Perception of Voicing Contrast in Labial-velar Consonants

Matthew Ajibade (maajibad@iu.edu)

Indiana University Bloomington

The properties of labial-velar consonants, predominantly found in West African languages (Ladefoged, 1964; Maddieson, 1984), have continued to garner interest from linguists from classical and instrumental viewpoints. Among other factors, labial-velar consonants are said to possess the distinct quality of being one phonological unit that is characterized by the simultaneous articulation of both labial and velar places of articulation (Ohala, 1979). While there is a consensus on the categorization of labial-velar stops as voiceless [k͡p] and voiced [g͡b], instrumental analyses (e.g., Connell, 1991; Ladefoged, 1964) have reported partial voicing or prevoicing for "voiceless" labial-velar [kp]. The current paper reports the preliminary results of an ongoing study that investigates the potential perceptual challenges posed by this prevoicing characteristic on the discrimination of [k͡p] and [g͡b]. 5 heritage speakers (HS) of Yoruba and an additional 5 native speakers (NS) of the language participated in an oddity task. An Oddity task was chosen instead of other perception tasks such as AX or ABX because of (i) its higher cognitive demand and reduced susceptibility to ceiling effects (Strange & Shafer, 2008) and (ii) a lower chance level of 25%, in contrast to AX and ABX tasks with a 50% chance level. The target contrast is [k͡p-g͡b], with [s-r] and [t-j] contrasts serving as controls. For each contrast, there were 6 possible ordering of *different* trials (AAB, ABA, ABB, BBA, BAB, BAA) and 2 *same* trials (AAA and BBB). In a sound-attenuated booth, participants listened to 3 stimuli and pressed 1, 2, or 3 on a computer keyboard to indicate the position of the odd item or press X to indicate the three items were the same. The stimuli are C**V**C nonce words that appear in a carrier phrase, with the target consonant in the intervocalic position. Nonce words were used to avoid the effects of native participants invoking lexical over phonological representation. We hypothesize that the presence of voicing in [k͡p] may interfere with non-native speakers' ability to perceptually discriminate between [k͡p] and [g͡b]. Preliminary findings support this hypothesis. Results show that heritage speakers of Yoruba are less perceptually sensitive to the distinction between [gb] and [kp] than native speakers of the language. More specifically, the results of the descriptive statistics show that the NS group has higher accuracy values for [k͡p-g͡b] contrast (M = 0.94, SD = 0.05) than the HS group (M = 0.67, SD = 0.14). A two-tailed t-test for independent samples showed that the difference between NS and HS with respect to the contrast was statistically significant, $t(8) = 3.85$, $p = .005$, 95% confidence interval [0.11, 0.42]. The present outcomes tentatively indicate that the voicing characteristic in [k͡p] exerts an interference on the discrimination of [k͡p] and [g͡b] which will need to be mitigated in the process of acquiring Yoruba.

References

Connell, Bruce. 1991. Accounting for the reflexes of labial-velar stops. In Martina Rossi (ed.), *Proceedings of the Xllth ICPhS,* 3, 110-113. Aix-en-Provence: Institute of Phonetics.

Ladefoged, Peter. 1964. *A phonetic study of West African languages: An auditory-instrumental survey*. Cambridge: Cambridge University Press.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Ohala, John J. 1979. Universals of labial velars and de Saussure's chess analogy. In Eli Fischer-Jorgensen, Jorgen Rischel, & Nina Thorsen (eds.), Proceedings of the 9th International Congress of Phonetic Sciences, 2, 41-47. Copenhagen: Institute of Phonetics.

Strange, Winifred & Shafer, Valerie. 2008. Speech perception in second language learners: The re-education of selective perception. In Jette G. Hansen Edwards & Massimiliano I. Zampini (eds.), *Phonology and second language acquisition*, 153-191. Amsterdam: John Benjamins.

Left Edge Word Stress in Faifi Arabic

Abdullah Alfaifi (University of Bisha) and Stuart Davis (Indiana University
ahalfifi@ub.edu.sa and davis@indiana.edu

As noted by Watson (2011), primary stress in Arabic dialects references the right edge of the word, where, typically, stress falls on a final superheavy syllable or, lacking that, a heavy penult if there is one. Dialects differ in stress when the final two syllables are light; for example, if the antepenult is heavy, some dialects like Cairene stress the penult (e.g. [mak.'ta.ba] *'library'*) while others like Hijazi stress the antepenult (['mak.ta.ba] *'library'*). Further, dialects differ as to whether primary stress must fall on one of the last 3 syllables as in Hijazi (Bokhari 2020) or whether stress can be pre-antepenultimate as in San'ani ['ra.ga.ba.tih] *'his neck'* (Watson 2011). Dialects that reference the left edge of the word for primary stress are rare, an exception being Jewish Yemeni Arabic (Shachmon 2022) with word-initial stress that fluctuates. Another exception is stress fluctuation in postpausal position in San'ani (Watson 2002) where an initial syllable can receive primary stress irrespective of its weight or that of the following syllables as in ['ta.maam] *'okay'* and ['ka.tabt] *'I/you$_{M.S.}$ wrote'*. In non-post-pausal position, these words have stress on the final superheavy syllable. The focus of our paper is stress assignment in Faifi Arabic (FA), a group of dialects spoken in southwestern Saudi Arabia, which preserves a set of ancient features like the use of prefixal /m-/ as the definite article (e.g. [mgalam] '*the pen'*) and obligatory indefinite marker /-in/ as in [galam-in] '*a pen'*. Primary stress in FA seems word-initial by default whether the following syllables are light: ['ṭa.la.ʕa.ba.ha] *'he took her out'*; closed (even if closed by a geminate) ['na.baʃ.ʃa] *'you searched'*; or containing a long vowel ['śi.lii.hu] *'eat$_{Imp.\ F.S.}$ it$_{M.S.}$'*. This suggests that main stress in FA is determined from the left edge of the word making it rare among Arabic dialects. While initial stress is the most prominent pattern, there are specific cases where stress is attracted to other syllables. Cases include words starting with [ʔa] followed by a heavy syllable: stress falls on the 2nd syllable as in [ʔa.'taj.tim] *'you$_{PL.}$came'*, [ʔa.'ṭaʕ.nɑɑ.him] *'we obeyed them$_{M.}$'* (cf ['ba.naj.tim] 'you$_{PL}$ built'; and ['la.giṭ.nɑɑ.him] *'we picked them up'*). Similarly, initial [ʔVC] syllables are never stressed. Rather, stress falls on the 2nd syllable even when light, as in [ʔim-'ṭi.fil] *'the baby'*, [ʔaʕ.'fii.him] *'I pardon them'* (cf. ['gim.bi.la] *'a bomb'*, ['niʕ.fii.him] *'we pardon them'*). In trisyllabic words with an initial [Ca] syllable, stress falls on the 2nd syllable only when it is superheavy ([ma.'ṭɑɑʕ.min] *'restaurants'*, cf. ['da.ras.tim] *'you$_{PL}$ studied'*). In words longer than 3 syllables with an initial [Ca] syllable, stress falls on the 2nd syllable only if it is heavy or superheavy (e.g. [ma.'ʃɑɑ.jix.him] *'their$_{M.}$ sheiks'* and [na.'ʃart.bi.hin.na] *'I took them$_{F.}$ out'* (cf. ['na.ʃa.ra.him] *'he sawed them'*, ['sa.ma.kat.hin.ne] *'their$_{F.}$ fish'*). Stress fluctuation is observed in words with 5 or more syllables, where stress may fall on one of the 1st three syllables, depending on their shape, but not beyond the 3rd syllable (['naḍ.ḍɑɑ.rɑɑt.hin.na] - [naḍ.'ḍɑɑ.rɑɑt.hin.na] - [naḍ.ḍɑɑ.'rɑɑt.hin.na] *'their$_{F.}$ glasses'*. Thus in FA, the 3-syllable window is at the left edge of the word. We offer a foot structure analysis where moraic trochaic feet are built from the right edge of the word but where the leftmost foot is assigned main stress thus favoring word-initial stress. We suggest the tendency for word-initial syllables with glottal stops to repel stress relates to the epenthetic nature of the initial glottal stop. Stress is on the 2nd syllable in a word like [ma.'ṭɑɑʕ.min] because the initial syllable is unfooted when before a superheavy. Stress is initial on ['da.ras.tim] since the first two syllables form a moraic trochee given contextual coda weight (the 2nd syllable in ['da.ras.tim] is monomoraic) and final extrametricality. We connect the FA preference for word-initial stress with San'ani post-pausal initial stress noted by Watson, and to fluctuating initial stress in Jewish Yemeni Arabic (Shachmon 2022) contending that these are ancient patterns. In this way, the FA stress pattern can be considered archaic reflecting an ancient preference for initial stress.

Title: Suprasegmental information influences adaptation to speech in children and adults
Authors: Ruth M. Altmiller and Kristin J. Van Engen
Emails: ruth.altmiller@wustl.edu, kvanengen@wustl.edu
Affiliation: Washington University in St. Louis

In this project, we expand upon previous work showing that adults expend more listening effort when listening to L2-accented speech (i.e., speech produced by an individual using their second language) compared to L1 speech (McLaughlin and Van Engen, 2020), and that that difference in effort diminishes over time (Brown et al., 2020). Here, we use task-evoked pupillary response to ask whether children (n = 35, ages 5;0 – 8;11) show similar patterns of listening effort and adaptation for L1- and L2- accented speech as adults, and investigate whether acoustic-phonetic measures of prosody predict listening effort and adaptation to L1 and L2 accents in both children and adults. Using Growth Curve Analysis, we found that, while children initially recruit extra cognitive resources for L2-accented speech, this difference diminished over the course of the experiment to the degree that accent condition did not predict pupil response in later trials, indicating that children adapt more completely to L2-accented speech than do adults. Additionally, we found that three prosodic measures (relative word duration, pitch stability, and pitch range) were related to speech adaptation in both age groups, but with opposite directionality for the L1- and L2- accent conditions. We suggest that differences in background knowledge regarding accent and socio-indexical features of speech may play a role in age-related differences in listening effort. Furthermore, our results indicate a need models of speech perception and adaptation to include a role for suprasegmental information.

Cross-linguistic development of laryngeal gestures in acquisition

Luke Annear (annear@wisc.edu), UW – Madison
Joe Salmons (jsalmons@wisc.edu), UW – Madison
Charlotte Vanhecke (cvanhecke@wisc.edu), UW – Madison

Despite progress in understanding patterns of phonological development in diverse languages, the acquisition of laryngeal gestures and contrasts remains poorly understood. McLeod and Crowe (2018), for example, summarize patterns in consonant acquisition in 27 languages, but group languages such as English and French together as each having stop obstruent series /b,d,g/ and /p,t,k/ despite significantly different laryngeal phonetics and developmental patterns (see e.g. MacLeod et al. 2011). We argue that this conflation of differing laryngeal systems obscures cross-linguistic patterns in the development of laryngeal contrasts. The present study explores the development of such patterns in systems with two-way contrasts conventionally represented with <b, d, g> and <p, t, k> in the roman alphabet.

Adopting Laryngeal Realism (Iverson and Salmons 1995) and privativity (Avery and Idsardi 2001), we understand two-way laryngeal contrasts in terms of unspecified and specified pairs. Some languages specify [Glottal Width] ([GW], or aspirated) members and others specify [Glottal Tension] ([GT], or voiced) members. This understanding provides new insight into production patterns and order of acquisition. For example, Dutch (like French) specifies [GT] on /b/ and no laryngeal specification on /p/. English on the other hand specifies [GW] on /p/ and no specification for /b/. Dutch children tend, early on, to show patterns of devoicing initial /b, d/ (van Haaften, 2020), whereas English children tend to exhibit patterns of deaspiration ("voicing") of /p, t, k/ (Smit, 1993). This suggests that children first produce laryngeally unspecified obstruents (Dutch /p/, English /b/) before later mastering specified obstruents for their language (Dutch /b/, English /p/).

Initial data from 8 languages reviewed in Crowe and McLeod (Belgian Dutch, Italian, Portuguese, Quebecois French with [GT] and Cantonese, English, Icelandic, Mandarin with [GW]) indicate that viewing laryngeal contrasts in terms of unspecified/specified pairs better predicts age of acquisition (below right, $F(1,43) = 3.669$, $p = .06$, $R^2 = .078$) than conventional voiced/voiceless classification (below left, $F(1, 43) = 0.14$, $p = .71$, $R^2 = 0.003$).

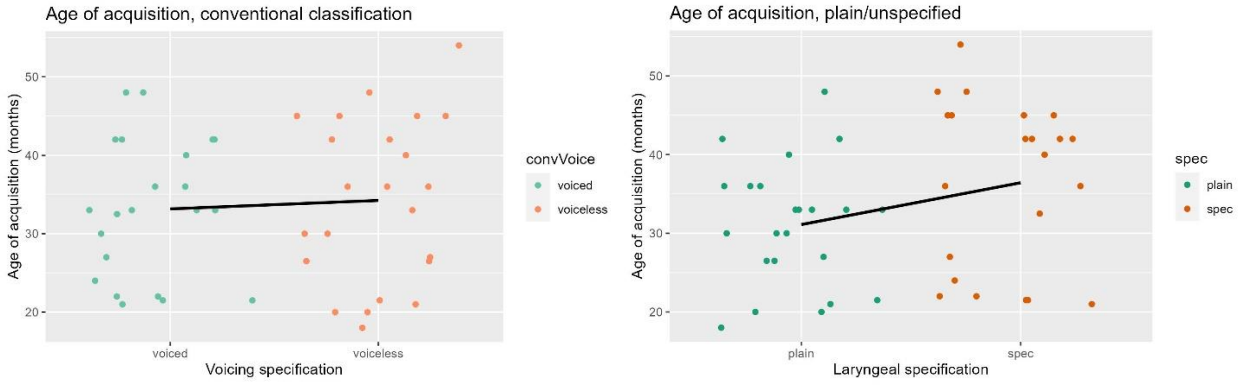


Figure 1 – Age of acquisition of initial stop consonants with conventional voicing classifications (left) compared to unspecified/specified classifications (right).

Perception of Spanish Intonation Contours of Declarative sentences by L2 speakers and Spanish speakers of other varieties.

Izaro Bedialauneta Txurruka
University of Illinois at Urbana Champaign
izarobe2@illinois.edu

Spanish declarative questions and statements often contrast only in their prosody, e.g. "¿Bebe agua?" *does he/she drink water* and "bebe agua" *he/she drinks water* are lexically and syntactically the same, but they differ in the intonation. These intonation contours, however, vary depending on the Spanish-speaking country and region. For example, in Castilian (CS), neutral questions usually end with a rise (H%) (Figure 1), whereas in Buenos Aires Spanish (BAS) they most commonly end with a circumflex contour (HL%) (Figure 2). Declaratives in both varieties end with a fall (L%) (Sosa 1999, Hualde, 2003, Gabriel et al. 2010, Estebas-Vilaplana et al., 2010). Dominican Spanish (DS) shows cross-linguistically fewer common patterns. In DS there is a H% boundary tone in statements (Figure 4) and a L% boundary tone in questions (Willis, 2010; Hualde et al., 2015) (Figure 3).

We report on an experiment where participants were auditorily presented with statements and questions produced by speakers of the three Spanish varieties just described. A total of 60 audio files were presented. 20 stimuli consisted of sentences containing two accentual phrases. In addition, the experimental stimuli included sound files containing either the first or the second accentual phrase of the complete sentences (20 examples of each). Listeners were asked to click 'yes' or 'no' to answer to the question 'is this a question?' after hearing each stimulus in Qualtrics. 18 L2 speakers with American English as their L1, and 24 native speakers (9 from Buenos Aires, 7 from Dominican Republic and 8 from Spain) participated in the experiment. Previous studies on the perception of Spanish intonation contours by L1 American English speakers have shown that perception of declarative question contours to be specially challenging (Brandl et al. 2020, Trimble, 2013, Casillas et al., 2022), but some contours are more problematic than others. H% has been found to be the easiest boundary tone to interpret as question, followed by HL%, with L% ending the hardest (Casillas et al., 2022). We hypothesized that participants would be guided by their L1 and in cases where a contour is absent in their L1, they would rely on universal tendencies. DS statements and questions were thus predicted to be particularly difficult to identify. Regarding incomplete sentences, our hypothesize was that the final contour would be easier to identify than the first half, but that a relatively high beginning might also be sufficient for interrogativity to be conveyed (Face, 2007).

A binomial logistic regression showed a significant difference between participant groups. L2 speakers were not as accurate as the other groups. Moreover, as predicted, DS sentences were significantly the most challenging ones for our listeners, whereas there was no difference in accuracy in the perception of CS and BAS sentences. Since the circumflex interrogative contour of BAS is not part of the American English inventory, we hypothesize that L1 English participants relied on universal tendencies for the identification of this contour (i.e. the presence of a high tone associated with the last accent). Lastly, participants performed significantly differently depending on whether they heard just the initial accentual phrase, the final accentual phrase, or the whole sentence, with whole contours being better identified than incomplete sentences.

References:

Brandl, A., González, C., & Bustin, A. (2020). The development of intonation in L2 Spanish: A perceptual study. *Hispanic linguistics: Current issues and new directions*, 12-31

Casillas, J. V., Garrido-Pozú, J. J., Parrish, K., Arroyo, L. F., Rodríguez, N., Esposito, R., & Taveras, K. (2022). Using intonation to disambiguate meaning: The role of empathy and proficiency in L2 perceptual development.

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. *Transcription of intonation of the Spanish language*, 17-48.

Face, T., (2007). The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de fonética experimental*, 186-225.

Gabriel, C., Feldhausen, I., Pešková, A., Colantoni, L., Lee, S. A., Arana, V., & Labastía, L. (2010). Argentinian spanish intonation. *Transcription of intonation of the Spanish language*, 285-317.

Hualde, J. I. (2003). El modelo métrico y autosegmental. *Teorías de la entonación*, *155*, 184.

Hualde, José I. & Pilar Prieto. 2015. Intonational variation in Spanish: European and American varieties. In Sónia Frota & Pilar Prieto (eds.), Intonational Variation in Romance, 350- 391. Oxford: Oxford University Press.

Sosa, J. M. (1999): La entonación del español, Madrid, Cátedra.

Trimble, J. C. (2013). Perceiving intonational cues in a foreign language: Study abroad and its effect on the perception of sentence type in two dialects of Spanish. In C. Howe, S. Blackwell, & M. Quesada (Eds.) Selected proceedings of the 15th Hispanic Linguistics Symposium. Somerville, MA: Cascadilla Proceedings Project.

