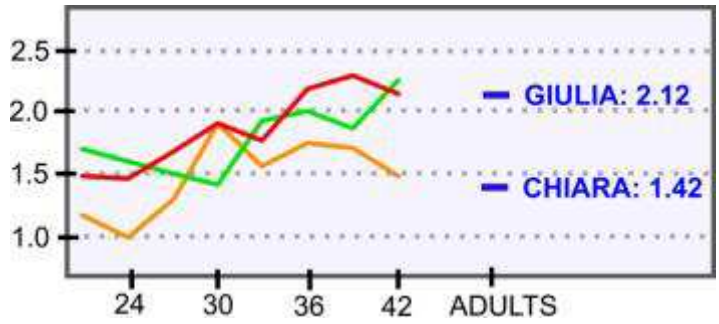


Figure 1. *Ratio between Stressed and Unstressed vowels ([a]) duration for the three subjects across the months, compared with adults' data*. BS: red line; CN: green line; LM: yellow line.

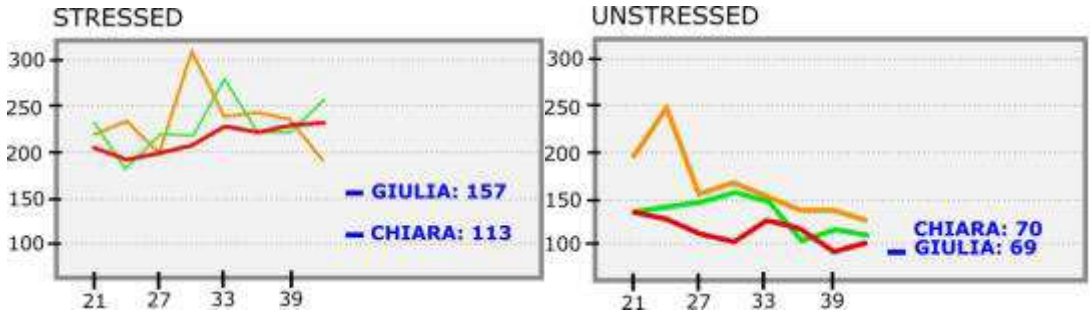


Figure 2. *Duration (in ms) of Stressed (left) and Unstressed vowels (/a/) for the three subjects across the months, compared with adults' data (Giulia and Chiara)*. BS: red line; CN: green line; LM: yellow line.

[1] Sluijter, A.M.C. & van Heuven (1996). Spectral balance as an acoustic correlate of linguistic stress. JASA, 100 (4), 2471-2485. **[2]** Arciuli, J., & Colombo, L. 2016. An acoustic investigation of the developmental trajectory of lexical stress contrastivity in Italian. *Speech Communication*, 80, 22-33. **[3] Ortega-Llebaria & Prieto (2010). Acoustic correlates of stress in Central Catalan and Castilian Spanish.** *Language and Speech*, 54(1), 73-97. **[4] Bocci, G., & Avesani,** C. 2011. Phrasal Prominences do not need Pitch Movements: Postfocal Phrasal Heads in Italian. *Interspeech*, 1357-1360. **[5] Erikson, A., Bertinetto, P.M., Heldner, M., Nodari, R., & Lenoci, G**. 2016. THe acoustics of lexical stress in Italian as a function of stress level and speaking style. *Interspeech 2016, San Francisco, USA, September 8-12*, 1059-1063. **[6] Kehoe, M. Stoel-Gammon, C., & Buder, E. H.** 1995. Acoustic correlates of stress in young children's speech. *Journal of Speech, Language and Hearing Research*, 38(2), 338-350. **[7] Olivucci, F., Pasqualetto, F., Vayra, M., & Avesani**, C. 2016. Lo sviluppo dell'accento lessicale nel bambino in età prescolare: una prospettiva fonetico-acustica. *Collana Studi AISV*, vol.II, AISV, Milano. **[8] Pollock, K. E., Brammer, D. M., & Hageman, C. F**. 1993. An acoustic analysis of young children's productions of word stress. *Journal of Phonetics* 21,183-203. **[9] Schwartz, R. G., Petinou, K., Goffman, L., Lazawki, G., & Cartusciello, C**. 1996. Young children's production of syllable stress: An acoustic analysis. *The journal of the Acoustical Society of America*, 99(5), 3192-3200. **[10] Tamburini, F**. 2009. Prominenza frasale e tipologia prosodica: un approccio acustico. *Atti del XL Congresso SLI*, 437-455. **[11] Terken, J.** 1991. Fundamental frequency and per- ceived prominence. *Journal of the Acoustical Soci-ety of America* 89(4), 1768–1776.