Willis, E. (2010). Dominican Spanish intonation. *Transcription of intonation of the Spanish language*, 123-153

Figures:

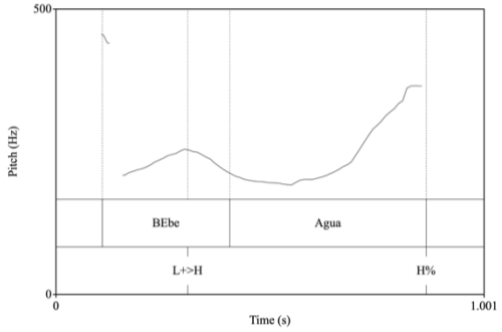


Figure 1: CS question contour

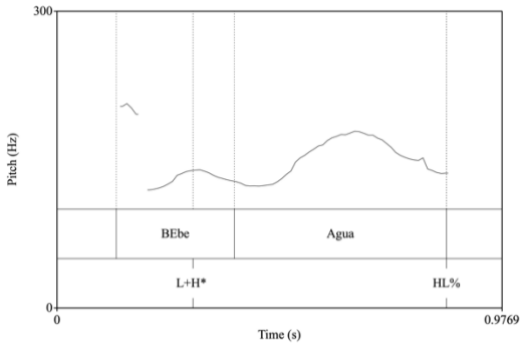


Figure 2: BAS question contour

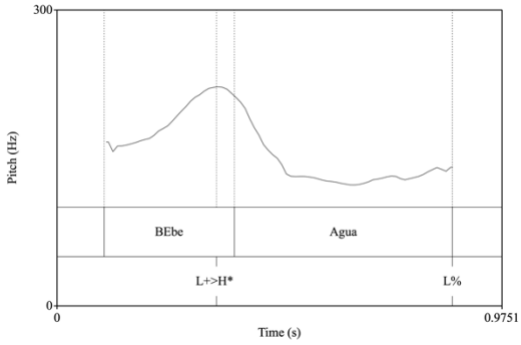


Figure 3: DRS question contour

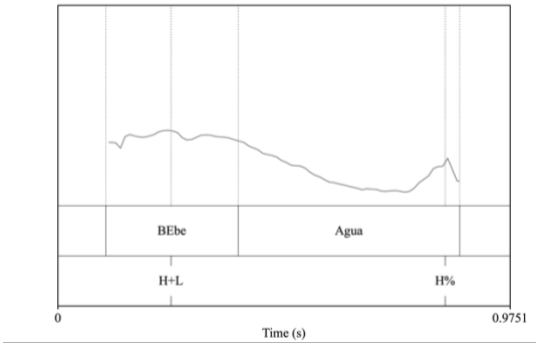


Figure 4: DRS statement contour

Ander Beristain, Ph.D. | Saint Louis University | ander.beristain@slu.edu

Asymmetrical control of aerodynamic and temporal features of stops in heritage speech

Background. The production of stop consonants (SCs) involves a closure in the vocal tract, building up air pressure, and releasing that closure (Stevens, 1993). Thus, acoustic (temporal) and articulatory (aerodynamic) phonetic details are implemented in interarticulatory timing. Moon and Folkins (1991) find that, in speech motor control, speakers may adjust acoustic and aerodynamic variables independently based on the target sound and its perceptual similarity with phonological neighbors. One of the most widely used variables in SC research is VOT, and it is known that world languages vary depending on the range of this feature. For instance, Spanish has a short VOT and is unaspirated, and English has an aspirated and longer VOT in stressed syllables (Amengual, 2012). Whereas the acoustic properties of SCs have been thoroughly investigated, aerodynamic studies (which pertain to the intrinsic nature of SC production) are scarcer. Cho et al. (2002) have shown the correlation between aspirated-unaspirated stops and greater-lesser airflow degree, respectively. However, no study has provided a simultaneous aerodynamic and acoustic description through a (heritage) bilingual lens, where it has been proposed that early bilingualism is more conducive towards cross-linguistic phonetic system adjustability (Flege & Bohn, 2021).

Research Question and Hypotheses. (RQ1) Can bilingual speakers adjust acoustic-aerodynamic features cross-linguistically? (RQ2) Are aerodynamic and acoustic variables controlled comparably? Based on previous research, (H1) cross-linguistic adjustment is expected in early bilingual speakers, and (H2) due to the intrinsic nature of SCs, the temporal (acoustic) properties should correlate with physiological (aerodynamic) characteristics.

Methodology. 16 second-generation heritage speakers of Spanish (HSS) participated in two read-aloud tasks in Spanish and English. Additionally, 10 adult L1 Spanish (L1Sp) and 16 adult L1 English (L1En) speakers served as control groups in their L1s. Simultaneous acoustic and aerodynamic data were collected using a head-mounted microphone and two pressure transducers connected to a vented mask. Each speaker produced real words with a stressed-syllable initial /t/ (n=40 in Spanish, n=50 in English). VOT was inspected in each SC. For the aerodynamic analysis, that region was further divided into 10 equidistant points to observe its dynamics. The acoustic analysis included 1,040 tokens in Spanish and 1,600 in English, while the aerodynamic analysis included 10,400 datapoints in Spanish and 16,000 in English. The acoustic data were fit into linear mixed-effect models with VOT duration as dependent variable, while the aerodynamic data used GAMMs for the dynamic analysis of oral airflow. All models included GROUP as fixed factor, and SPEAKER and WORD as random factors. GAMMs also had smooth terms and factor smooths.

Results. HSS exhibited significant adjustability cross-linguistically in both domains. Regarding ACOUSTICS, HSS had significantly shorter VOTs as opposed to L1En in English (β=9.84, t(40)=2.96, p<.01), while their values resembled those of L1Sp in Spanish (p=.82) (see Table 1). Regarding AERODYNAMICS, HSS behaved like L1En speakers in English (β=0.02, t=.70, p=.48) (see Figure 1) and like L1Sp speakers in Spanish (β=-0.003, t=0.15, p=.87) (see Figure 2).

Conclusion. The current study presents novel data in which HSS control acoustic and aerodynamic variables independently. The overcompensation of greater oral airflow in a shorter durational period may be due to the perceptual saliency of aspiration in contrast to its durational cues.

Ander Beristain, Ph.D. | Saint Louis University | ander.beristain@slu.edu

Table 1. VOT durations (sd) in ms

	English	Spanish
L1En	69 (17)	--
HSS	64(20)	17(4)
L1Sp	--	18(5)

Figure 1. Aerodynamic results for English (left: GAMM curve; right: difference curve)

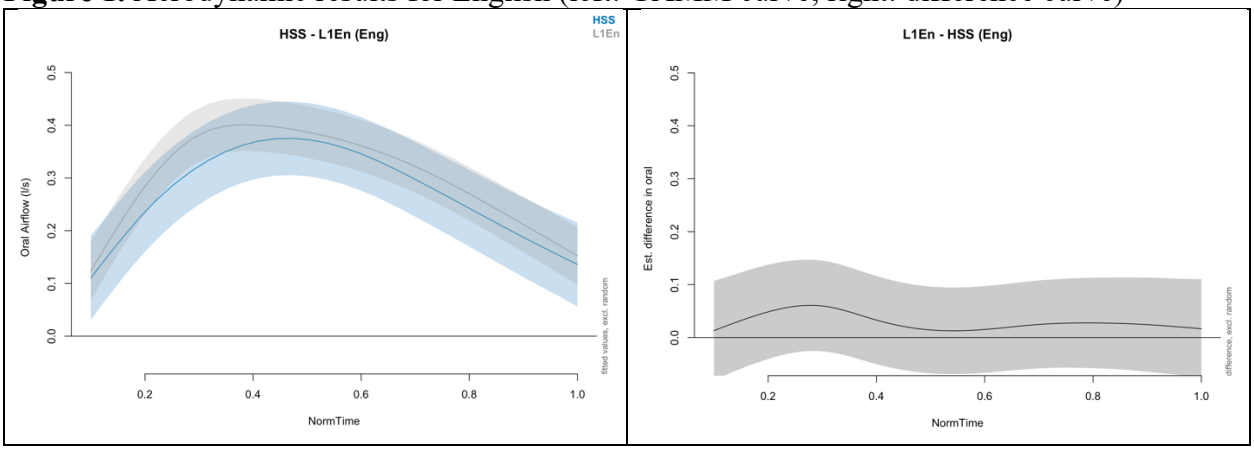
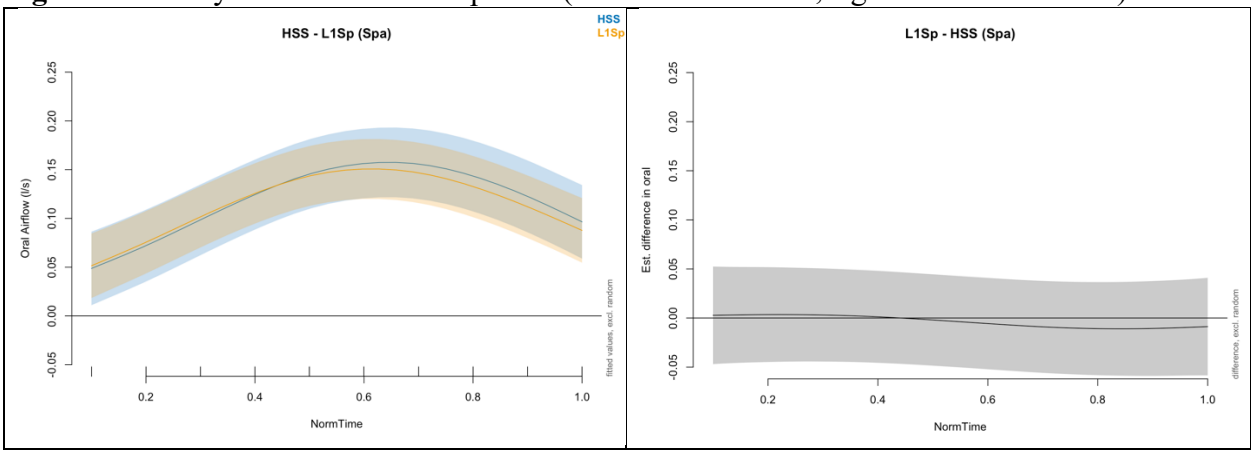


Figure 2. Aerodynamic results for Spanish (left: GAMM curve; right: difference curve)



Maximal Contrastivity in Swedish: How minimal is phonological contrast?

David Bolter, Denison University
bolterd@denison.edu

The term MAXIMAL CONTRASTIVITY has been applied in the literature to describe the contrast between fortis and lenis stops in Swedish, where the contrast is said to involve both [voice] and [spread glottis]. Herein, I broaden this term to two other areas of Swedish phonology and discuss the implications this has for phonology.

The debate surrounding Maximal Contrastivity, as I term it, is hardly new. For example, Hockett [1] writes about the difficulties of and ultimately, the impossibility of dividing phonological constituents into those that are 'determining' and those that are 'determined'. Such a view fundamentally challenges the notion of DISTINCTIVE FEATURES, at least in the sense that they are conceived in works such as Jakobson et al. [2] or Chomsky & Halle [3]. Yet, distinctive feature theory persists, with more recent debates centering on the relationship between PERCEPTUAL CUES and DISTINCTIVE FEATURES. Wang & Bilger [4] have argued that perceptual evidence calls into question the existence of distinctive features. Nonetheless, Parker [5] and many others have suggested (or simply tacitly assume) that the notion of distinctive features is a phonological rather than a perceptual one, with voicing contrasts like those described above occupying the perceptual/ phonetic domain and therefore not in a way contradicting the phonological notion of distinctive features.

In Standard Swedish, Helgason & Ringen [6] and others have argued that the Swedish fortis/ lenis contrast (viz. /tʰuːg/ 'took' vs. /duːg/ 'died') is maximally contrastive in that both voicing and aspiration are active. As Riad [7] and others have noted this makes Swedish unique amongst the Germanic languages, although it may be shared with Southern American English [8]. Yet, it seems that maximal contrastivity can be found in other domains of Standard Swedish phonology. Swedish, like other Germanic languages, contrasts a series of long, tense vowels with a series of short, lax vowels (viz. /viːn/ 'wine' vs. /vɪnː/ 'win!'). There is considerable debate about the modelling of the tense / lax contrast in Germanic languages. For Swedish, Behne et al. [9] have argued that length is the most crucial cue. However, in addition to length and tenseness, it has been observed that Swedish long, tense vowels also have a marked diphthongization (see Ekland & Traunmüller [10]). Thus, the so-called tense vowels of Swedish are distinguished by (at least) three properties: vowel quantity, vowel quality, diphthongalness. Perhaps, it is the case that a perceptual study might reveal that one is given a bigger weight than the other two, but many Swedish speakers realize all three. Finally, I consider the contrast between Accent 1 and Accent 2 in Swedish (viz. /¹anːdɛn/ 'the duck' vs. /²anːdɛn/ 'the spirit'). Although the contrast is usually discussed in terms of tonal contour differences between the two categories, Accent 2 words also differ from Accent 1 in having a root that is bisyllabic. Thus, one may also ask whether the contrast is fundamentally one of bisyllabicity or one of tone. In these three areas, it is difficult to choose one phonological feature as primary and another as secondary or epiphenomenal and we may surmise that similar facts can be found in a great many languages.

To conclude, I offer a potential solution to the debate on cues vs. distinctive features. As I see it, one of the fundamental problems in this debate is the fact that the terminology used for the two domains is often one and the same. For example, [voice] might be used in an abstract phonological or as phonetic term referring to vocal cord vibration. If one is serious about maintaining a distinction between cues and features, perhaps one solution would be to change the vocabulary of the distinctive features to avoid overlap with perceptual cues. Thus, instead of using [palatalization], one could call this *soft* / *hard*. From a distance, this might seem like a radical approach, but at a practical level many phonologists already do this. For instance, many phonologists use fortis as a shortcut for /p t k/ in a voicing system and /pʰ tʰ kʰ/ in an aspirating system. In this regard, I make common cause with advocates of SUBSTANCE FREE PHONOLOGY e.g. Blaho [11].

References

[1] Hockett, C.F. (1955). *Manual of phonology* Baltimore: Waverly Press.

[2] Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis : the distinctive features and their correlates*. Cambridge, Mass.: M.I.T. Press.

[3] Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

[4] Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *The Journal of the Acoustical Society of America*, *54*(5), 1248-1266.

[5] Parker, F. (1977). Distinctive features and acoustic cues. *The Journal of the Acoustical Society of America*, *62*(4), 1051-1054.

[6] Helgason, P., & Ringen, C. 2008. Voicing and aspiration in Swedish stops. *Journal of phonetics*, *36*(4), 607-628.

[7] Riad, T. (2014). *The phonology of Swedish.* Oxford, United Kingdom: Oxford University Press.

[8] Hunnicutt, L., & Morris, P. A. (2016). Prevoicing and aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics*, *22*(1), 24.

[9] Behne, D., Czigler, P., & Sullivan, K. P. (1997). Swedish quantity and quality: a traditional issue revisited. *Reports fron the Department of Phonetics, Umeå University*, *4*, 81-83.

[10] Eklund, I., & Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica*, *54*(1), 1-21.

[11] Blaho, S. (2008). *The syntax of phonology: A radically substance-free approach*. University of Tromsø: Dissertation.

Mora Theory explains what X-Slot Theory cannot: false onsets in Slovak

Scott Borgeson, Michigan State University
borges11@msu.edu

Mora Theory and X-Slot theory are two competing views of syllable structure that are theoretically distinct, but notoriously difficult to tease apart empirically, as both can capture the general facts of syllable structure with equal degrees of stipulation. However, the two differ in one key respect: in Mora Theory, it is theoretically possible to disentangle the notion of a segment's phonological weight from that of its position within the syllable. This is because weight is a matter of linkage to moras, whereas position within the syllable can be defined by other means, such as association with an Onset (o) or Nucleus (ν) node (a modification to the original theory that, I argue, is needed anyway to capture the existence of moraic onsets). By contrast, in X-Slot Theory, the notions of position and phonological weight are inseparable: all segments are linked to X-slots, and if those X-slots are in turn linked to the nucleus (N), then they will by definition contribute weight to their syllable – thus, the notion of non-weightful nucleus is a contradiction in terms.

In this talk, I will show that weightless nuclei are indeed attested in Slovak, and therefore that Mora Theory is empirically superior to X-Slot Theory. In Slovak, vowels are shortened after the palatal glide /j/; the resulting syllable counts as light for the purposes of Slovak phonology (namely, it fails to trigger the well-known Rhythmic Law, a ban on two adjacent heavy syllables). I argue that this phenomenon is best captured as follows. First, note that /j/ is universally disfavored as an onset owing to its high sonority. Slovak resolves the issue by moving /j/ out of onset position and into the nucleus (i.e., it disassociates it from the onset node and associates it with the nucleus node). There, it remains non-moraic and does not contribute weight to the syllable. At the same time, syllable nuclei in Slovak are maximally binary-branching, so /j/ cannot co-exist with two moras in the same nucleus (i.e., it cannot be followed by a long vowel); instead, one of them must delete (i.e., the vowel must shorten). This analysis relies on the possibility of /j/ being a non-moraic nucleus consonant, which I term a *false onset*: we know it must be non-moraic because the resulting syllable counts as light, and that it must be nucleic because it competes for space with moras. This configuration is possible under the version of Mora Theory advanced here, where moraicity and segment position are distinct, but *not* under X-Slot Theory, where all segments necessarily bear X-slots, and X-slots within the nucleus necessarily count for weight.

Lax mid vowels in Eastern Andalusian Spanish: Data from perception

Maialen Casquete de la Puente, University of Illinois at Urbana-Champaign
maialen2@illinois.edu

Eastern Andalusian Spanish (EAS) is characterized by processes of vowel laxing and harmony when word-final consonants are lenited. However, few studies address categorical perception of the resulting vowels. Studies that do are limited to paroxytonic words with /s/ in coda (e.g. ['nɛnɛ] *nenes* 'kids'), which is the only context having been described as triggering laxing (Corbin, 2006; Jiménez & Lloret, 2007; Herrero de Haro, 2017; Henriksen, 2017)

This study addresses perception of the lax and tense vowels in EAS by speakers of the dialect, and North Central Peninsular Spanish (NCPS) speakers. For the stimuli, I recorded a female EAS speaker pronouncing 58 oxytonic words. Words either ended in <-r> or <-s>, where the final consonant had been lenited (*beber* and *bebés*, [bɛ'bɛ]), or in a bare, stressed vowel (*bebé* [be'be]). The underlying consonant and vowel stimuli are considered lax [ɛ] (F1 ~750Hz, F2 ~1850Hz) and tense [e] (F1 ~480Hz, F2 ~2500 Hz) respectively. That is, [ɛ] was lower, and more retracted. Similarly, lax [ɔ] (e.g. *color*) had an F1 of ~720Hz, and [o] (e.g. *coló*) had an F1 of ~440Hz, meaning the latter was higher. Regarding differences between consonantal endings <-er> and <-es>, lenition on the latter seemed to leave a longer aspiration following the vowel (~30ms). F2 was also slightly higher (+~150Hz).

The experiment was carried out in Qualtrics. A total of 19 EAS and 19 NCPS speakers participated in the study. 29 experimental words (+14 fillers) were visually presented to participants. After reading each word on the screen, participants listened to two recordings, from four possible different contrasts: 1. <-er> vs. <-es> (e.g. *beber* vs *bebés*), 2. <-e> vs. <-er> (*bebé* vs *beber*) , 3. <-e> vs. <-es> (*bebé* vs *bebés*), 4. <-o> vs <-or> (*coló* vs. *color*). Lastly, participants had to choose which one of the two audios best represented the written stimulus.

For data analysis, binomial tests were conducted for each condition in order to compare results to a random chance distribution. EAS participants selected the target recording in Conditions 2-4 (grouped as *consonant* vs. *vowel*). They were able to identify the recording in 91.7% of the trials (p. value <.001). However, they had trouble distinguishing Condition 1 (underlying <-er> or <es>), which they identified in 58% of the cases (p=.08). The difference in aspiration and F2 for the <-er> ending was unsufficient to perceive both lax vowels as categorically different. NCPS speakers performed above chance in Conditions 2 and 3 ([ɛ] vs. [e]), succeeding in 69.5% of the trials (p<.001), as well as in Condition 4 ([o] vs. [ɔ]), 61% (p=.007). As expected, they did not perform successfully in Condition 1. Results are presented in Tables 1 and 2.

Written stimuli alternated between the two possible choices. It is particularly salient that NCPS participants, unfamiliar with the EAS contrast tested, selected the target recording when the written stimulus ended in a consonant, but not a vowel. When they saw a word ending in a vowel (e.g. *canté, cantó*), they selected the recordings for tense [e] and [o] in 55% of the trials, which was not sufficient to argue for a non-random chance distribution (p=.098). We note that the final vowels of recordings with no underlying consonant were produced as higher vowels than one typically finds in NCPS (Chládková, Escudero & Boersma, 2011; Cervera, Miralles & González-Àlvarez, 2001). That is, both open and tense vowels may have been unfamiliar to our NCPS listeners.

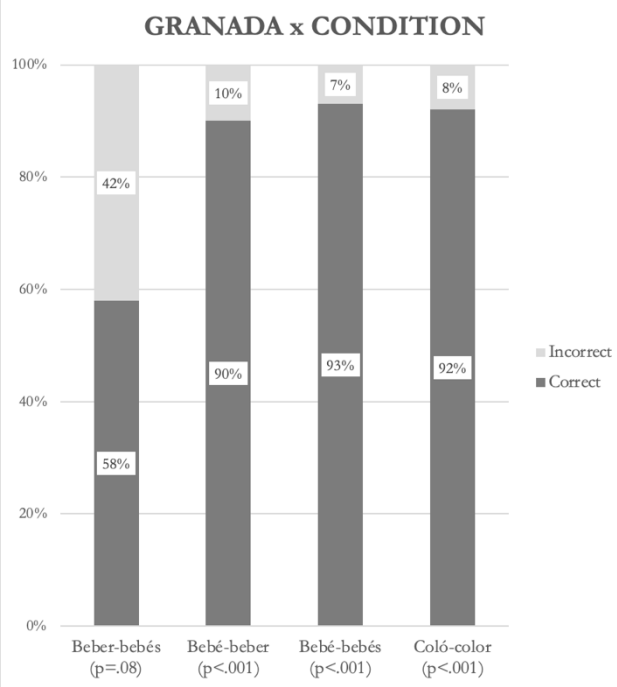


Table 1: Correct/incorrect responses by EAS speakers, by condition

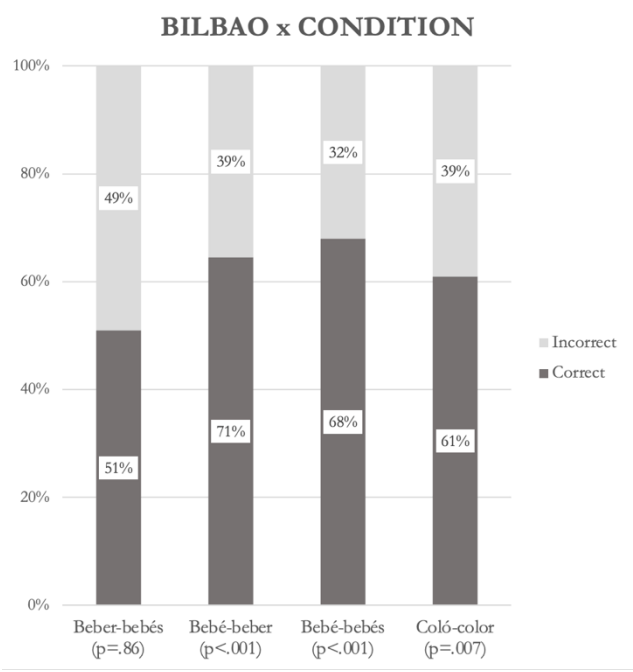


Table 2: Correct/incorrect responses by NCPS speakers, by condition.

References

Cervera, T., Miralles, J. L., & González-Àlvarez, J. (2001). Acoustical analysis of Spanish vowels produced by laryngectomized subjects. *Journal of Speech, Language, and Hearing Research, 44,* 988-996.

Chládková, K., Escudero, P., & Boersma, P. (2011). Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *The Journal of the Acoustical Society of America*, *130*(1), 416-428.

Corbin, L. (2006). The phonetics and phonology of s-lenition and vowel laxing in eastern Andalusian Spanish. *Williams College, Williamstown, MA, BA honors thesis*.

Henriksen, N. (2017). Patterns of vowel laxing and harmony in Iberian Spanish: Data from production and perception. *Journal of Phonetics*, *63*, 106-126.

Herrero de Haro, A. (2017). The phonetics and phonology of Eastern Andalusian Spanish: A review of literature from 1881 to 2016. *Íkala, revista de lenguaje y cultura*, *22*(2), 313-357.

Jiménez, J., & Lloret, M. R. (2007). *Andalusian vowel harmony: Weak triggers and perceptibility* [Conference presentation]. Treball presentat a la Old-World Conference in Phonology, Rhodes, TN, United States.

Title: An OT Approach to the Inventory of Mid Vowels in Taiwan Mandarin

Pin-Yu Chen
Indiana University Bloomington
pinychen@iu.edu

Abstract:

This study focuses on the mid vowels in Taiwan Mandarin Chinese using Optimality Theory (OT) (McCarthy & Prince, 1993, 1994, 1995; Prince & Smolensky, 1993). Taiwan Mandarin (TM) is spoken in Taiwan, and the pronunciation and lexical usage are very similar to Standard Mandarin but different in many ways (e.g., [fəŋ] does not exist in TM but in SM). Several studies have discussed the vowels in Mandarin, and the discussion of underlying vowel phonemes and the surface representation of these phonemes have not reached a definitive conclusion (C. Cheng, 1973; R. Cheng, 1966; Lin, 1989; Wan & Jaeger, 2002; Wu, 1994). The fact that Mandarin uses a logographic writing system makes it more challenging to argue for the selection of the underlying representation (UR). In the surface representation, TM has six mid vowels, [e], [ɛ], [ə], [o], [ɔ], and [ɤ]. Previous studies mentioned above have proposed that either /ə/ or /ɤ/ is the UR of these vowels. This study examines whether OT can explain the distribution of surface representations assuming schwa as the underlying phoneme. Among all studies, I follow the analysis of Wan & Jaeger (2003) which chose /ə/ as the UR and was coined within the Feature Geometry approach and utilized speech errors from corpora to analyze the distribution of the vowels. Speech errors could give us evidence of how a vowel alternates in different contexts, which cannot be captured by the normal distribution of the SR but provides good sources for arguing the existence of constraints. The OT approach can further explain the phonological process. The constraints of mid vowels in TM are very sensitive to contextual markedness, and constraints related to assimilation rank higher than the other constraints, such as faithfulness constraints. **AGREE_CPLACE** constrains the vowel to agree with its following glide and consonant (nasal) places, including rounding, RTR, and backness. **AGREE GPLACE_** confines vowels to agree with the place of its preceding glide. **Front Laxing** shows that vowels need to be laxed following a front vowel/glide. **IDENT RD** and **IDENT Bck** retain the place of the vowel to the UR. **+RTR$** constrains the vowel to be a lax vowel when it is at a syllable-final. **IDENT RTR**, on the other hand, wants the vowel to pertain to its original RTR. The constraints related to the coda assimilations outrank assimilations to the onset, and glides are more influential to consonants, which predicts the direction of the assimilation. The ranking proposed is as follows: AGREE_CPLACE>>AGREE GPLACE_, Front Laxing>>IDENT RD, IDENT Bck, +RTR$>>IDENT RTR The results of this study allow us to have a better understanding of how vowels in Mandarin change and might also benefit L2 studies to see if L2 speakers transfer the constraints to their interlanguage. Further work is needed to account for the distribution of the high and low vowels.

Reference

Cheng, C. C. (1973). *A Synchronic Phonology of Mandarin Chinese*, Mouton, The Hague.

Cheng, R. L. (1966). Mandarin phonological structure, *Journal of Linguistics*, *2*(2), 135–262.

Lin, Y. H. (2007). *The sounds of Chinese with audio CD (Vol. 1)*. Cambridge University Press.

McCarthy, J., & Prince, A. (1993). *Prosodic morphology I: Constraint interaction and satisfaction*. Unpublished manuscript, University of Massachusetts at Amherst and Rutgers University, New Brunswick, NJ.

McCarthy, J., & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. *Proceedings of the North East Linguistics Society*, *24*, 333–379.

McCarthy, J., & Prince, A. (1995). Faithfulness and reduplicative identity. In J. Beckman, L. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory*, *University of Massachusetts Occasional Papers*, *18*, 249–384.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Unpublished manuscript, Rutgers University, New Brunswick, NJ.

Wan, I. P., & Jaeger, J. J. (2003). The phonological representation of Taiwan Mandarin vowels: A psycholinguistic study. *Journal of East Asian Linguistics*, *12*(3), 205-257.

Wu, Y. (1994). *Mandarin Segmental Phonology*, PhD dissertation, University of Toronto.

Cue re-weighting in the variation of vowel length contrasts in Daiya

Xiaobei Chen
Department of Linguistics, The University of Chicago
xiaobeichen@uchicago.edu

Cue re-weighting can lead to two significant outcomes: (i) the phonologization of redundant acoustic cues alongside the reduction of co-varying primary cues, as illustrated by Kiparsky's (1995) 'priming effect'; and (ii) the disappearance of the original contrast due to the loss of the primary cue, when the redundant cues lack sufficient saliency to distinguish minimal pairs. These variations can co-occur within the same type of phonological contrasts under distinct phonological conditions.

In Daiya, a Southwest Kam-Tai language, the contrast between diphthongs /V_1:V_2/ and /V_1V_2:/ (where V_1 = /a/) relies on two main cues, relative length patterns and vowel formants, as revealed by acoustic analyses. Perception experiments involving the same group of participants show that native speakers, especially younger individuals, tend to identify /V_1V_2:/ using spectral features as the salient cue. In comparison, the /V_1:V_2/ diphthongs are more susceptible to length: When V_2 is /u/, the duration ratio of the two vowels proves pivotal. Conversely, with /i/ as V_2, the formants play a more significant role, and the saliency is influenced by initial types.

Although the relative vowel durations were historically regarded as a robust cue in Kam-Tai languages, the primary cue in Daiya has shifted over time, with formants outweighing duration ratios. The /aːi/-/aiː/ distinction persists, suggesting a tendency among the younger generation to shift towards /ai/-/ei/, thereby transforming distinctive features into vowel qualities. On the other hand, the /aːu/-/auː/ contrast might fade, merging into the same rhyme /au/. These hypotheses align with Luo's (1984) observations on vowel length changes in other Dai dialects.

Notably, production and perception changes manifest asymmetry. Perceptual variations demonstrate greater internal consistency. In articulation, both formants and duration cues of /aːu/-/auː/ are diminishing, whereas the formants of /aːi/-/aiː/ are gaining prominence. In the realm of perception, formants are increasingly emphasized, regardless of the V_2 distinction. Given that the production variations are more powerful in this case, the perceptual cue re-weighting cannot trigger the 'priming effect' in the /aːu/-/auː/ distinction.

References:

Kiparsky, Paul. "The phonological basis of sound change." *Handbook of phonological theory*, edited by John A. Goldsmith, Basil Blackwell, 1995, pp. 640-670.

Luo, Meizhen 罗美珍. "傣语长短元音和辅音韵尾的变化" [The Variation of Long/Short Vowels and Consonant Codas in Dai]. *民族语文 Minority Languages of China*, vol. 6, 1984, pp. 20-25.

Tonal polarity in Hadza nominal morphology

Jeremy Coburn, Indiana University

jecoburn@iu.edu

Abstract:

Tonal polarity is a process in which a given tone-bearing unit (TBU), typically of an affix, is assigned a tonal value which is opposite to an immediately adjacent tone. Examples of tonal polarity are primarily attested in African languages, including languages of the Afro-Asiatic, Niger-Congo, and Nilo-Saharan language families. However, non-African languages have also been analyzed with tonal polarity, e.g., Mundurukú (Picanço 2002), Thadou (Hyman 2007), Tenyidie (Meyase 2021). Phonologists have long debated whether tone polarity is an actual process whereby a tonally-unspecified TBU is assigned a tonal value, or rather a dissimilatory epiphenomenon related to other linguistic processes (e.g., Obligatory Contour Principle). Hadza—a language isolate spoken by ~1,200 people (Brian Wood, pc.) in Tanzania—has been described as having two contrastive tone levels, H vs. L (Sands 2013). The mora has been identified as the TBU (Coburn et al., forthcoming), but relatively little is known about the tonal phonology of the language. Based on fieldwork conducted in Tanzania in 2022, this study finds that some nominal suffixes are underlyingly toneless /∅/ and undergo tonal polarity. The plural gender suffixes /-beʔe/ and /-biʔi/ and the plural copula suffixes /-pʰeʔe/ and /-pʰiʔi/ are assigned a tonal value opposite to the root-final tone, with a few exceptions. These results support the classification of tonal polarity as a case of underspecification rather than dissimilation. Hadza is, therefore, more accurately described as contrasting H vs. L vs. ∅.

The rhythmic pattern of Vietnamese quadrisyllabic reduplicative words: An investigation into their syllable duration

Phuong Dang, The Ohio State University, dang.275@osu.edu

There is an ongoing debate whether Vietnamese, a tonal monosyllabic language, provides evidence of prosodic structure above the level of the syllable. On the one hand, some studies support the argument that Vietnamese appears to have a flat prosodic structure or at least lacks specific prosodic levels. The claim that each Vietnamese syllable has equal energy (Emeneau, 1951) and the finding that there is no acoustic prominence in disyllabic words with different morphosyntactic structures in utterance-medial and final contexts (Brunelle, 2017) may imply a flat structure in the language. Relatedly, Schiering et al. (2010) claim that the levels of prosodic words and feet are skipped in Vietnamese, and that phonological phrases dominate syllables directly. On the other hand, the description of different levels of stress in Vietnamese in a pause group by Thompson (1965) and Tran (1967) supports the presence of some rhythm in the language even though it is not clear whether a pause group is equal to a prosodic word, foot, or phonological phrase. Nguyen and Ingram (2006, 2007b) and Nguyen (2010), who measure the acoustic properties of the syllables in disyllabic and longer words, support the presence of final prominence in an iambic pattern, unlike Brunelle (2017). Note, however, that Nguyen (2010) restricts her polysyllabic words to (rather unnatural) nonce words. Notably, Pham (2008) shows evidence of prosodic words by the finding that function words tend to be cliticized to their host on their left.

The debate described above motivates further investigation into rhythmic patterning in Vietnamese. The existence of a strong and weak pattern between the syllable and the phonological phrase would provide evidence for a non-flat structure and thus justify the presence of an intermediate level. This paper focuses on examining the rhythmic pattern of polysyllabic words via their syllable duration. Methodologically, native speakers of the Southern dialect were asked to read a constructed list of 10 quadrisyllabic existing reduplicative words embedded into a carrier sentence *Nói ABCD ngay 'Say ABCD now."* These words are morphosyntactically either right-headed or left-headed. Their syllable duration was measured and statistically analyzed.

The results presented in this paper were collected from the 3 repetitions of 10 quadrisyllabic reduplicative words by 5 speakers. Figure 1 shows the mean duration of each syllable calculated for all speakers. First, it may be seen that the investigated reduplicative words form an iambic pattern in which the second and fourth syllables are longer than the first and third syllables. However, one-way ANOVA analyses do not show significant statistical support for this pattern. Second, there may be no effect of the morphosyntactic difference on the rhythmic pattern of these reduplicative words. The expectation of the longer duration for the first and second syllables in the left-headed words and for the third and fourth syllables in the right-headed words is not observed in the results. Instead, there is a similarity in these patterns in which the first and second syllables tend to be longer than the third and fourth syllables.

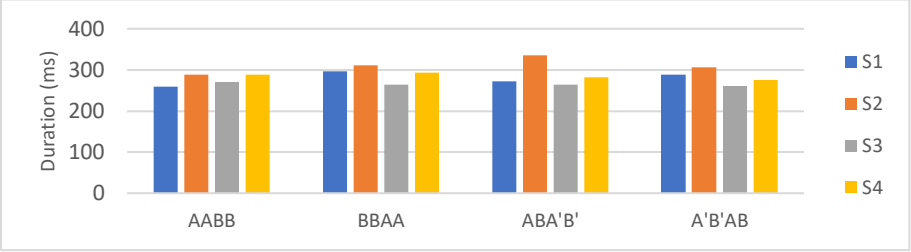


Figure 1. Mean syllable duration for all speakers[1]

[1] AABB is left-headed full reduplication, BBAA is right-headed full reduplication, ABA'B' is left-headed partial reduplication, and A'B'AB is right-headed partial reduplication

L1 Differences in Cross-language Perceptual Mapping and What this Says about Second Language Identification

Kenneth de Jong (*Indiana University*), Yu-Jung Lin (*College of the Holy Cross*), Yen-chen Hao (*University of Tennessee*), and Hanyong Park (*University of Wisconsin Milwaukee*)

Commonly cited models of bilingual phonology, e.g. the Speech Learning Model and the Perceptual Assimilation Model (L2), attempt to explicate the role of a first language phonology in second language phonological capabilities. Both models predict a L2 developmental pattern from relations between the L1 and L2. This paper presents two attempts at testing this general approach to understanding English-as-a-Foreign-Language learners' identification skills, one group with Korean as L1 and the other with Taiwan Mandarin as L1. The two language groups were chosen due to differences in their phonological inventories; Korean does not have non-sibilant anterior fricatives, while Mandarin has /f/. The groups also both have experience with written systems that transparently indicate consonant identities using non-Roman character sets, *Hangul,* the standard Korean orthography, and *Zhuyin Fuhao*, a written phonetic annotation system used in Taiwan.