Perception of non-native prosody: the case of prosodic prominence

Valentina Schettino[1], Francesco Cutugno[2] and Antonio Origlia[2]
[1]*University of Naples "L'Orientale",* [2]*University of Naples "Federico II"*

In this work, we intend to investigate phenomena involved in the perceptual processes related to prominence [1]. More specifically, we study relations between acquisitional processes and prominence perceptual mechanisms. We question which strategy concerning prominence perception is used by listeners when their language is uttered by non-native speakers. This paper is included in the theoretical frame depicted in [2].

In order to do that, we carried out an experiment using a corpus containing prosodically annotated material in German L1 and Italian L2 (for any further detail see [3]). We examined a sample of the data of about 500 sentences in Italian L2. Using the PromDrum technique [4] enhanced through the DTW procedure [5], we let 9 Italian native speakers drum the degree of perceived prominence of 61 different sentences. Utterances were selected trying to locate the files in which intonation and stress patterns differed from the norm: if a stress was put on the "wrong" syllable, or the pitch shape diverged from usual Italian patterns and/or alignment, the file was considered to be a good element of investigation.

The first empirical observation we made is that drummers exhibit different behaviours depending on the fluency level of the L2 talker. Indeed, the highest levels of perceived prominence, corresponding to the higher drumming intensity, are marked on syllables with typical acoustic indicators of prominence in Italian, i.e. mainly duration [6]. This can be seen in figure 1, where the proposed stimulus was associated to a speaker of the A level. On the contrary, we find that stimuli produced by higher levels speakers are perceived integrating also linguistic knowledge coming from other related domains, more than acoustics alone. As regards pragmatics, for example, positional factors do not seem to be relevant in A-level files (figure 2), while they do play a role with C-level input files (figure 3). Also information structure is confirmed as a less relevant domain influencing perception of prominence with A-level input files, but not with more fluent productions (figure 1 and figure 4). This tendency is confirmed throughout the considered corpus.

The conclusion we can draw from this first evaluation of peculiar cases in Italian L2 produced by German native speakers is that perceptual processes are influenced by the quality of the input audio files on the fluency level. Indeed, the complex interplay of acoustics, information structure, pragmatics and the related prosodic domains of intonation and stress shaping prosodic prominence is also influenced by the competence of the speaker, that plays a big role in the production phase. The more what they hear is similar to a native-like production, the more they make full use of their additional linguistic knowledge.

For these reasons, we suggest to rethink perceptual processes linked to prominence as adaptive phenomena, in which an interface "filters" the relevant elements useful on the basis of given contexts. If this is true, comparative analyses between degrees of perceived prominence of L1 and L2 material could help disentangling the complex bundle relative to this phenomenon, furthering our understanding of the complex interplay of top-down expectancies and bottom-up features in different contexts.

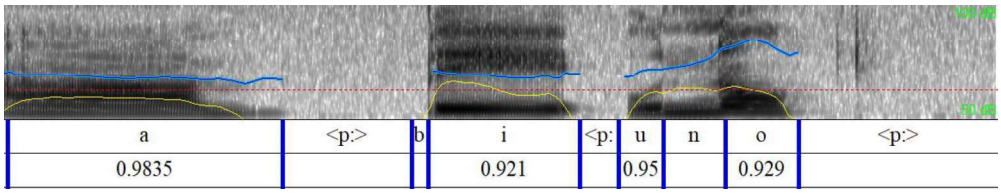


Figure 1. *Role of duration, A-level.*

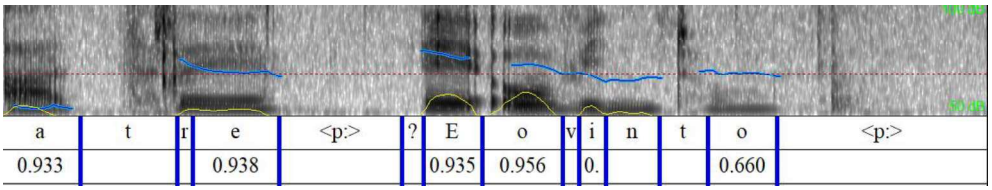


Figure 2. *Influence of positional factors, A-level.*

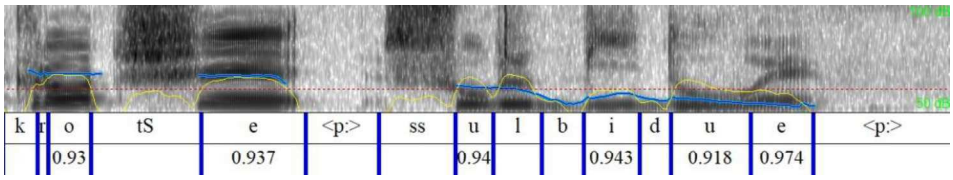


Figure 3. *Influence of positional factors, C-level.*

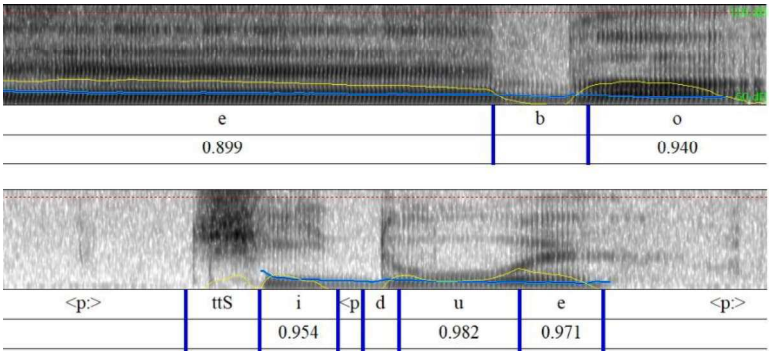


Figure 4. *Influence of the information structure domain, C-level.*

[1] Terken, J. 1991. Fundamental frequency and perceived prominence. Journal of the Acoustical Society of America 89(4), 1768–1776.

[2] Wagner, P., Origlia, A., Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., Escudero Mancebo, D., Gili Fivela, B., Lacheret, A., Ludusan, B., Moniz, H., Chasaide, A. N., Niebuhr, O., Rousier-Vercruyssen, L., Simon, A. C., Simko, J., Tesser, F. & Vainio, M. (2015). Different Parts of the Same Elephant: a Roadmap to Disentangle and Connect Different Perspectives on Prosodic Prominence, Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, Scotland, online at http://www.icphs2015.info/pdfs/Papers/ICPHS0202.pdf [last accessed 17/11/2018].

[3] Schettino, V. (2015). Prosogit: a corpus for prosodic studies in German L1 and Italian L2. AION, 25(1/2), 237-255.

[4] Samlowski, B., & Wagner, P. (2016). PromDrum-Exploiting the prosody-gesture link for intuitive, fast and fine-grained prominence annotation. Proceedings of Speech Prosody 2016.

[5] Schettino, V., Origlia, A., Cutugno, F. (accepted). Dynamic Time Warping and Prosodic Prominence, Proceedings of the XIV AISV Conference, Bolzano, 25-27/01/2018.