Both groups performed two tasks with productions of a substantial portion of the English consonant system placed in various prosodic positions: onset, coda, intervocalic post-stress and intervocalic pre-stress positions. Both tasks asked listeners to identify each consonant from a large set of possible alternatives, the first task asking them to use orthographic labels from *Hangul* or *Zhuyin Fuhao*, and the second task asking them to use labels from the Roman alphabet with English key-words. From this larger set, we analyze voiced and voiceless, labial and coronal, non-sibilant fricatives and plosives (2 x 2 x 2 = 8) in the four prosodic positions (=32) as cases to examine the relationship between L1 mapping and L2 identification. Analytic processes, presented in Park & de Jong (2008, 2017), use L1 labeling responses to estimate the probability of mapping English productions onto L1 categories, and then also the reverse probability of mapping the L1 categories onto English categories. These probabilities are then combined to estimate the likelihood of English identification, if listeners rely only on the L1 categories. We expect predictions to be accurate for consonants that are 'similar' across the languages, but to underestimate accuracy for consonants that are 'novel' to the L1, due to the listeners' learning of the novel segment.

We present differences in the predicted identification patterns for the two languages, showing that having the additional /f/ category creates pervasive differences in the mapping pattern for many of the consonants, and that just having more matching categories in the L1 does not necessarily lead to predictions of better L2 identification. Comparing with actual identifications shows that the accuracy of the "novel" categories is generally much higher than what is predicted, as expected with Flege's model; however, this analysis shows that English /f/ is not a "novel" category since it maps well onto Mandarin /f/, but is predicted to exhibit the same patterns as found for other non-sibilants which are "novel" to the Mandarin listeners. We will consider various aspects of the model which might account for its discrepancy from actual observations.

Park, H, and K.J. de Jong (2008). "Perceptual Category Mapping between English and Korean Prevocalic Obstruents: Evidence from Mapping Effects in Second Language Identification Skills." *Journal of Phonetics,* **36**: 704 - 723.

Park, H, and K.J. de Jong (2017). "Prosodic Position Effects on Perceptual Category Mapping between English and Korean Obstruents." *Journal of Phonetics,* **62**: 12 - 33.

Productivity and Acoustic Realization of Mandarin Tone 3 Sandhi in L2 Learners

Xiao Dong (dong1@iu.edu); Chien-Jer Charles Lin (chiclin@indiana.edu)
Indiana University Bloomington

Introduction The acquisition of phonological rules in language learning has been widely recognized as a challenge. Learners not only need to internalize the underlying and surface forms but also develop the ability to apply these rules productively to unfamiliar words (e.g., Shannon & Rachel, 2020). Our study delves into understanding this linguistic complexity by investigating the second language (L2) acquisition of a productive phonological alternation rule in Mandarin Chinese, Mandarin Tone 3 sandhi, where the initial low tone (Tone 3 or T3) of a low tone sequence changes to a rising tone (Tone 2 or T2).

Research questions We explore whether L2 learners of Mandarin can acquire this rule and effectively apply it to nonce words (Yang, 2016; Qin, 2022). We also probe the roles that proficiency, word familiarity, and explicit instruction play in its productivity and realization.

Method A wug test was used for data collection, engaging 15 native speakers (Group 1 or G1), 16 intermediate-low level learners (G2), and 17 indeterminate-mid level learners (G3). The test materials are twelve T2T3 and T3T3 minimal pairs, with six pairs being real words and six pseudos. During the test, participants heard the two monosyllables of each disyllabic unit separately and produced them together as a disyllabic word. Two types of measurement were included: native speaker judgment (Sandhi vs. Non-sandhi) and acoustic features measurement. In addition, learners proficiency level and their familiarity with the test words were measured in the test.

Results & Discussion Regarding the application rate of the T3 sandhi rule to pseudowords, native speakers exhibited a 100% application rate, affirming the rule's productivity (Zhang & Lai, 2010; Zhang & Peng, 2013). Intermediate-low and intermediate-high levels demonstrated average rates of 80% and 83.3%, respectively, suggesting successful rule abstraction. Subsequent acoustic analysis reinforced these findings, revealing only subtle nuances in pitch contours. Word familiarity emerged as a significant factor influencing sandhi realization for both learner groups, but in a way contrary to our expectations. Sandhi tone 3 production mirrored T2 better in low familiarity words. This may be due to the influence of the frequency of the first T3 syllables occurring as underlying T3. Notably, our investigation did not unveil any significant effect of explicit instruction on the acoustic realization of sandhi for either learner group.

Key References

Zhang, J., & Lai, Y.-W. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. Phonology, 27(1), 153–201. http://www.jstor.org/stable/40783035

Qin, Z. (2022). The Second-Language Productivity of Two Mandarin Tone Sandhi Patterns. Speech Communication, 138, 98-109

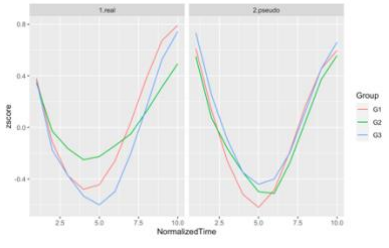


Figure 1: T3 sandhi realization for different groups across word conditions

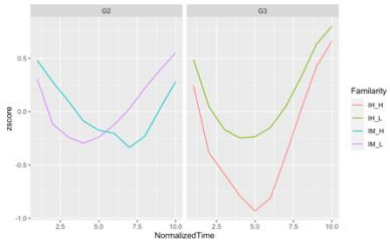


Figure 2: T3 sandhi realization by familiarity across learner groups

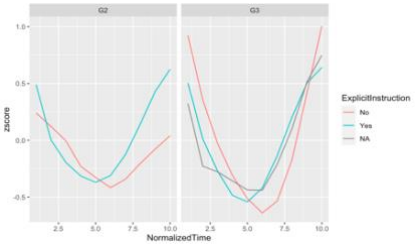


Figure 3: T3 sandhi realization by instruction type across leaner groups

Loss of Consonantal Quantity Distinctions from Middle High German to Early New High German

A Diachronic OT Analysis

Janine Emerson, Indiana University (Germanic Studies)
javonk@iu.edu

Modern Standard German does not have a quantity distinction between short and long consonants. This is especially striking, considering the fact that Old High German (OHG) shows more geminates than any of the other Germanic languages at the same time (cf. Braune & Heidermanns 2018: 132, § 91). How can such a process of drastic degemination be accounted for?

While New High German (NHG) only retains a contrast in vowel length, earlier stages, such as OHG and Middle High German (MHG) still distinguished between vowel and consonant length (see Page 2020: 98, 109). The aim of this paper is to capture this shift in quantity distinctions from MHG to NHG from the perspective of Optimality Theory (OT), as there is no OT account that traces this development throughout the history of German as a whole.

Focusing on intervocalic geminates, I will establish a ranking for each relevant stage: one for Early MHG, Late MHG, and Early New High German (ENHG). My choice of constraints is based on Kristofferson (2011) who explains changes from Old Norse (in which syllable weight and stress were independent) to current North Germanic dialects (in which stressed syllables are obligatorily bimoraic) with the help of a constraint for minimality requirements. Throughout the course of my analysis, I will show that the loss of consonant quantity distinctions in German, however, cannot be captured by solely relying on such a constraint. Rather, it is necessary to incorporate a constraint against geminates in word-final position for the stage of Early MHG, and one against geminates in all positions to describe the state of affairs in ENHG. Furthermore, I will argue that the loss of intervocalic geminates is connected to the emergence of ambisyllabic consonants in ENHG.

References

Braune, Wilhelm & Frank Heidermanns. 2018. *Althochdeutsche Grammatik I.* Berlin: De Gruyter.

Kristoffersen, G. 2011. Quantity in Old Norse and modern peninsular North Germanic. *Journal of Comparative Germanic Linguistics* 14, 47–80.

Page, B. Richard. 2020. Quantity in Germanic Languages. In Putnam, Michael T. & B. Richard Page (eds.), *The Cambridge Handbook of Germanic Linguistics*. 97–118, Cambridge: Cambridge University Press.

Evidence for Selective Prosody: The Case of Ambiguous Gapping in Hijazi Arabic

Aisha Fuddah, Hammad Alshammari, & Hanyong Park
Department of Linguistics, University of Wisconsin-Milwaukee
afuddah@uwm.edu, alsham58@uwm.edu, park27@uwm.edu

We investigated how Hijazi Arabic speakers dealt with syntactic ambiguity involving gapping in their production. Gapping is an elliptical construction in which the verb in a non-initial conjunct of coordinated structures is elided, sometimes leading to ambiguity. For example, "Bill took chips to the party and Susan to the game" can be interpreted as either "[Bill took chips to the party]$_{IP}$ and [Susan ~~took chips~~ to the game]$_{IP}$", which is a gapping interpretation, or "Bill took [chips to the party]$_{VP}$ and [Susan to the game]$_{VP}$", a non-gapping interpretation (Carlson, 2001). To the best of our knowledge, there has been no production study on how speakers resolve the structural ambiguity resulting from gapping. Thus, we conducted a production experiment with the aim of exploring whether speakers utilize prosodic cues in this situation.

We had 24 native speakers of Hijazi Arabic read a set of three types of sentences: unambiguous gapping (Type A), unambiguous non-gapping (Type B), and ambiguous potentially gapping (Type C). All sentences followed the construction pattern of [Argument1][Verb][Argument2][Adjunct][and][Arugment3][Adjunct], where gapping was realized within the constituent involving [Argument 3], if present. Due to the fact that none of the participants recognized the gapping interpretation in Type C sentences, we divided them into two groups. Half of the participants initially produced Type C sentences with a non-gapping interpretation, followed by the gapping interpretation after its introduction. The remaining participants followed the reverse order. Our analysis focused on pause locations, pause duration, stressed-syllable final lengthening, mean and max F0, and mean intensity.

The results revealed two key findings. First, Type A and Type B sentences exhibited no differences in the properties we examined. This suggests the absence of overt prosodic distinctions between gapping and non-gapping structures. Second, gapping and non-gapping Type C sentences displayed distinct patterns in terms of pause duration before "and". Notably, gapping reading sentences exhibited longer pause duration (M = 383ms) compared to non-gapping ones (M = 32ms). Further analysis demonstrated that final lengthening occurred more frequently in the stressed syllable before "and" in the gapping reading sentences (92%) than in non-gapping sentences (12%). The results of both pause and final lengthening indicated that prosodic boundaries aligned with syntactic boundaries in Type C sentences.

In summary, our study suggests that speakers employ prosodic cues to resolve syntactic ambiguity, although they might not utilize them extensively when utterances have only a single interpretation. As such, we argue that prosody operates as a filter to confirm or reject syntactic parsing when ambiguity is recognised (Snedeker & Trueswell, 2003).

References

Carlson, K. (2001). The effects of parallelism and prosody in the processing of gapping structures. *Language and Speech*, *44*(1), 1-26.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and language*, *48*(1), 103-130

Colombian Spanish Intonation:
A comparison between Colombian Caribbean Spanish intonation in cities and towns

Estefanía Galindo Navarro – Ph.D. Student | galindo3@wisc.edu

This exploratory study provides a comparative acoustic analysis of the phrase-final intonational patterns of polar questions (PQs) produced by speakers of Colombian Spanish. Previous research has pointed out variation in final rises or falls in question intonation across Spanish dialects, with speech style and a series of social factors, such as gender, age, level of education, and urban or rural living conditions, all playing an influential role. Within the last decade, scholars have carried out valuable research in different Colombian regions involving the intonation of vocatives (Braun et al., 2018), declaratives and interrogatives (Muñetón & Dorta, 2015), assertive statements with expressive force, and exclamatory sentences involving commands and requests (Velásquez-Upegui, 2015, 2016).

The Autosegmental Metrical (AM) model (Ladd, 2008; Pierrehumbert, 1980) and its Spanish-specific transcriptional offshoot, Spanish in the Tones and Break Indices (Sp_ToBI), have been the typical frameworks used to classify the pitch accents and boundary tones of various utterance types and pragmatic conditions in Spanish. Regarding Caribbean Spanish, an example of a nuclear (i.e., phrase-final) pitch accent and boundary tone that has been attested in Spanish PQs intonation is H+L*L% (i.e., falling movement) for Dominican Spanish, while it has also been mentioned that there is a certain intonational similarity in questions produced in Dominican, Puerto Rican, and Canarian Spanish since, in these three Spanish varieties, questions can be produced with a falling nuclear pitch accent (see, e.g., Armstrong, 2017; Willis, 2010; among others). Furthermore, within Colombia, Roberto (2023) studied the Spanish spoken in Bucaramanga, mentioning that speakers from this region produced both PQs and wh-echo questions with a rising boundary tone H%, while the boundary tone was falling in other types of utterances.

The novelty of the present study is that it is an initial attempt to draw a comparison between the intonational variation of PQs (specifically: offers, invitations, imperatives) in urban and rural settings by gathering data through a discourse completion task (see sample task items below) conducted in a city and five other towns located along the Caribbean coast of Colombia, a region whose intonational patterns remain understudied to date. Twenty-four speakers differing in age (young, mid, old), gender (male, female), level of education (low, mid, high), length of residence in their cities or towns, and level of bilingualism (monolingual, bilingual), were split into four different groups (i.e., rural, urban, experience in both settings, and bilinguals) and recorded. Our preliminary examination of the data (analysis in progress), which were acoustically inspected in Praat (Boersma & Weenink, 2022) and transcribed based on Sp_ToBI, shows that the most frequent phrase-final configuration across participants, regardless of the urban vs. rural contrast, is L*L% (i.e., final gradual fall to a relative low). However, participants who have studied at the university level tend to prefer the bitonal boundary tones HH% or LH% (both indicative of a final rise).

Overall, this study represents a preliminary glimpse at both Colombian Caribbean Spanish's intonation and the urban vs. rural comparison, thus advancing our understanding of both Spanish intonation as well as sociolinguistic approaches to it. In order to expand upon these preliminary data, further data are currently being collected through fieldwork in the region

in question, as well as in regions of comparison (Cali, Manizales, Bogotá, Medellín, and Cartagena), some of which will be integrated into this presentation.

Sample task items:

1. . *"You have many bananas in your grocery bag. While you are walking on the street, you see an old woman who seems to be hungry. You ask her if she accepts a banana from yours".*
   *- **¿Le regalo un guineo?** "Would you accept a banana?"*
2. *"Your son's friends came over to play dices with him today. When they leave, you ask whether they are coming over to play again tomorrow"*
   *- **¿Vienen a jugar dados mañana?** "Are you coming over to play dice tomorrow?*

References

Armstrong, M. (2017). Accounting for intonational form and function in Puerto Rican Spanish polar questions. Probus, 29(1), 1-40. https://doi.org/10.1515/probus-2014-0016

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.3.03, Online: http://www.praat.org/

Braun, B., Feldhausen, I. & Huttenlauch, C. (2018). The purpose shapes the vocative: Prosodic realisation of Colombian Spanish vocatives. *Journal of the International Phonetic Association, 48*/1 C doi:10.1017/S0025100317000597

Ladd, D. R. (1996/2008). Intonational phonology. Cambridge: Cambridge University Press.

Muñetón, A, M. & Dorta, L, J. (2015). Intonation of declarative and interrogative statements in Colombian Spanish of Medellin: female vs. male voices. *Boletín de Filología*, Tomo L Número 2: 103-122

Pierrehumbert, J. (1980). The phonetics and phonology of English intonation. Ph.D. diss., Cambridge, MA: MIT dissertation.

Roberto, Y. J. (2023). The intonation of Spanish spoken in Bucaramanga (Colombia). *Estudios de Fonética Experimental*, 32, 43–65.

Velásquez-Upegui. E. (2015) ¿Entonación de enunciados aseverativos con fuerza expresiva o de enunciados exclamativos? A propósito del español hablado en Colombia. *Universidad Autónoma de Querétaro Moenia 2*, 111-130. ISSN: 2340-003X.

Velásquez-Upegui, E. (2016). Intonation of commands and requests in four Colombian dialects. *Lingüística y Literatura, (69)*, 31-49. https://doi.org/10.17533/udea.lyl.n69a01

Vilaplana, E. & Prieto, P. (2008). La notación prosódica en español. Una revision del Sp_ToBI. *Estudios de Fonéticas Experimental 18.* 263–283.

Willis, E. W. (2010) Dominican Spanish Intonation. Transcription of intonation of the Spanish language, 123-153

The Systematic Variation of Gestural Coordination Due to Sonority

Yunting Gu (guyuntin@msu.edu), Karthik Durvasula (karthikd@msu.edu)
Michigan State University

Overview: Previously, Crouch (2022); Crouch et al. (2023) suggested that sonority sequencing affects gestural overlap of CC clusters in Georgian. Other researchers also observed that the lag between a consonant and a vowel – CV lag – was significantly shorter in syllables beginning with a nasal stop than in those with an oral stop (Shaw & Chen, 2019; Gao, 2008). We use these prior results to infer a more general relationship between sonority and segment sequencing beyond consonant sequences to perhaps all segment sequences within a syllable, including CV. In the current study, we found that the CV lag in English correlates positively with the sonority difference between a C and V.

Methods: The hypothesis was evaluated using measurements from 3346 utterances (57 different speakers, 34 stimuli) from the Wisconsin X-ray Microbeam Database (Westbury et al., 1990). Specifically, 17 nonce-words *uhCa* (e.g. *uhya* [jɑ], *uhma* [mɑ]), 15 *sVd* words (e.g. *seed* [si], *sad* [sæ]) together with *back* [bɑ] and *been* [bi] were analyzed. The kinematic data were annotated in Matlab using the default settings of the *lp_findgest* algorithm of the *mview* package (Tiede, 2005). To evaluate the hypothesis, descriptive plots were generated by the *tidyverse* package (Wickham et al., 2019), and mixed effects modeling was conducted by the *lme4* (Bates et al., 2014) and *lmerTest* packages (Kuznetsova et al., 2017) in R (R Core Team, 2017). To compute the CV lag, each timestamp of the consonant was subtracted from the corresponding timestamp of the vowel. The sonority difference was quantified by subtracting the C sonority from the V sonority using the sonority scale in Parker (2002, 2008, 2011).