[6] D'Imperio, M. (1998). Acoustic-perceptual correlates of sentence prominence in italian questions and statements. Journal of The Acoustical Society of America 104.

Prosodic prominence versus frequency effects on the acquisition of CCV branching onsets in Brazilian Portuguese

Andressa Toni[1]

[1]*University of São Paulo*

Terken & Hermes (2000)[1] describe the quality of "being prominent" roughly as "standing out from the environment". In this study, we compare the effects of two different ways of standing out from a linguistic environment – a qualitative way, by prosody (via stress); and a quantitative way, by frequency (via number of occurrences). The comparison is addressed by observing the effects of prosodic prominence and frequency prominence on the development course of CCV branching onsets ($Consonant_1$+$Consonant_2$+Vowel) in Brazilian Portuguese (BP). In this language, most occurrences of CCV syllables are unstressed: the dictionary corpus of Viaro & Guimarães-Filho (2007)[2] presents 27,767 CCV syllables, from which 70.85% are pretonic, 10.68% are post-tonic and 18.47% are stressed. Following the same pattern, the speech corpus of Mendes (2013)[3] presents 44.25% pretonic CCV syllables, 26.56% post-tonic and less than 30% stressed CCV occurrences. The distribution on these two corpora points to opposite prominence patterns towards CCV: prosodically prominent CCV syllables are not the most frequent, quantitatively prominent occurrences of CCV. Our goal is to observe which (if any) of those prominence patterns can be reflected on the acquisition of branching onsets in BP. Branching onsets are pointed as both articulatorily and phonologically challenging for the child, being fully developed only by 5 years old[4] – although common words containing CCV syllables may figure as targets in child speech even before 2 years old[4]. Until CCV is fully acquired, branching onsets are often produced by repair strategies meant to simplify the CCV structure to CV, as in (1); or to modify the segmental content and structure of CCV, as in (2):

(1) /**gɾu**dej/ 'I sticked' → [**gu**ˈdej], [**gu.ɾu**ˈdej], [**gu**.deˈ**ɾej**]

(2) /**gɾu**dej/ 'I sticked' → [**guɾ**ˈdej]; /**gɾu**dado/ 'It's sticked' → [**glu**ˈda.dʊ]

By analyzing the frequency and stress patterns of CCV targets in child productions, we aim to observe if stressed CCV syllables (most prosodic prominent context) or pretonic CCV syllables (most frequent, quantitatively prominent context) would present higher rates of adult-like productions. We also aim to observe if different stress contexts are more likely to present different repair strategies (as in (1) or (2)). Longitudinal data from 3 children was verified with Praat and CCV productions were categorized into *Target-like*, *CCV>CV Simplification* (as in (1)) and *Other repairs* (as in (2)). Stress patterns were extracted with the FreP tool[5]. Results show that, in general, most target-like CCV syllables produced by the child are stressed, closely followed by pretonics – as well as simplified CCV>CV occurrences and other CCV repairs (cf. Table 2). This is due to the stress distribution of child CCV targets: children's productions are more equally distributed regarding lexical stress, with a difference of less than 10% stressed syllables over pretonic syllables – in comparison, the Dictionary corpus had a difference of 50 percentual points between stressed and pretonic CCV syllables, favoring pretonics. However, when compared to the Child Directed Speech of their caretakers, children's CCV targets have similar stress distributions to adult's CCV syllables (cf. Table 1). By analyzing the proportion of target-like, CCV>CV simplifications and other repair strategies on the total of stressed, pretonic and post-tonic CCV syllables (cf. Table 3), we observe that around the age of CCV acquisition, at 5;0 years old, children present 74% of target-like CCV syllables on the stressed context, while no more than 50% is presented on pretonic and post-tonic contexts. No difference related to stress is observed regarding repair strategies: CCV>CV simplification is favored on the three stress conditions. The results on Tables 1, 2 and 3 indicate that prosodic prominence stands out more than quantitative prominence on language acquisition data: frequency and stress point towards the same direction on child speech and child directed speech, and stressed CCV syllables are more likely to present target-like productions after age 5;0 compared to unstressed branching onsets.