Results: It was found that CV lag based on gestural onset statistically significantly increased with the sonority difference for all the stimuli in an appropriate mixed effects model. While an analysis with *all* the data has higher statistical power, it does assume that the sonority scale used is linear and not just relative (which is contrary to most phonologists' belief), and it collapses across different articulators or gestures. For these reasons, we looked at sets of stimuli that control for the place of articulation or gesture of the consonant, such as 1) lip aperture: [wɑ, mɑ, bɑ, pɑ]; 2) tongue tip: [lɑ, nɑ, zɑ, dɑ, sɑ, tɑ]; 3) bV: [bi, bɑ]. Crucially, the analyses showed statistically significant positive relationships.

Implications: The results suggest a sonority-based speech production model, which entails that gestures sequentially map from abstract sequential segmental representations within a syllable according to the sonority differences with the adjacent segment. This model can account for the results in the current study and previously observed correlations in gestural timing discussed above. The study also provides the potential basis of a unified explanation for several markedness constraints in human language, if we make a further assumption that larger lags are preferred. Since a sonority rise has a larger gestural lag than a sonority plateau and a sonority fall, the preference towards a larger gestural lag surfaces as the Sonority Sequencing Principle (SSP), which states that, cross-linguistically, in syllable onsets, a sonority rise is preferred over a plateau, which in turn is preferred over a sonority fall (Clements, 1990; Greenberg, 1965; Berent et al., 2007; Ren et al., 2010; Zhao & Berent, 2016). Similarly, the above view could also explain the Sonority Dispersion Principle (Clements, 1990; Parker, 2011) and the cross-linguistic tendency that CV syllables are much more common (more "unmarked") than VC syllables (Ohala, 1990; Tabain et al., 2004; Nam et al., 2009).

Lexical stress patterns of Chinese proper nouns spoken by American English speakers

Jianfeng Steven Guo (jg31@illinois.edu)
University of Illinois at Urbana-Champaign

This study attempted to explore the lexical stress patterns of Chinese proper nouns spoken by native American English speakers. It is part of a larger investigation to see how the pronunciation of these nouns by second language learners of Mandarin evolves as they acquire Mandarin tones and intonation.

We recorded four native American English Speakers (one male and three females) with no knowledge of Chinese. (We will eventually record six speakers in total). The stimuli consist of 140 Chinese names, of which half are place names and half proper names. Each name category has 35 disyllabic names and 35 trisyllabic names. These stimuli were presented on a PC screen in a random order using a counterbalanced block design (each name category a block) with E-Prime 3 [1]. The participant was asked to read them first with a question intonation and then with a declarative intonation (e.g., Beijing? Yes, Beijing.). For this study, only the declarative productions of trisyllabic proper names and disyllabic place names of one speaker were analyzed.

Acoustic measures, including duration, mean F0, and mean intensity, were extracted from the sonorant parts of each syllable in the names using a script [2] in Praat [3]. For the current analysis, only mean F_0 and mean intensity were used. (Duration was not used because the sonorant parts often have different numbers of phonemes in each syllable in the same name, making the comparison less justified.)

For the trisyllabic proper names, one-way ANOVA was conducted with mean F0 and mean intensity respectively as a function of syllable position. The results showed significant differences among the three syllable positions in both measures ($p < 0.001$). Regarding mean F0, it was found by the Tukey test that three syllable positions in the names were all significantly different ($p < 0.001$). The mean F0 of the first syllable is on average about 80.6 Hz higher than the last syllable and 39.3 Hz than the second syllable. The second syllable is on average 41.3 Hz higher than the last syllable. About the mean intensity, the results of a non-parametric test showed significant differences among the three syllable positions. The mean intensity of the first syllable is on average about 2.3 decibels higher than the second syllable ($p = 0.049$), but about 2.6 decibels lower than the last syllable ($p = 0.02$). The last syllable has a mean intensity of about 3.7 decibels higher than the second syllable ($p < 0.001$) on average.

For the disyllabic place names, the results of two-sample t-tests showed that compared with the second syllable, the first syllable is on average about 66.1 Hz significantly ($p < 0.001$) higher in mean F0 and about 4.2 decibels greater ($p < 0.001$) in mean intensity.

The evidence that the first syllable tends to have both higher mean F0 and intensity in the Chinese disyllabic place names implies that these disyllables prefer initial stress. Whereas in the Chinese trisyllabic proper names, the first syllable tends to have the highest F0, but not the highest mean intensity. Since past studies [4] found F_0 property to be the most reliable feature in predicting stress level while intensity is relatively unreliable, the results here indicated that the first syllable is most likely to carry the primary stress.

In summary, this study found that both Chinese disyllabic place names and trisyllabic proper names prefer initial stress. It concurred with past findings that in English, words beginning with strong syllables are predominant in its vocabulary [5, 6].

References

[1] Psychology Software Tools, Inc. *E-Prime 3.0*. 2016. URL: https://support.pstnet.com/.

[2] Yi Xu. "ProsodyPro—A tool for large-scale systematic prosody analysis". In: Laboratoire Parole et Langage, France. 2013.

[3] Paul Boersma. "Praat, a system for doing phonetics by computer". In: *Glot. Int.* 5.9 (2001), pp. 341–345.

[4] Matthew Gordon and Timo Roettger. In: *Linguistics Vanguard* 3.1 (2017), p. 20170007. DOI: doi:10.1515/lingvan-2017-0007. URL: https://doi.org/10.1515/lingvan-2017-0007.

[5] Anne Cutler and David M Carter. "The predominance of strong initial syllables in the English vocabulary". In: *Computer Speech & Language* 2.3-4 (1987), pp. 133–142.

[6] Rolf Carlson et al. "Phonetic and orthographic properties of the basic vocabulary of five European languages". In: *STL-QPSR 1* 1985 (1985), pp. 63–94.

Speech acoustics in preschool-aged children with and without speech sound disorders

Malachi Henry mahhenry@indiana.edu
Megan Diekhoff merediek@indiana.edu
Steven Lulich slulich@indiana.edu
Department of Speech, Language and Hearing Sciences, Indiana University

Abstract: Native speakers of English typically acquire most consonants by age 6 (Crowe & McLeod, 2020). Some children fail to follow typical developmental trajectories, and may be diagnosed with a speech sound disorder (SSD) which is a detriment to their perceived intelligibility. In comparison to their typically developing (TD) counterparts, children with SSD produce speech sounds with less accuracy (Preston et al., 2015). Although vowels account for approximately half of the acoustic stream of speech, the majority of the literature tends to focus on consonant productions. The knowledge base relating to vowels in disordered and developing speech comes from perceptual analyses such as phonetic transcription and measures of articulation (Howard & Heselwood, 2013; Stoel-Gammon & Pollock, 2008). However there is evidence that suggests perceptual methods such as subjective transcription of speech errors are less reliable and valid when applied to the study of vowels (Cox, 2008; Howard & Heselwood, 2013). Analyzing children's speech acoustically poses several challenges due to the complex and concurrent developmental processes that potentially impact acoustic measures. Previous studies of typical speech development in children have contributed to a growing base of normative data in speech acoustics. Reduced Vowel Space Area (VSA) has been observed in numerous communication impairments, including cerebral palsy and other childhood neurological disorders, dyslexia, residual speech sound disorders, and Down syndrome (Bertucci et al., 2003; Bunton & Leddy, 2011; Higgins & Hodge, 2002; Hustad et al., 2010; Liu et al., 2005; Narasimhan et al., 2016; Spencer et al., 2017). Koenig (2001) suggested that children with typically developing speech may produce voiceless consonants with longer Voice Onset Times (VOT) when compared to adults.The aim of the present study is to utilize acoustic measurements of VSA and Voice Onset Time VOT to quantify differences in acoustic characteristics among monolingual American English-speaking children with and without speech sound disorders. Participants included children who were reported to be monolingual native speakers of American English who were either diagnosed with speech sound disorders (N = 9) or were children with typically developing speech (N = 17) between 3;0 to 5;7 years of age. Single word recordings were gathered from the Speech Exemplar & Evaluation Database (SEED; Speights et al., 2020). Word-level tokens of phonetic contrasts were used for analyzing VOT using bilabial (“pig”), alveolar (“teeth”), and velar (“king”) voiceless stops in word-initial position. Measurements of time in milliseconds between the stop closure, release, and onset of voicing were calculated. VSA was mapped using three point vowels: [i], [u], and [ɑ]. VSA was calculated by identifying formant frequencies (F1 and F2) measured at the midpoint of each vowel segment, with three tokens for each vowel. Results revealed that children with SSD were observed to have smaller VSAs than typically developing counterparts, consistent with the first hypothesis. Additionally when mapped to the acoustic space, greater variability was observed in the children with SSDs. There was no significant difference in VOT between the two groups.

Awareness of Glottalization among American English Speakers

Sarah Holmstrom, University of Wisconsin- Madison, slholmstrom@wisc.edu

While much has been written about the stigmatization of glottal /t/ in British English (see Fabricius 2002, Foulkes et al. 2005, Docherty and Foulkes 2005), it has been claimed that glottalized /t/ in American English carries no such stigma (Eddington and Taylor 2009). Recent studies (Roberts 2006, Eddington and Taylor 2009, Eddington and Chandler 2010, Holmstrom 2021) have demonstrated that /t/-glottaling is increasing across various American English dialects, but evidence suggests that listeners show low awareness of this feature. The systematic pattern of adoption and low awareness of this ongoing sound change suggest that it fits Labov's criteria for a *change from below* (1966, 1994, 2021). However, no study has directly tested awareness of glottal forms among American English speakers. This project aims to assess how well listeners can distinguish between glottal and non-glottal forms of coda /t/.

An ABX discrimination task evaluated listeners' ability to differentiate words that are identical except for the variant of word-final /t/ they contain. Four target words, *heat, hat, hoot,* and *hot*, were used with different word-final /t/ endings: [t], [ʔ], a laryngealized vowel, and complete elision. There were a total of 24 target word comparisons with 31 filler words. Eight listeners participated in a pilot survey and further data is expected before the conference. Initial results show that listeners consistently had an easier time discriminating between filler words than target words. Several of the filler word comparisons only differed by one phoneme, such as *hat-had*, which indicates that listeners were less attuned to the phonetic variation in /t/ than they were to phonemic differences. This result supports previous findings that listeners are more easily able to discriminate between sounds that are separate phonemes in their native language (Werker 1995). Of the target words, the listeners had the easiest time distinguishing between words containing a phonemic difference: words ending in [t] compared to words with complete /t/-elision. However, they had the hardest time discriminating between words ending in [ʔ] vs. elision; these words also contain a phonemic difference (presence or absence of a final phoneme). This suggests difficulty hearing glottals at the end of a word. As Table 1 shows, pilot results point to a range of successful identification. Overall, these initial results indicate that while listeners may have some ability to hear glottal forms of /t/, they may be less aware of these phonetic differences than they are of more socially salient forms of variation. Further data is needed to confirm these initial results, but they suggest that word-final /t/ glottalization in American English is a sound change below the level of awareness, meeting Labov et al.'s (2011) definition of an *indicator*.

Table 1: Accuracy of listeners on ABX task for target and filler words

Comparison	Correct	Incorrect	Total
t vs glottal	19 (73.1%)	7 (26.9%)	26
t vs elision	20 (76.9%)	6 (23.1%)	26
t vs laryngeal	18 (64.3%)	10 (35.7%)	28
glottal vs laryngeal	24 (85.7%)	4 (14.3%)	28
glottal vs elision	15 (57.7%)	11 (42.3%)	26
laryngeal vs elision	16 (64%)	9 (36%)	25
fillers	185 (95.4%)	9 (4.6%)	194

Phonetics and phonology in stop lenition: Evidence from Romance languages

José Ignacio HUALDE
University of Illinois at Urbana-Champaign
jihualde@illinois.edu

An important question that often arises in the study of sound alternations is whether a given phenomenon should be characterized as a phonological rule or, instead, as a matter of phonetic implementation (or, indeed, whether we should make a distinction at all between phonetics and phonology). In particular, the evidence for pervasive variation and gradience in the lenition of onset obstruents in Spanish, Sardinian, Italian and other Romance languages that acoustic studies over the last few decades have revealed (Cole et al. 1991, Martínez-Celdrán & Regueira 2008, Hualde et al. 2011, Figueroa Candia 2016, etc) has often led to a reconsideration or rejection of earlier (generative) phonological analyses, in favor of a purely phonetic approach to lenition (most forcefully by Katz 2021, but see also Mascaró 1991, Broś et al. 2021, among others). In view of the gradience shown by the facts, representing a phenomenon such as Spanish spirantization or Campidanese voicing lenition as a simple alternation between two distinct allophones or as a change in the value of a binary feature would seem overly simplistic. In this presentation I will argue, however, that there are a number of good arguments for maintaining the traditional phonological analysis of these processes (e.g. /bdg/ --> [+cont] in context C), in spite of gradience in the output. Arguably, each of the consonants involved in the process has two distinct articulatory targets. I will argue that to fully understand phenomena such as Spanish spirantization and similar lenition processes the results of acoustic studies need to be interpreted alongside other evidence, including facts of surface opacity, dialectal variation and sound change. The view that I will defend is that the traditional phonological analysis with two allophones in complementary distribution is correct, but that each of the two allophonic targets admits a window of realizations and the actual realization depends in part on a number of contextual factors, including stress, and the specific preceding and following context. Phonetics and phonology are both involved. Time permitting I will also consider other phenomena, such as vowel reduction, where similar questions arise.

References

Broś, K., M. Żygis, A. Sikorski &, Wołłejko. 2021. Phonological contrasts and gradient effects in ongoing lenition in the Spanish of Gran Canaria. *Phonology* 38:1–40. **Cole**, J., J. I. Hualde &. K. Iskarous. 1999. Effects of prosodic and segmental context on /g/-lenition in Spanish. In O. Fujimura, B.D. Joseph & B. Palek (ed.) *Proceedings of Lp'98*, 575-589. Prague: The Karolinum Press. **Figueroa** Candia, M. A. 2016. Lenition in the production and perception of Chilean Spanish approximant consonants: Implications for lexical access models. Doctoral diss,, UCL. **Hualde**, J. I., M. Simonet & M. Nadeu. 2011. Consonant Lenition and Phonological Recategorization. *Laboratory Phonology* 2: 301–329. **Katz**, J.. 2021. Intervocalic lenition is not phonological: evidence from Campidanese Sardinian. *Phonology* 38: 651-692. **Martínez-Celdrán**, E. & Regueira, X, L.. 2008. Spirant approximants in Galician. Journal of the International Phonetic Associatiion 38: 51–68. Mascaró J. 1991. Iberian spirantization and continuant spreading. *Catalan Working Papers in Linguistics*, 1, 167–179.

Constructing a Phonetically Balanced Speech Corpus for Mandarin Chinese

Mingyue Huo (mhuo5@illinois.edu) and Yan Tang (yty@illinois.edu)
Department of Linguistics, University of Illinois at Urbana-Champaign

Phoneme distribution within a corpus is critical in speech-related research and applications. For instance, bias in speech perception studies may arise from the variability caused by the disproportionate occurrence of different phonemes within the stimulus set. Therefore, it is desirable to use a speech corpus that matches the real-world phoneme distribution as closely as possible. The *Harvard Sentences*, also known as the *IEEE Sentences* [1], has been widely used in speech and linguistic research. There are 72 lists, each of which contains 10 English sentences. These sentences were designed to have relatively low word-context predictability, and are constructed to ensure that each phoneme occurred a certain number of times in each list as well as in the corpus. More recently, a Spanish version of the Harvard corpus – the Sharvard corpus – was created [2] using a similar design. For Mandarin Chinese – the language spoken by the greatest number of native speakers, however, there is no public corpus comparable to *Harvard Sentences* to the best of our knowledge. Moreover, Mandarin's tonal nature introduces additional complexities when devising balanced phonemic representations.

In this study, we report the process of constructing a phonetically balanced speech corpus for Mandarin Chinese. First, 3000 sentences were randomly selected from the base corpus *AISHELL-3* [3] with the length of 8~11 Chinese characters, in accord with the syllable number of Harvard Sentences (mean=8.9, P25=8.0, P75=10.0). We calculated the perplexity of each sentence with language model GPT-2, where a higher perplexity indicated a less predictable word-context. A subsequent manual checking ensured that all the sentences are syntactically plausible, yielding 800 qualified sentences with the relatively high perplexity score. Then, all sentences were transcribed into phoneme and tone sequence using *phkit*, a python toolkit based on Tsinghua phoneme set containing a total of 218 toned phonemes. A comparison of phoneme distribution was conducted among the 800 sentences and two large-scale Mandarin corpora *Zaixian* [3] and *THCHS-30* [4], which were assumed to capture real-world distribution due to their large vocabularies. We manually made up vocabularies and substituted some words in sentences until the corpus covered all Mandarin toned phonemes and its phoneme distribution approximated the two reference corpora, as shown in Figure 1, with cosine similarity of 0.987 and 0.978 respectively. Finally, the corpus was recorded by two male and two female speakers. We hope this phonetically balanced low-predictable corpus could benefit linguistic research, especially the speech perception subfield.

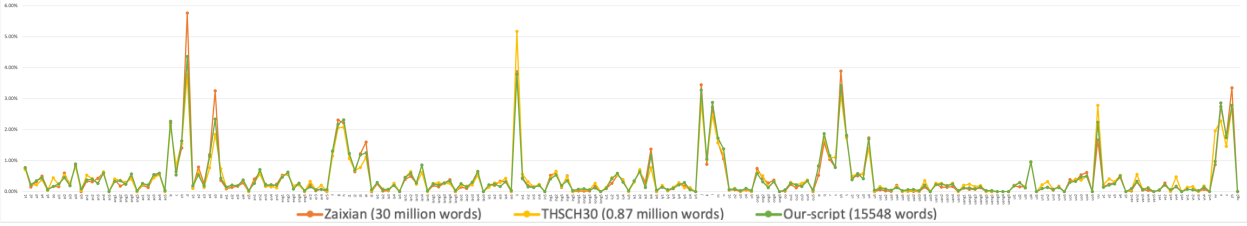


Figure 1: Toned phoneme distribution (in percentage) of 3 corpora: *Zaixian* with 30 million words, *THCHS-30* with 0.87 million words, and our corpus with 15548 words. High cosine similarity indicates close phoneme distribution.

References:

[1] Rothauser, E. H. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225-246.

[2] Aubanel, V., Garcia Lecumberri, M. L. and Cooke, M. (2014). 'The Sharvard Corpus: A phonemically-balanced Spanish sentence resource for audiology', International Journal of Audiology, 53(9) 633–638.

[3] Shi, Y., Bu, H., Xu, X., Zhang, S., & Li, M. (2020). Aishell-3: A multi-speaker mandarin tts corpus and the baselines. *arXiv preprint arXiv:2010.11567*.

[4] Website: http://www.cncorpus.org/

[5] Wang, D., & Zhang, X. (2015). Thchs-30: A free chinese speech corpus. *arXiv preprint arXiv:1512.01882*.