Table 1: Distribution of CCV syllables in Brazilian Portuguese adult corpora					
Data	*Total Words*	*Total CCV*	*% stressed CCV*	*% pretonic CCV*	*% post-tonic CCV*
Dictionary[2]	150,875	27,767	18.47%	70.85%	10.68%
Speech corpus[3]	363,848	30,114	29.19%	44.25%	26.56%
Child Directed Speech[4]	-	12,799	45.63%	28.96%	25.41%
Child speech[4]	-	4,266	40.6%	33.45%	25.95%

Table 2: CCV syllables in child speech: by type of production										
Age range	*Target-like CCVs*			*CCV>CV Simplification*			*Other repairs*			*Total CCV per age*
	Pretonic	*Stressed*	*Post-tonic*	*Pretonic*	*Stressed*	*Post-tonic*	*Pretonic*	*Stressed*	*Post-tonic*	
<2;0	10	90	0	8.70	65.22	26.09	66.67	33.33	0	104
2-2;11	25	25	50	22.40	45.12	32.48	18.75	68.75	12.5	1017
3-3;11	37.58	41.61	20.81	35.46	39.92	24.63	46.15	40.66	13.19	1,653
4-4;11	49.83	38.87	11.30	38.08	34.17	27.75	37.50	45.83	16.67	1042
5-5;6	30.38	46.15	23.46	41.21	23.03	35.76	45.83	16.67	37.50	449
All ages	39.48	42.49	18.03	31.79	40.06	28.15	40.23	43.10	16.67	4,266

Table 3: CCV syllables in child speech: by stress										
Age range	*Total CCV*	*% stressed CCV*			*% pretonic CCV*			*% post-tonic CCV*		
		Target-like	*CCV>CV*	*Others*	*Target-like*	*CCV>CV*	*Others*	*Target-like*	*CCV>CV*	*Others*
<2 ;0	104	12.86	85.71	1.43	9.09	72.73	18.18	0	100	0
2;0-2;11	1,017	0.65	94.61	4.74	1.32	96.03	2.64	1.84	96.93	1.23
3;0-3;11	1,653	9.35	85.07	5.58	9.35	83.64	7.01	7.93	89	3.07
4;0-4;11	1,042	31.37	65.68	2.95	34.72	63.19	2.08	14.35	83.97	1.69
5;0-5;6	449	74.07	23.46	2.47	50	43.04	6.96	47.29	45.74	6.98
All ages	4,266	17.96	77.71	4.33	20.25	74.84	4.90	11.92	85.46	2.62
Total		**1,732**			**1,427**			**1,107**		

[1] TERKEN, J. M. B., & HERMES, D. J. (2000). The perception of prosodic prominence. In M. Horne (Ed.), Prosody: Theory and experiment. Studies presented to Gösta Bruce (pp. 89-127). Dordrecht: Kluwer.

[2] VIARO, M. E.; GUIMARÃES FILHO, Z. O . (2007) Análise quantitativa da freqüência dos fonemas e estruturas silábicas portuguesas. Estudos Lingüísticos (São Paulo), v. XXXVI, p. 28-36.

[3] MENDES (2013). Projeto SP2010: Amostra da fala paulistana. Disponível em <http://projetosp2010.fflch.usp.br>

[4] TONI, A (2017). Representação subjacente do ataque ramificado CCV na aquisição fonológica. *Master's dissertation*.

[5] MARTINS, F., M. VIGÁRIO & S. FROTA. 2009. FreP - Frequências no Português. Version 2.0. Software

Prosodic distance: Using variation to weight measures of prominence contrast

Laurence White[1], Claire Delle Luche[2] and Caroline Floccia[3]
[1]*Newcastle University,* [2]*University of Essex,* [3]*University of Plymouth*

Patterns of cross-linguistic variation in the temporal distribution and marking of prominence have long been used in attempts to construct prosodic typologies. Thus, the concept of 'rhythm class' was based on an impressionistic (and empirically groundless) hypothesis about stress-based isochrony in some languages contrasting with syllable-based isochrony in others [1]. More recently, empirical measures of durational contrast between more and less prominent syllables have been based on segmentation of utterances into vocalic and consonantal intervals [e.g., 2, 3]. These so-called "rhythm metrics" (or "contrastive rhythm metrics" [4]) tend to agree in showing some languages, such as Iberian Spanish, at the less contrastive end of a continuum and others, such as Southern British English, as the more contrastive end [e.g., 5]. Some metrics lack discriminative power, however, or are highly correlated with speech rate [4, 5], and no simple combination of metrics reliably distinguishes all languages [6]. Some have argued that metrics of contrast should embrace more than just durational cues to prominence [7], whilst others, more trenchantly, have claimed that the unsurprisingly variation in scores between speakers and spoken materials effectively renders such metrics otiose [e.g., 8].

Despite within-language variation, the magnitude of prominence marking does appear to be a perceptually salient factor in between-language comparison, potentially important for first and second language acquisition through its impact on word segmentation [9] and language discrimination [10]. Here we try to derive a composite metric of cross-linguistic prosodic distance using between- and within-language variation, indexed by F-ratios, to weight a combination of acoustic measures related to prominence contrast and linguistic structure.

We recorded four female speakers reading aloud five sentences for each of 13 languages: Cantonese, Dutch, English, French, German, Greek, Hindi, Italian, Mandarin, Polish, Portuguese, Spanish and Welsh. The utterances were hand-segmented into vocalic and consonantal intervals using standard criteria [5] to extract six acoustic variables: median pitch (Hz); pitch range (semitones); coefficient of variation of vowel interval duration (VarcoV); the percentage of utterance made up of vowels rather than consonants (%V); articulation rate (syllables/second); ratio of final vowel duration to average vowel duration (FinVR).

To derive weighted distances in the six-dimensional acoustic space, we ran six one-way ANOVAs, using each of the above measures as the dependent variable and language as the independent variable. We take the resulting F-ratios as a relative measure of how well each acoustic parameter distinguishes between the 13 languages: median pitch – $F(12,247) = 10.15$; semitone range – $F(12,247) = 3.88$; VarcoV – $F(12,247) = 12.08$; %V – $F(12,247) = 16.95$; articulation rate – $F(12,247) = 24.59$; FinVR – $F(12,247) = 3.05$. These F-ratios were used to scale each of the acoustic dimensions, after they were all first normalised as z-scores. Thus, after z-score normalisation and F-ratio weighting, between-language variation in articulation rate (F-ratio 24.59) contributed much more to overall distance than, for example, semitone range (F-ratio 3.88). Because this approach was developed to inform a study of bilingual lexical acquisition in toddlers learning English and another language [11], we took (British) English as our reference and calculated the Euclidean distance between English and the other 12 languages in our scaled six-dimensional space. The prosodic distances thus derived appear in line with expectation (Fig. 1). Welsh, in close contact with English, is nearest, then English's Germanic cousins Dutch and German. Most Romance languages, along with Hindi, are more distant than Polish and Portuguese, which have been characterised as prosodic intermediates, but all are closer to English than the tone languages Cantonese and Mandarin. We consider the benefits and shortcomings of this preliminary new approach, discussing how the prosodic distances thus derived relate to studies of speech perception and language acquisition.