Rhotic Deletion and Demographics in the Salvador, Bahia, Variety of Brazilian Portuguese

Trey Jagiella

Indiana University Department of Linguistics

fjagiell@iu.edu

Brazilian Portuguese has two rhotic phonemes: the alveolar flap /ɾ/ and another variable phoneme historically identified as the long version of the rhotic and currently as a velar, uvular, or glottal fricative, or as an alveolar trill or approximant. The surface forms of this phoneme vary both within and across dialects. Deletion, especially word-finally, is also common. This word-final deletion has been attributed to African influences on Brazilian Portuguese in existing literature on the language. This project examines rhotic production in the city of Salvador, which is the largest city in Brazil with a population predominantly self-identified as being of African descent (Brazilian census categories *preto* "black" and *pardo* "brown" or "mixed race"). Given these factors, there is a question of how race may be tied to deletion of the rhotic in this variety.

For this project, thirty-five participants (self-identified as 21 *preto*, 6 *pardo*, 7 white, 1 indigenous) read aloud predetermined stimuli of isolated tokens and sentences (total rhotic phones/deletions = 4964). Among the demographic variables tested (age, gender, race, socioeconomic class), only race had a significant difference in deletion rate, with participants identifying as *preto* or *pardo* deleting the phoneme more frequently than those identifying as white. This distinction is further complicated by the fact that most Brazilians, regardless of identification, have a mix of African, Indigenous American, and European ancestry, and a lack of consistent classification of individuals into racial categories. While deletion is most common word-finally, it occurs in all environments where the phoneme is found. However, the difference in deletion rate by race is only significant in higher deletion environments (coda position). Analysis is currently ongoing as to demographic differences in surface forms of the rhotic when it is not deleted. Additionally, a secondary dataset from São Paulo to compare with the data from Salvador is being compiled.

The Northern Cities Shift isn't (yet?) receding in Northwest Indiana

Brian José -- Indiana State University
bjose.ling.phd@gmail.com

The Northern Cities Shift (NCS) is a clockwise-like rotation of several of the short vowels taking place throughout much of the Great Lakes region of the United States, where it has been described as the defining characteristic of the Inland Northern dialect of American English (Labov et al 2006: 121). However, a fair amount of recent research has found various elements of the NCS to be receding in a number of Inland Northern communities: most often and most consistently the TRAP /æ/ and/or the LOT /ɑ/ vowels (e.g., Dinkin 2011a & 2022 in Cooperstown NY, D'Onofrio & Benheim 2020 in Chicago IL, Driscoll & Lape 2015 in Syracuse NY, Nesbitt 2018 & 2021 and Wagner et al 2016 in Lansing MI, Thiel & Dinkin 2020 in Ogdensburg NY). Here, in the pursuit of two goals, I present an analysis of reading-passage data for 27 white and/or Latina speakers from the Calumet Region in extreme northwestern Indiana: 13 men and 14 women, recorded in 2007 and ranging in age from their 20s to their 70s.

The first goal is to provide a fuller account of the NCS in this community than has been documented in prior studies of it (Gordon 2000, José 2016, Labov et al 2006). The results confirm previous findings that the NCS is, indeed, attested in NW Indiana but that local participation is not robust. For instance, these 27 speakers' collective score on five NCS metrics (Labov 2007) is 2 / 5, with an average individual score of 1.6 / 5. Both of these values are comparable to the average score of 1.9 / 5 in NY's Hudson Valley, and they're intermediate between the average scores of 3.7 / 5 for all of the Inland North in the *Atlas of North American English* (*ANAE*) and 0.6 / 5 for *ANAE* communities outside of the Inland North (Dinkin 2011b: 77).

The second goal is to determine if there are any indications that the NCS is receding in NW Indiana, as it is in some other communities, as noted above. However, there's almost no such evidence. The only indication that the NCS might be receding here is a marginal effect of STRUT /ʌ/ shifting *up* and back in the vowel space, partially contrary to the *downward* and/or backward trajectory that it 'should' follow under the force of the NCS. At the same time, the relative proximity of STRUT /ʌ/ and LOT /ɑ/ on the F2 dimension (cf. Labov et al's 2006 *UD Index*) and the relative proximity of TRAP /æ/ and DRESS /ɛ/ on the F1 dimension (cf. Labov et al's 2006 *EQ Index*) are both indicative of continuing advancement of the NCS through apparent time, rather than any weakening of it. Furthermore, it's generally the case, with just a few exceptions, that even when a vowel exhibits effective (statistical) stability, it is nonetheless creeping along through apparent time in a direction consistent with the NCS; so, even where there isn't evidence of ongoing advancement of the NCS, there is often evidence against its retreat. Of course, how the NCS may have developed in NW Indiana since this data was recorded some 15 years ago remains to be determined; therefore, IRB approval for a new round of fieldwork that will seek to address that issue has just been secured.

References

Dinkin, A. 2011a. Weakening resistance: Progress toward the low back merger in New York State. *Language Variation and Change* 23: 315-345.

Dinkin, A. 2011b. Nasal short-a systems vs. the Northern Cities Shift. *U-Penn Working Papers in Linguistics* 17.2: 71-80.

Dinkin, A. 2022. Generational phases: Toward the low-back merger in Cooperstown, New York. *Journal of English Linguistics* 50: 219-246.

D'Onofrio, A.; Benheim, J. 2020. Contextualizing reversal: Local dynamics of the Northern Cities Shift in a Chicago community. *Journal of Sociolinguistics* 24: 469-491.

Driscoll, A.; Lape, E. 2015. Reversal of the Northern Cities Shift in Syracuse, New York. *U-Penn Working Papers in Linguistics* 21.2: 41-47.

Gordon, M. 2000. Phonological correlates of ethnic identity: Evidence of divergence? *American Speech* 75: 115-136.

José, B. 2016. Chicagoland and Northwest Indiana. *Proceedings of CLS* 50: 287-301.

Labov, W. 2007. Transmission and diffusion. *Language* 83: 344-387.

Labov, W.; Ash, S.; Boberg, C. 2006. *The Atlas of North American English*. Berlin. Mouton.

Nesbitt, M. 2018. Economic change and the decline of raised TRAP in Lansing, MI. *U-Penn Working Papers in Linguistics* 24.2: 67-76.

Nesbitt, M. 2021. The rise and fall of the Northern Cities Shift: Social and linguistic reorganization of TRAP in twentieth-century Lansing, Michigan. *American Speech* 96: 332-370.

Thiel, A.; Dinkin, A. 2020. Escaping the TRAP: Losing the Northern Cities Shift in Real Time. *Language Variation and Change* 32: 373-398.

Wagner, S.; Mason, A.; Nesbitt, M.; Pevan, E.; Savage, M. 2016. Reversal and re-organization of the Northern Cities Shift in Michigan. *U-Penn Working Papers in Linguistics* 22.2: 171-179.

Prosodic realization of coda voicing in ambiguous English sentences

Dong Jin Kim (kim252@uwm.edu) & Hanyong Park (park27@uwm.edu)
Department of Linguistics, University of Wisconsin-Milwaukee

The research aims to explore the extent to which English coda voicing is signaled by vowel duration in different prosodic locations involving prominence and boundary marking. While prominence marking enhances the phonemic distinction in a stressed syllable, boundary marking is associated with pre-boundary lengthening effect. Does the vowel duration difference indicating coda voicing become larger under prominence? Does the pre-boundary lengthening lead to a clearer vowel duration difference? What occurs when prominence and boundary effects conflict with each other? To address these questions, we examined vowel durations involving coda voicing in English sentences with relative clause attachment ambiguity.

The sentence “The next card shows the ‘bop’ on the school that is blue” can mean either that the word “bop” is blue or that the “school” is blue. This ambiguity arises because the relative clause “that is blue” can modify either the nearest noun phrase “the school” or the preceding noun phrase “the bop.” Native speakers resolve this ambiguity through the locations of prosodic boundaries and prominence [1]. When the “school” is blue, that word becomes more prominent, realized by pitch accents, and the prosodic boundary is positioned between “the bop” and “on.” If the “bop” is blue, this word becomes more prominent, and the prosodic juncture is placed between “school” and “that.” We had eight native speakers of English read sentences of this type, in which 12 minimal pairs of contrasting coda voicing appearing in two prosodic locations (e.g., “bop” and “school”). Following an illustrated description guiding them towards a specific interpretation, participants produced the sentences.

Preliminary results indicated that participants exhibited contrastive boundary marking depending on their intended meaning. Pauses were longer at the anticipated location corresponding to the intended meaning. Both prominence and boundary effects were observed. The duration difference was greater in prominent positions compared to non-prominent positions, and also at the domain-final compared to the middle of constituent. When prominence and boundary effects conflicted, such as a non-prominent, domain-final location vs. a prominent, non-domain-final location, boundary effects seemed to be more pronounced than prominence effects, though further analysis is necessary. We will provide detailed results if our submission is accepted to present at the conference.

[1] H. Baek, “A cross-linguistic comparison on the use of acoustic cues for ambiguity resolution,” *The Journal of the Acoustical Society of America*, vol. 145, no. 3, pp. 1926–1927, Mar. 2019, doi: 10.1121/1.5102002.

Faithfulness to foot structure in OT: a restrictive approach

Björn Köhnlein (OSU, koehnlein.3@osu.edu) & Oscar Coppola (OSU, coppola.31@buckeyemail.osu.edu)

Goal. Aiming to contribute to a long-standing debate in phonological theory, this presentation provides the tenets of a theory of faithfulness to metrical structure with underlying metrical trees. We propose to restrict faithfulness to maximal heads (= DTEs) and assume that the lowest-level head in the emerging hierarchy must be a vowel. As we show, our approach derives attested unpredictable patterns of surface metrification while avoiding unattested ones.

Background. It is often assumed that metrical structure cannot be underlying and/or contrastive – at least not above the level of the mora (Krämer 2012 for overview). Most and foremost, this position derives from the observation that syllabification appears to be predictable across languages. However, the existence of prosodic systems with unpredictable lexical or morphological stress requires some way of representing prominence underlyingly. Traditional approaches commonly solve this problem with either underlying diacritics or features that, in some versions, are then translated into surface foot structure (such as Alderete 1999, Revithiadou 1999 for diacritics, de Lacy 2020 for a feature [stress]).

Our approach. As discussed in, e.g., de Lacy (2020), protecting metrical boundaries by faithfulness can lead to a range of undesirable predictions, and therefore, the only tenable solution might be to exclusively protect maximal heads. Building on HEAD-MATCH constraints introduced in McCarthy (1995, 2000) for output-output correspondence and adopted in, e.g., Köhnlein (2011, 2016) and Morrison (2019) for input-output correspondence, we indeed argue that only maximal heads of underlying metrical trees (but not boundaries or dependents) are protected by faithfulness:

(1) HEAD-MATCH (DTE, where the DTE hierarchy = PW–FT–σ–μ–V): Assign a violation mark for every underlying maximal head that is not a maximal head on the surface

By virtue of (1), HEAD-MATCH constraints only protect candidates that conform to the specified hierarchy 'PW–Ft–σ–μ–V'; e.g., an underlying foot node is only protected by faithfulness if it is the head foot of a PW on the surface. Furthermore, underlying trees can be underspecified, functioning as, e.g., a floating foot without a prespecified location. Since the lowest element evaluated by HEAD-MATCH must be a vowel, a stored consonantal DTE (as in /kik^{DTE}a/; 'DTE' is shorthand for a tree) would not be protected by faithfulness, which eliminates unattested surface syllabifications of the type [ki.k.a]. As we demonstrate, our approach correctly derives phenomena such as unpredictable lexical stress and foot-based accentual oppositions but rules out unpredictable secondary stresses (not a maximal head), moraic stress (second mora of a syllable cannot be a DTE), and *unpredictable* stress on syllabic consonants (consonants cannot be DTEs, though *predictable* stress can still be on syllabic consonants; see Bell 1978).

Alternatives. Diacritic or featural solutions to the underlying representation of prominence that use foot structure on the surface sometimes rely on assumptions that are not that different from our proposal. For instance, Alderete and Revithiadou restrict associated lexical accents to vowels, and de Lacy (2020: 2) notes that "[i]t is possible that [stress] depends on the presence of [+vocalic]," comparable to our 'DTE → V' claim. Conversely, unless stipulated otherwise, the possibility of underlying metrical structure follows from *Richness of the Base* (e.g., Prince & Smolensky 1993) and is furthermore compatible with the notion of *homogeneity of inputs and outputs* (Moreton 2004 for discussion). Along those lines, our approach could be argued to be more parsimonious than diacritic/featural storage solutions (at least within OT) – *unless* it could indeed be shown that we can abandon metrical constituency altogether and specify diacritics/features both underlyingly *and* on the surface (as proposed in, e.g., Scheer & Szigetvári 2005, van der Hulst 2012).

Breaking away: "De-unifying" diphthongisation in Laurentian French

Jeffrey Lamontagne (Indiana University Bloomington; jlamonta@iu.edu)

Introduction. Laurentian (Quebec/Canadian) French exhibits variable diphthongisation, a stigmatised process that occurs at markedly reduced rates in formal speech (see e.g. Côté 2012; Bigot 2021). Vowels eligible for diphthongisation share a key phonological description: they are phonologically long on the surface, regardless of whether they are phonemically heavy (e.g. nasal vowels, /ɛː/) or not (e.g. vowels lengthened by /v z ʒ ʁ/). This generalisation would appear to motivate treating diphthongisation as a unified process even though that process is not available in the same segmental or syllabic contexts for all vowels. However, leveraging sociophonetic methodology to probe both linguistic and social conditioning, we argue that treating diphthongisation as multiple processes with superficial similarities offers a better understanding of both diphthongisation and the broader phonological system.

Methodology. We draw on 80 714 final-syllable vowels from spontaneous interviews conducted in Laurentian French on two television shows that differ in style while both displaying formal registers of the dialect (see Villeneuve 2017), computing Cartesian distances as a measure of degree of diphthongisation. Values were then submitted to mixed-effects linear regression to test differennces associated with phonemes, the syllable shapes (open, closed by a lengthening consonant, otherwise closed) and sociolinguistic context (more or less formal interview style).

Results. Nasal vowels are the only vowels to exhibit clear diphthongisation in final open syllables, consistent with phonologically length being variably retained for these vowels in this context (see Lamontagne 2022), and no difference is observed between the two shows. However, the degree of diphthongisation in these contexts in generally smaller than what is found for long vowels in closed final syllables and especially for vowels that underwent phonological lengthening (other than unbacked /a/, which is not a possible target). Surprisingly, nasal vowels in closed syllables exhibit more diphthongisation on the more formal show, reversing the pattern generally observed for oral vowels (where significant differences are found), though both /ɑ̃/ and /ɑ/ exceptionally exhibit this reversed pattern in lengthening contexts and the nasal pattern in other closed syllables. Phonological lengthening appears to predictably merge /ɛ/ into /ɛː/, both in terms of degree of diphthongisation and in terms of sociolinguistic conditioning. Finally, the interviewer of both shows follows the expected pattern whereby diphthongisation is less robust in the more formal show, whereas guests typically did the opposite where differences are found.

Discussion. While one might expect diphthongisation to pattern similarly across vowels and contexts, our results are for a distinction between several broad categories: (a) open-syllable nasal-vowel diphthongisation, (b) closed-syllable nasal-vowel diphthongisation, (c) closed-syllable oral-vowel diphthongisation, (d) lengthened-vowel diphthongisation, and finally (e) low-vowel diphthongisation (regardless of nasalisation). The latter low-vowel case, the otherwise consistent nasal-oral divide and the merger of lengthened /ɛ/ into phonemic /ɛː/ suggests that our results illustrate the main groupings operating in the dialect also reflected other phonological behaviours (see Lamontagne 2021). For example, /ɑ̃/ uniquely exhibits fronting instead of diphthongisation in final open syllables (in parallel with /a/ backing), and this separation from other nasal vowels is replicated in /ɑ̃/'s social conditioning mirroring /ɑ/'s. Other nasal vowels, on the other hand, relatively clearly pattern distinctly from heavy oral-vowel counterparts. Put together, this study suggests a reinterpretation of the organisation of the phonemic oppositions in Laurentian French and demonstrates the utility of system-wide comparisons for probing abstract phonological representations.

We're worrying about the wrong thing; the Matched Guise Technique and social speech perception.

Kyler Laycock, The Ohio State University & Kevin B. McGowan, the University of Kentucky

laycock.21@osu.edu kbmcgowan@uky.edu

Sociophonetic perception (Sumner et al. 2014, Drager 2010) is often studied using versions of the matched guise technique (MGT). Lambert et al. (1960), for example, found that bilingual Montrealer's voices evoked quite different social judgments in French vs English guises; providing evidence that listeners are able to perceive and connect social information in the voice to ideological framing of social types. In social speech perception research, cross-modal audio/visual matched guise studies are common in which visual information serves as a 'guise' for identical voice recordings; researchers sometime disregard the social information in voices (Rubin 1992) and sometimes take the combination of voice and visual stimuli into account (Gnevsheva 2017, Campbell-Kibler 2016, McGowan 2015). But uniting these researchers is the methodological assumption that the connection of voice to social type happens, indeed *must* happen, below the level of conscious awareness. Researchers attempt to deceive participants about the intentional use of guise to elicit evidence of social evaluation in language attitudes, segmental perception, memory, etc. It is generally assumed that the matched guise technique works at all because listeners are unaware of the guise manipulation.

This paper reports an audiovisual matched guise experiment with both standard 'hidden' and a novel 'unhidden' guise conditions. The basic task is a replication of Strand & Johnson (1996) in which listeners are asked to identify an ambiguous word as 'sack' or 'shack' on a continuum with the initial fricative ranging from purely [s] to purely [ʃ] under manipulated beliefs about the gender identity of the speaker (Tripp & Munson, 2022). Numerous previous replications have found that listeners perceive more of the ambiguous continuum as [ʃ] when they believe the speaker identifies as a woman and more as [s] when they believe the speaker identifies as a man and that, furthermore, this effect is bi-directional, with fricative type influencing perception of gender for an ambiguous voice (Bouavichith et al., 2019). Participants in the 'unhidden condition were briefed about the guise manipulation. They were instructed that the man or woman in the photo did not represent the voice they were listening to. When face and voice were not incongruous, participants in the hidden and unhidden guises exhibited the Strand & Johnson effect to nearly numerically identical extents. This result suggests that participants need not believe a link exists between a voice and a purported social category for visually-cued social information to influence segmental perception. However, when gender of the voice and face were incongruous, the expected effect was reversed; listeners' perception of the [ʃ]-[s] continuum tracked phonetically, rather than visually, cued social information. We explore the implications of this result for the MGT and for theories of social awareness and speech perception more broadly.