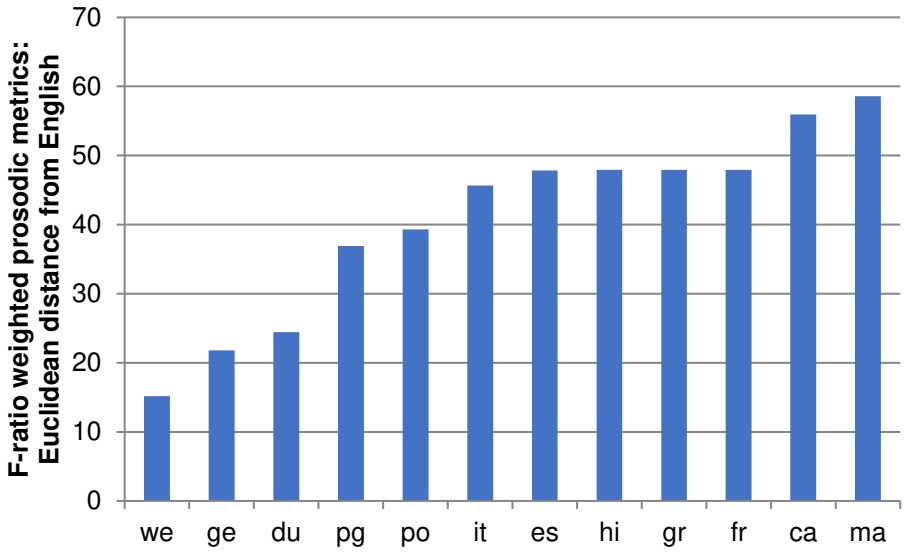


Figure 1. *Combined acoustic metrics (median pitch, semitone range, %V, VarcoV, articulation rate, final vowel ratio) weighted according to F-ratios: Euclidean distance from English of Welsh, German, Dutch, Portuguese, Polish, Italian, Spanish, Hindi, Greek, French, Cantonese and Mandarin*

References

[1] D. Abercrombie, *Elements of General Phonetics*. Edinburgh: Edinburgh University Press, 1967.

[2] F. Ramus, M. Nespor, and J. Mehler, 'Correlates of linguistic rhythm in the speech signal', *Cognition*, vol. 73, no. 3, pp. 265–292, 1999.

[3] E. L. Low, E. Grabe, and F. Nolan, 'Quantitative characterizations of speech rhythm: syllable-timing in Singapore English', *Lang. Speech*, vol. 43, no. 4, pp. 377–401, 2000.

[4] L. Wiget, L. White, B. Schuppler, I. Grenon, O. Rauch, and S. L. Mattys, 'How stable are acoustic metrics of contrastive speech rhythm?', *J. Acoust. Soc. Am.*, vol. 127, no. 3, pp. 1559–1569, 2010.

[5] L. White and S. L. Mattys, 'Calibrating rhythm: First language and second language studies', *J. Phon.*, vol. 35, no. 4, pp. 501–522, Oct. 2007.

[6] A. Loukina, G. Kochanski, B. Rosner, E. Keane, and C. Shih, 'Rhythm measures and dimensions of durational variation in speech', *J. Acoust. Soc. Am.*, vol. 129, no. 5, pp. 3258–3270, 2011.

[7] R. E. Cumming, 'Perceptually informed quantification of speech rhythm in pairwise variability indices', *Phonetica*, vol. 68, no. 4, pp. 256–277, 2011.

[8] A. Arvaniti, 'The usefulness of metrics in the quantification of speech rhythm', *J. Phon.*, vol. 40, no. 3, pp. 351–373, May 2012.

[9] L. D. Sanders, H. J. Neville, and M. G. Woldorff, 'Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues', *J. Speech Lang. Hear. Res.*, vol. 45, no. 3, pp. 519–530, 2002.

[10] L. White, S. L. Mattys, and L. Wiget, 'Language categorization by adults is based on sensitivity to durational cues, not rhythm class', *J. Mem. Lang.*, vol. 66, no. 4, pp. 665–679, May 2012.

[11] C. Floccia *et al.*, 'Vocabulary of 2-year-olds learning English and an additional language: norms and effects of linguistic distance', *Monogr. Soc. Res. Child Dev.*, vol. 83, no. 1, pp. 1–135, 2018.