Investigating the Perceptual Center in Cantonese

Yu-Jung Lin[a], Bihua Chen[b], Kenneth de Jong[b]

a. Department of World Languages, Literatures, and Cultures, College of the Holy Cross (Email: yjunglin@gmail.com)

b. Department of Linguistics, Indiana University, Bloomington (Emails: bc1@iu.edu; kdejong@indiana.edu)

In the field of speech rhythm analysis, the perceptual center (P-center) of syllables acts as a foundational metric for determining syllabic duration. P-center research to date has predominantly centered on Indo-European languages, known for their proclivity to have consonant clusters. Languages such as Cantonese, where the syllable-initial position encompasses no more than a singular consonant, present less variability in duration. A pivotal study by Chow et al. (2015) indicated that in Cantonese, contrary to prior assumptions, the metronome beat is aligned with the syllable-initial consonant onset, rather than the vowel onset as is typical of previous studies. This behavior resonates with findings in sensorimotor synchronization paradigms like finger tapping, highlighting the potential use of the syllable-initial consonant onset as a key articulatory reference in speech synchronization.

Lin & de Jong (2023), however, finds Mandarin speakers to behave like speakers in previous examinations of European languages, despite the typological similarity with Cantonese. The current paper attempts to replicate and extend the Chow et al. (2015) study to include additional phonological elements, notably onglides and offglides. The current study more carefully matches the protocol used our Mandarin study, our study aims to determine if a similar pattern emerges in Cantonese. Also, Lin & de Jong (2023) finds that Mandarin speakers group pre-vocalic approximants with the following vowel and align the onset of the approximant to the metronome. The current study seeks to determine if a similar pattern holds with Cantonese speakers as well.

With 10 native Cantonese speakers as participants, we conducted a syllable repetition task across two distinct speech rates: 60 bpm and 120 bpm. Preliminary data analysis is in progress, with results anticipated to be ready by the time of the conference. Findings will be compared with those from our Mandarin and English projects, aiming to shed light on the intricacies of P-center across different languages. This study will not only enhance our understanding of Cantonese syllabic structure and rhythm but also to refine our theories of P-center in world languages.

References:

Chow, I., Belyk, M., Tran, V., & Brown, S. (2015). Syllable synchronization and the P-center in Cantonese. *Journal of Phonetics, 49,* 55–66. https://doi.org/10.1016/j.wocn.2014.10.006

Lin, Yu-Jung, and K.J. de Jong (2023). The Perceptual Center in Mandarin Chinese Syllables. *Journal of Phonetics,* **99:** 101245. https://doi.org/10.1016/j.wocn.2023.101245

Title: Arabic Heritage Speakers' Perception of Arabic Emphatic-Plain Contrasts and the Influence of Vowel Context and Emphatic Position in a Word

Authors: Maaly Al Omary, Jae Yung Song
Affiliation: University of Wisconsin-Milwaukee
Emails: malomary@uwm.edu, songjy@uwm.edu

Abstract:

Studies examining emphasis perception among native English and native Arabic speakers have shown that factors such as vowel context play a significant role in the perception of emphasis, often more so than the emphatic consonant itself. However, the impact of such factors has not been explored among Arabic heritage speakers. This study aims to investigate three key factors: (1) whether Arabic heritage speakers, compared to second language (L2) learners, benefit from their vernacular Arabic input in identifying emphatics, (2) whether vowel context, and (3) emphatic position (initial vs. final) serve as perceptual cues in perceiving emphatic consonants. A word identification task was conducted, analyzing accuracy scores for emphatic phonemes among two groups: 6 Arabic heritage speakers and 4 English-speaking learners of Arabic. The findings revealed that Arabic heritage speakers exhibited higher accuracy in perceiving emphatic consonants compared to L2 learners. Arabic heritage speakers demonstrated a stronger emphasis response in the context of /u/ as compared to /æ/ and /i/. In contrast, L2 learners of Arabic recorded the highest emphasis responses in the vowel context /æ/, followed by /u/ and /i/. Moreover, Arabic heritage speakers relied on the emphatic position in the final position of a word to identify emphatics, while L2 learners relied more on the emphatic initial position. These results suggest that Arabic heritage speakers benefited from their vernacular input in identifying emphatics, and that vowel context and emphatic position served as perceptual cues, albeit with significant differences between the two groups.

Looking at the *Loi de Position* through *Verlan* in Parisian French

Corentin Mazet (Indiana University)
cmazet@iu.edu

Mid-vowel alternation of the pairs /e/-/ɛ/, /ø/-/œ/, and /o/-/ɔ/ is one of the central areas of research in French phonology. This phenomenon, described by Tranel, among others, (Tranel 1987), was noticed as early as the 16th century (Straka 1981), and started being formally studied in the beginning of the 20th century (Fouché 1935, for instance). In the following centuries, the topic has been revisited multiple times and analyzed through the lens of several theories of phonology. However, no consensus on the process has ever been reached. The most widespread analysis is that of the *Loi de Position* (*LdP*) - a position law - which proposes a selection of the phoneme within the pair based on its position in the syllable. Several instances of this theory have been fleshed out; they can be summarized as such: closed syllables favor the lax version of the mid-vowel pair and open syllables favor the tense version. This analysis, although shared by many under one form or another, has never been confirmed by sufficient data. Several obstacles prevent it, such as the plurality of varieties of French and the large amount of variation within varieties. The limits of this analysis have given way to a number of other theories, all of which have pitfalls. Among them are analyses centered around etymological length (Morin, 1986, 1988), around synchronic length (Montreuil, 1995), around weight (Côté, 2008), and analyses focusing on phonetics aspect (Storme, 2017).

In this work, I am attempting to analyze mid-vowels through *Verlan*, a well-known language game of French, that has made its way into daily colloquial French and constitutes an ideal filter through which we can look at mid-vowels. First, this language game changes the syllabic structure of words, which allows us to test several analyses offered in the literature. In addition, *Verlan* words are associated with an informal style, which could be crucial in understanding the variation between speakers. Focusing on Parisian French, I analyze attested *Verlan* words containing at least one mid-vowel through the various lenses put forward in previous works.

The results of this study seem to support a syllabic-structure analysis: more than 95% of the *Verlan* words analyzed in Parisian French follow the strictest version of the *LdP* (closed syllable = lax mid-vowel, open syllable = tense mid-vowel). This is the case for all three mid-vowel pairs, in all types of syllable structures, and with a variety of words. Some examples are even more compelling; for instance, some words allow two *Verlan* counterparts, with two syllable structures, showing a perfect alternation in mid-vowels (/site/ <cité> 'hood' -> [te.ci] <téci> or [tɛs] <tess>). Some others can be verlanized twice with vowel insertions and constitute more data in favor of the *LdP* (/aʁab/ <arabe> 'Arab' -> [bœʁ] <beur> -> [ʁø.bø] <rebeu>).

Empirical studies remain to be conducted to confirm the data put forward in this work. However, if the data were to be confirmed, several conclusions could be drawn. First, it seems as though the *LdP* analysis holds true and is productive in Parisian French. Furthermore, other analyses, such as the etymological one, cannot account for the data put forward here. Finally, the fact that *Verlan* words seem ideal for the *LdP* could lead to interesting analyses on the representation of such processes.

References

Côté, M.-H. (2008). Loi de position: Où est la syllable? *Journées « Phonologie Du Français Contemporain »*.

Fouché, P. (1935). La prononciation actuelle du français. In A. Dauzat, *Où en sont les études de français* (pp. 13–34). Bibliothèque du français moderne.

Montreuil, J.-P. (1995). Weight and length in Conservative Regional French. *Lingua*, *95*, 77–96.

Morin, Y.-C. (1986). La loi de position ou de l'explication en phonologie historique. *Revue Québécoise de Linguistique*, *15*(2), 199–231. https://doi.org/10.7202/602567ar

Morin, Y.-C. (1988). Loi de position? *Revue Québécoise de Linguistique*, *17*(1), 237–242. https://doi.org/10.7202/602622ar

Storme, B. (2017). The *loi de position* and the acoustics of French mid vowels. *Glossa: A Journal of General Linguistics*, *2*(1). https://doi.org/10.5334/gjgl.300

Straka, G. (1981). Sur la formation de la prononciation française d'aujourd'hui. *Travaux de Linguistique et de Littérature*, *19*, 161–248.

Tranel, B. (1987). *The Sounds of French: An Introduction* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO978051162064

Pursuing a Rule-Based Approach to English Intonation

Joseph Moran
University of Wisconsin-Madison
jmoran8@wisc.edu

This project investigates mental representations of English intonation patterns and the process by which these patterns get applied to a given utterance. English is one of many languages that make use of distinct intonational patterns (i.e., tunes) to affect the meaning of utterances. For example, changing the intonation of the final word in the utterance below distinguishes between a declarative (falling tone) and interrogative (rising tone):

(1) She lives in Kentucky.

Much work has been done in the phonological literature to describe and account for the variety of tunes which may be associated with a given utterance, most notably the tones and break-indices (ToBI) notation system (Pierrehumbert, 1980). ToBI provides a robust system for transcribing intonation patterns according to an autosegmental-metrical (AM) framework. Furthermore, results from perception and production tasks suggest that these tones are mentally organized into a hierarchy of distinctiveness based on a primary distinction between high-rising and non-high-rising tunes (Cole et al., 2023). Despite the utility of the ToBI system, questions remain about how the mental representations of these tunes become realized in the externalization of language. Much of the contemporary work on English intonation is situated within a constraint-based phonological (CBP) account where the syntax-phonology interface is governed by optimality-theoretic (OT) constraints that both favor syntactic-phonological constituent isomorphism and rely on reference to prosodic units (i.e., syllables) which do not fall cleanly within the domain of either linguistic module. This project pursues a rule-based phonological (RBP) account of English intonation where the application of tunes to an utterance is analyzed from the perspective of the prosody-semantics interface with a strong commitment to modular specificity (Dalrymple & Mycock, 2011). This analysis of tunes further develops the precedence-based phonology (PBP) account of phonology (see Idsardi & Raimy, 2022; Papillon, 2020) by describing how tonal events described in the ToBI system can be represented and incorporated into phonological graphs. Under this analysis, tunes are analyzed as distinct 'blueprint' tunes that get inserted into the broader derivation in an instance of linearization from the semantics/pragmatics module to the phonology module. In doing so, this project pursues an analysis of tune-utterance association analogous to the ALIGN/WRAP constraints of Match Theory (Selkirk & Kratzer, 2020).

Title: “Did it hurt when you fell from Heaven?”: Variation in intonation and prosody in flirting

Author: Catherine Mott, **Affiliation:** University of Kentucky, **Email:** cimo225@uky.edu

The sociophonetic information we receive influences us. Is someone angry? Are they being sarcastic? Are they simply stating a fact? The speaker’s intonational patterns help identify what is going on within a situation (Armstrong et al., 2022). Though the phonetics of flirtation received some attention (Ranganath et al., 2009; Jurafsky et al., 2009), a clearly-established definition of what flirtation sounds like does not exist. In this project, I aim to answer these questions: what is the intonation pattern of flirtation in American English? What do these intonational patterns look like?

The data for this project was collected from students at the University of Kentucky through a Qualtrics survey with audio recording extension Addpipe. Of the total 57 eligible responses recorded, the age range was between 18-24, whose self-identified gender was split between male and female solely. The participants read aloud a set of 15 sentences with similar sentence structure pulled from HINT (Nilsson et al.,1994) in three guises: as a declarative sentence (as a control), as if they were angry (Combs, 2020), and as if they were being flirtatious. After each recording is force aligned using Montreal Forced Aligner (MFA), I will use pitch contours in Praat (with support from ToBI) to determine the intonational pattern for each sentence, and compare the individual recordings of each sentence to that sentence’s group to see the possible intonational patterns.

Understanding phonetic and phonological cues of flirtation tells us about the ability to encode social behavior in language. Flirting is an inherently social act, and its presence may color interactions differently, leading to a deeply interdisciplinary avenue that connects linguistics to the fields of anthropology, psychology and sociology, among others. This work will also provide a direction to continue looking at this pattern presented with perception within sociophonetics as one of the deeper pieces of the puzzle is: if there is an intonational pattern, how is it received and perceived?

References

Armstrong, M., Breen, M., Gooden, S., Levon, E., & Yu, K. M. (2022). Sociolectal and Dialectal Variation in Prosody. Language and Speech, 65(4), 783–790. https://doi.org/10.1177/00238309221122105

Combs, A. (2020). Relevant Angry Affect Slows Response Time To Commands.

Jurafsky, D., Ranganath, R., & McFarland, D. (2009). Extracting social meaning: Identifying interactional style in spoken conversation. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 638-646).

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085-1099.

Ranganath, R., Jurafsky, D., & McFarland, D. (2009). It’s not you, it’s me: Detecting flirting and its misperception in speed-dates. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 334-342).

Phonetic Evidence for Pitch Contour Contrast in Low German

Tianyi Ni (ni.386@osu.edu)
Department of Linguistics, The Ohio State University

Low German (LG), a Continental West Germanic language, is one of the few languages in the world that is described as having a ternary contrast between short, long, and overlong vowels in stressed syllables, as is illustrated in (1). A phonetic difference in terms of vowel quality is also declared by previous research (e.g., Kohler & Tödter 1984, Kohler 2001). Short vowels are always lax, while long and overlong vowels are usually tense. In other words, overlong vowels exhibit no differing quality with regard to the corresponding long vowels (von Essen 1958).

(1) a. [zɪt] 'sit-1.sg.pres.' [ziːt] 'side' [ziːːd̥] 'silk'
b. [rʏs] 'rust' [riːs] 'rice' [riːːz̥] 'giant'

It has sometimes been proposed that diverse tonal contours accompany the distinction between long and overlong vowels (e.g., Prehn 2012 for summary), but no modern phonetic studies have been able to confirm the existence of systematic tonal contrasts; scholars such as Kohler (2002) even doubt the existence of phonologically relevant pitch contrasts altogether.

This paper discusses the issue in East Frisian dialects of LG. On the basis of phonetic pilot measurements (from 10 LG speakers, and 5 are male), we provide phonetic evidence for the existence of a tonal contrast for East Frisian dialects (for instance, in non-final position of declarative sentences). We further argue that both pitch and duration are potential production cues for speakers to long and overlong vowels; speakers might make use of one of these cues or use both at the same time. For example, vowel duration is not a salient cue for speaker 6. A nonparametric test is applied upon all the non-final focused tokens in declarative carrying sentences (10 tokens for one single speaker with 5 long and 5 overlong vowels), whose result demonstrates that the vowel durations between long and overlong vowels are not significantly different (p-value=0.7104), while the pitch contour contrast exists, as shown in Figure 1. On the other hand, speaker 7, shown in Figure 2, uses vowel duration as a stronger cue (p-value=0.03788*), while a tonal contrast is not found. This may suggest that at least for some speakers, the system might transition from a three-way durational contrast to a pitch-based opposition, similar to claims made for Estonian (Lehiste 2003) and Franconian (Köhnlein 2015).

Figure 1. Lexical pitch contour visualization of the disyllabic minimal pair [riːtn̩] 'to tear' vs. [riːːd̥n̩] 'to ride' (left); vowel duration long vs. overlong (right); extracted from speaker 6

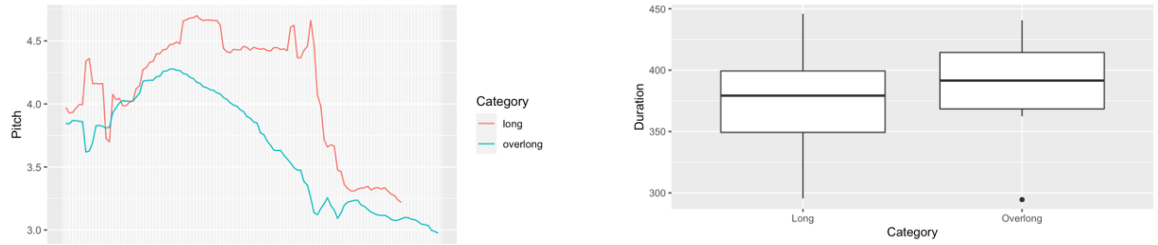
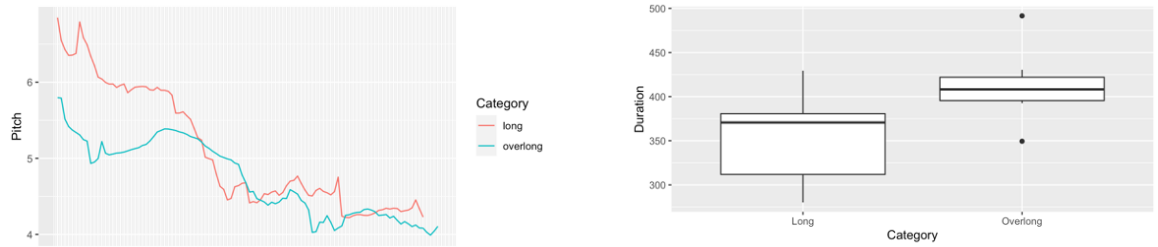


Figure 2. Lexical pitch contour visualization of the disyllabic minimal pair [riːtn̩] 'to tear' vs. [riːːd̥n̩] 'to ride' (left); vowel duration long vs. overlong (right); extracted from speaker 7



Solving complex problems: Analyzing palatal lateral change in Occitan

Kaitlyn Owens (Indiana University – Bloomington; kaitowen@iu.edu)

OVERVIEW: Whereas palatal consonants have been argued to be a single subsegment (e.g. Lipski, 1989; Backley, 2011), this study proposes they may behave as two subsegments in some languages. I analyze /ʎ/ because it is difficult to articulate exclusively in the palatal zone (Recasens, 2013), making it susceptible to variation. /ʎ/ is also often realized with a following [j] (e.g. [ʎj]; Zampaulo, 2015) which may be reinterpreted as a segment, given that perception may lead to sound change (e.g. Ohala, 2003). While many studies on /ʎ/-loss have focused on Spanish *yeismo* (e.g. Colantoni, 2004), this study focuses on two Occitan varieties: Gascon and Languedoc. Whereas [j] is replacing /ʎ/ in Gascon due to contact with French (Mooney & Hawkey, 2018), /ʎ/ depalatalizes word-finally (Oliviéri & Sauzet, 2016) in Languedoc. I probe whether diverging changes in Occitan may indicate different subsegmental representations of /ʎ/.

METHODOLOGY: I analyze 181 tokens of /ʎ/ (Table 1) in oral narratives performed by Gascon (1M, 1F) and Languedoc (1M, 1F) speakers in the *OcOr Corpus* (Vergez-Couret & Carruthers, 2018). I marked each token impressionistically in *Praat* as either [j], a lateral [ʎ l], or both [ʎj lj]. Laterals with low F1 (below ~450 Hz) are categorized as [ʎ] and those with high F1 (above ~550 Hz) as [l] because [ʎ] phones notably have low F1 (e.g. Tabain et al, 2014). /ʎ/ is never adjacent to a non-final consonant, and no word-initial tokens were followed by a non-palatal vowel.

	Word-initial	**Intervocalic**				**Word-final**			
	#_I	**I_I**	**I_V**	**V_I**	**V_V**	**I_#**	**V_#**	**I_s#**	**V_s#**
Gascon	[ʎj] 7 [lj] 1	[ʎj] 2 [j] 2	[j] 7 [ʎj] 5 [lj] 1	[ʎj] 27 [lj] 4 [ʎ] 1 [j] 1	[ʎj] 16 [lj] 3	[j] 2	[ʎj] 2 [j] 2	[ʎj] 1 [j] 1	[lj] 1
Languedoc	None	[ʎj] 3	[ʎj] 40 [ʎ] 1 [l] 1 [j] 1	[lj] 1	[ʎj] 8 [l] 3 [lj] 1 [j] 1	[ʎ]9 [l]9 [j]1 [ʎj] 3[lj] 1	[l] 5 [j]1 [lj] 1	[ʎ] 4 [l] 1	None

Table 1 Token distribution by variety and phonological context. **I** denotes /i e/; **V** denotes all other vowels

RESULTS: [j] is more common in Gascon than Languedoc. Further, [j] is more frequent in Gascon if the preceding vowel is palatal, as in Spanish *yeismo* (e.g. Lipski, 1989). Although results show word-final depalatalization in Languedoc, half of the lateral tokens in the data are palatal. Additionally, the Languedoc speakers in our data set do not pronounce [s] in tokens which end in /s/, while the Gascon speakers did. Given that Languedoc generally does not permit word-final clusters and Gascon does (Oliviéri & Sauzet, 2016), I explored a subsample of tokens without /ʎ/ that end in complex codas. By doing so, I confirm that the Languedoc speakers in our data set simplify complex codas while Gascon speakers do not. In our Languedoc data, however, segments which are part of complex codas may remain if they are re-syllabified into the following word (1 & 2; neither *dins* nor *jusqu'ins* are nominal).

(1) /dins tuts/ [din.tut] *dins tots* 'in all'

(2) /ʒyskins al/ [ʒy.skin.sal] *jusqu'ins al* 'until the [morning]'

To our knowledge, cross-word resyllabification has not been attested in either Occitan variety.

ANALYSIS: Our data suggest that the depalatalization process described in the literature on Languedoc does not fully capture what occurs when /ʎ/ is word-final; [ʎ] is produced nearly as often as [l]. Rather, this work proposes that treating /ʎ/ in Languedoc as two segments rather than a single complex segment allows this process to be included in coda simplification. In the presentation, I will use CVCV analysis to demonstrate that Languedoc's depalatalization rule can be reanalyzed as one of coda simplification. This presentation: (a) offers a phonological analysis of /ʎ/-loss in modern varieties of Occitan, which is the first to my knowledge, (b) gives further insight regarding how phonetic cues may lead to phonological change, and (c) suggests that complex segments such as /ʎ/ may be reinterpreted as two separate phonological segments.

The velar nasal in French, new member of the French system?

Arielle Roadman (Indiana University), Corentin Mazet (Indiana University)

alebouli@iu.edu, cmazet@iu.edu

English borrowings into French are increasingly studied (e.g. Poplack et al., 1988; Chelsey, 2010; Ten Hacken & Panocová, 2020), with *-ing* forms becoming numerous (e.g *parking, bowling…*). While they were borrowed without awareness of morphological structure (Gottlieb & Furiassi, 2020), this awareness seems to be growing which may be introducing the velar nasal into the French phonemic system. We thus ask the question what are the phonological repercussions of *-ing* in Hexagonal French? Is the velar nasal the newest member of the French phonological system?

Historically, the velar nasal is absent from the French phonemic inventory (Walter, 1983; Walker, 2001), but the decreasing phonetic adaptation of English *-ing* borrowings has left its status unclear. Older studies reject it (Greenleaf, 1921), but more recently, studies attest the [ŋ] production without testing its phonemic status (Picone, 1996; Walker, 2001; Lewis, 2007). Some others view it as a marginal phoneme, restricted to the *-ing* suffix (see Bazell, 1952).

In this study, participants recorded themselves reading sentences which included attested and novel borrowings, which included the morpheme *-ing*, as well as other cases of English /ŋ/. Stimuli were further balanced for vowel backness in the non-*ing* condition and across conditions for the following context (consonant, vowel, sentence final). Productions were analyzed by trained phoneticians to identify the phoneme, and preliminary evidence shows that [ŋ] is produced consistently by speakers, with or without an added oral velar consonant [g]. This latter part is constrained by the phonological environment: when followed by a vowel or sentence final, [g] tends to be present, but when followed by a consonant, [g] tends to be absent. These results would suggest that the velar nasal is on its way to phonemicization as it is produced across word types, and does not require a strong association to the *-ing* morpheme to be produced.

Borrowings in Hexagonal French are understudied compared to those in other varieties of French, and, to our knowledge, no systematic test of the status of the velar nasal has been conducted before. The presence of [ŋ] in this variety of French leaves no doubt, but we show here that the phoneme holds a place and function in the phonemic inventory of the language. The use of the velar nasal is productive in the results of our study: we were able to find consistent production in attested loanwords as well as novel borrowings of various kinds (morphological *-ing*, non-morphological *-ing*, other cases of /ŋ/). These results supports an increasingly phonemic status of the velar nasal in Hexagonal French.

References

Bazell, C.E. (1952). Phonemic and morphemic analysis. *WORD, 8*(1), 33-38.

Casesnoves Ferrer, R., & D. Sankoff. (2004). "The Valencian Revival: Why Usage Lags behind Competence." Language in Society 33: 1–31. Print.

Chelsey. P. (2010). Lexical borrowings in French: anglicisms as a separate phenomenon. *French Language Studies. 20*, 231-251.

Coveney, A. (2001). *The sounds of contemporary French: articulation and diversity.* Exeter: Elm Bank Publications.

Dominicy, M. (2000). La dynamique du système phonologique en français. *Français moderne*, *68*(1), 17-30.

Furiassi, C. & Gottlieb, H. (2015). Pseudo-English, studies on false anglicisms in Europe. *Language Contact and Bilingualism, 9*, 1-35.

Goldsmith, J.A. (1995). *The Handbook of Phonological theory.* Cambridge, Massachusetts: Blackwell.

Greenleaf, J.H. (1921). Le [ɲ] ou "n mouillé". *The Modern Language Journal, 5*(7), 365-369.

Lewis, J. C. (2007). *The –ing suffix in French.* [Master's thesis, University of North Dakota].

Morin, Y. C. (2009). Histoire des systèmes phonique et graphique du français. *Romanische Sprachgeschichte/Histoire linguistique de la Romania*, *3*, 2907-2926.

Mott, B. (2015). The rise of the English -ing form in Modern Spanish: A source of pseudo Anglicisms. *Language Contact and Bilingualism, 9*, 175-795.

Pergnier, M. (1989). *Les anglicismes.* Paris: Presses Universitaires de France.

Picone, M. D. (1996). *Anglicisms, Neologisms and Dynamic French*, Amsterdam, Philadelphia: John Benjamins Publishing Company.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics, 26*(1), 47-104. 10.1515/ling.1988.26.1.47

Redman, R. (1978). L'adaptation phonétique des emprunts à l'anglais en français. *La linguistique, 14*(1), 111-124.

Rifelj, C. (1996). False friends or true? Semantic anglicisms in France today. *The French Review. 69*(3), 409-416.

Spence, N.C.W. (1987). *Faux amis* and *faux anglicismes* : problems of classification and definition. *Forum for Modern Language Studies. 23*(2), 169-183.

Ten Hacken, P., & Panocová, R. (2020). *Word Formation, Borrowing and their Interaction.* Edinburgh : Edinburgh University Press.

Walter, H. (1983). La nasale vélaire /ŋ/ un phonème du français?. *Langue française, 60,* 14-29.

Walker, D.C. (2001). *French sound structure.* Calgary: University of Calgary Press.

Walker, J. (2015). False Anglicisms in French: A measure of their acceptability for English speakers. *Language Contact and Bilingualism, 9*, 159-175.

Should phoneticians shift to Bayesian?

Yuhyeon Seo (seo86@purdue.edu), Purdue University

Scientific disciplines where quantitative analysis is paramount have been increasingly adopting Bayesian methodologies for statistical analysis (Cumming, 2013; Lee & Wagnmakers, 2013). In recent years, this trend has also emerged in the field of phonetic science, where empirical observations combined with statistical modeling play a key role in answering research questions. Pioneers have not only embraced this Bayesian shift but have also provided tutorials and instructions for phonetic research (Norris & McQueen, 2008; Vasishth et al., 2018). In this presentation, I will delve into Bayesian frameworks and highlight their inherent strengths, including incorporation of prior knowledge, modeling flexibility, interpretability, and analysis of individual differences. I will demonstrate these advantages through a case study investigating acoustical properties of Korean word-initial stops, produced by heritage speakers (HSs) (n = 27) and long-term immigrants (LTIs) (n = 23) in the US and L1-immersed Korean (L1-i) speakers (n = 29) in South Korea.

The goal of the study was to investigate L2 (English) cross-linguistic influence in speech production of L1 (Korean) stops by examining two acoustic correlates of the laryngeal contrasts in stops: voice onset time (VOT) and onset f0. While both languages employ the same acoustic cues for laryngeal stop contrasts, their exact phonetic implementations and perceptual importance differ. In Korean, onset f0 plays a primary role in differentiating lenis-aspirated and lenis-fortis stops, with VOT being a secondary or even an uninformative cue. In contrast, English primarily relies on VOT for demarcating voicing contrasts in stops, with onset f0 acting as a secondary cue. To juxtapose Frequentist and Bayesian approaches, two separate mixed-effects linear regression models were fitted in R (version 4.2.2, R Core Team, 2022) using both methods. The models examined VOT and onset f0 as dependent variables with Group (HSs, LTIs, L1-i: reference level), StopType (aspirated: reference level, lenis, fortis), and their interaction as fixed effects and by-subject intercepts and slopes and by-item intercepts as random effects: $Y(\text{VOT, onset f0}) = \beta0(\text{Intercept}) + \beta1(\text{Group}) + \beta2(\text{StopType}) + \beta3(\text{Group}\times\text{StopType}) + \gamma0(\text{Subject}) + \gamma1(\text{Subject}\times\text{StopType}) + \theta0(\text{item}) + \epsilon$). The ‘lme4’ package was used for the frequentist models, and the Bayesian models were built using the Hamiltonian Markov Chain Monte Carlo (MCMC) sampler function in Stan (Carpenter et al., 2017) and the ‘brm’ package (Bürkner, 2017). Each Bayesian model used 4 MCMC chains to draw 4,000 samples, with 1,000 warm-up iterations.

The comparative analysis suggests that Bayesian models offer advantages over the Frequentist counterparts. First, the Bayesian models successfully converged, while the frequentist model examining onset f0 resulted in a singular fit due to the complex parameter structure. Second, the incorporation of prior knowledge on the acoustic qualities of Korean stops increased the model’s predictability in Bayesian approach, as estimated by the Leave-One-Out Informative Criterion. Third, the Bayesian approach, with its capability to compute posterior distributions for all parameters including the random effects, is well suited for individual-level analyses (see Figure 1). Finally, the Bayesian framework, centered on parameter testing over traditional null hypothesis significance testing, allows for direct probability statements about parameters and hypotheses. However, I caution that Bayesian analysis is not a panacea. I suggest that the choice of framework should be guided by specific research questions, the nature of the data, and research goals. Both frameworks have unique strengths and can be employed in complementary ways.

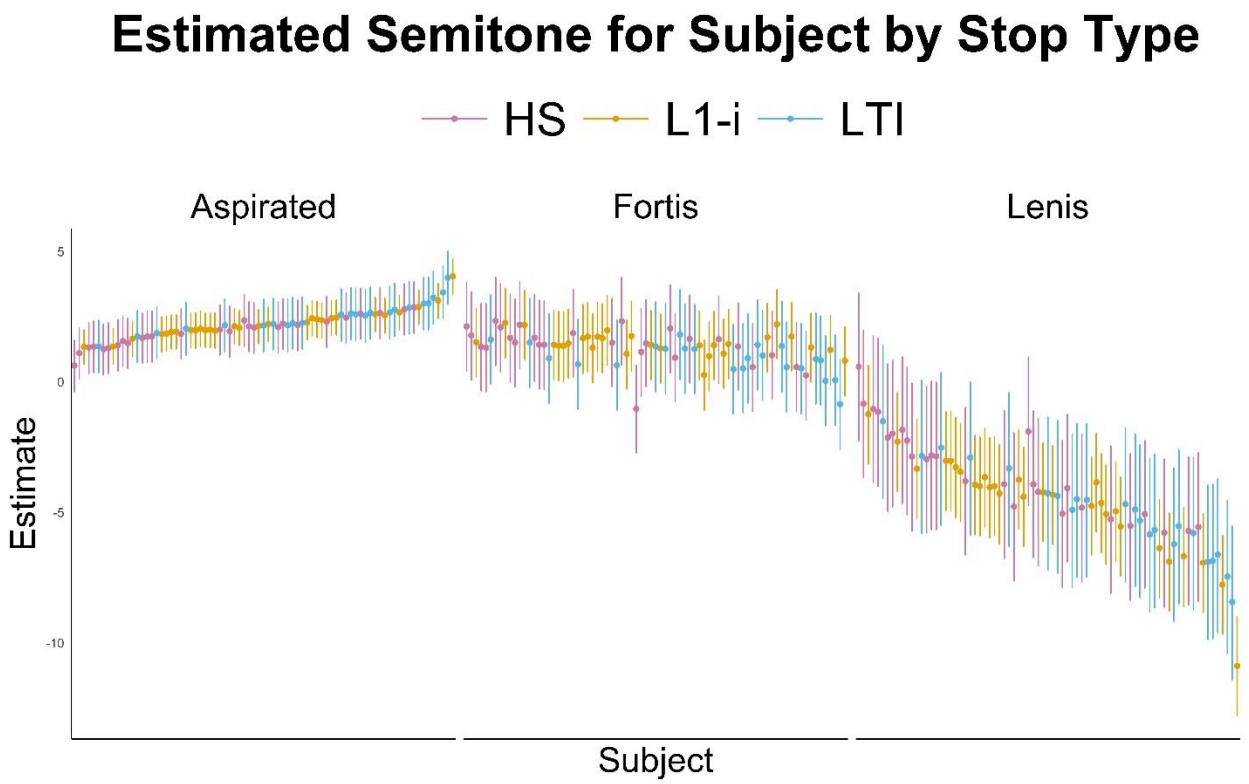


Figure 1. A caterpillar plot showing estimated semitones for each subject and each stop type by group, calculated based on posterior random effects.

Selected References

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1).

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.

R Core Team. (2022). R: A language and environment for statistical computing. URL https://www.R-project.org/

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161.

From Narrow Transcription to Dictionary Pronunciation (DP): A Trial with the Buckeye Corpus

Yucheng Song(songyuch@umich.edu)
San Duanmu(duanmu@umich.edu)
University of Michigan

The goal:

An important job in the description of a language is to compile a list of "dictionary pronunciations" (DP) of its words. In the traditional approach, we can obtain DPs by applying the phonemic analysis to narrowly transcribed words. In this study, we explore whether there is a more efficient way to obtain DPs.

Method:

We use the Buckeye Corpus as our data. The corpus contains about 40 hours of natural speech in American English (Mid-West variety), narrowly transcribed. Word tokens range from one to 12,348, and the number of different pronunciations of a word range from one to 355. We use an algorithm that involves four ordered rules, described below:

1. **Exclusions**: If a pronunciation of a word occurs at less than 1.5% of the token frequency of the word, we exclude it, because it is often a mispronunciation or a transcription error, or it is unlikely to be a DP.
2. **Longest**: Among different pronunciations of a word, keep those that are longest (having the most number of phones).
3. **Full vowel**: Among different pronunciations of a word, keep those that contain at least one full vowel, which is one of [ɑ æ ɔ ɛ eɪ i o u aʊ aɪ oɪ].
4. **Frequency**: Among different pronunciations of a word, keep those that have the highest frequency.

For convenience, we compare our results with standard dictionary pronunciations (SDP). An examination shows that for words whose type count is 1-10, only 59% of them contain SDP; for words whose type count is above 10, 96% of them contain SDP. Therefore, we focus on the latter set of words.

Results and discussions:

The predictions of our algorithm achieved a success rate of 85% (out of 82,214 tokens of 1,544 word types), which we consider encouraging. A number of future studies are worth exploring. First, we can examine the 'wrong predictions' and see whether some of them are possible dictionary pronunciations (either to be added to the one in the latter or to replace it). Second, the 'wrong predictions' are likely to point to ways to improve our algorithm and increase the success rate. Third, we should explore how to handle words with low occurrence tokens. Finally, we can examine allophonic rules in American English, which are likely to be far more than the twenty or so as given in Ladefoged & Johnson (2014).

Accommodation of Rising and Falling Intonation in American English

Thomas Sostarics and Jennifer Cole; Northwestern University
tsostarics@u.northwestern.edu jennifer.cole1@northwestern.edu

In American English, declarative sentences with a phrase-final falling intonational tune (e.g., *It's raining.* ToBI annotation: H*L-L%) are canonically interpreted as assertions—the speaker is conveying new information. The same sentence uttered with phrase-final rising intonation (ToBI: L*H-H%) is typically interpreted as a question—the speaker is seeking information. Recently, we have investigated whether variation in the question vs. assertion interpretation relates to variation in the edge tones (where pitch rises *towards*) or the pitch accent (where pitch rises *from*). Across several binary forced choice experiments (Sostarics and Cole 2023a; 2023b) we have found that variation in *Asking* vs. *Telling* interpretation is largely explained by the scaling of the final pitch target (i.e., *towards*, and not *from*). Yet, there are regions of our pitch continua where participants are less certain about their interpretation, with proportions of *Telling* responses close to chance (50%). It remains unclear whether the pitch contours at these steps are truly **ambiguous** or whether there are **distinct** meanings beyond the assertion/question contrast that participants couldn't directly judge within our binary-choice task.

Methods. We extend our previous paradigm where participants identified stimuli (short sentences) from a rising/falling pitch continuum as ***Asking*** or ***Telling*** to include (1) a third ***Other*** choice and (2) a free-text response where the participant describes their interpretation for some or all trials where they responded ***Other***. We investigate our forced choice results using a nested approach: what is the probability that a participant **accommodates** a pitch contour as either ***Asking* or *Telling***? Then, given those trials where participants **did** respond with either *Asking* or *Telling*, what is the probability of a ***Telling*** response specifically? We predict that when the uncertainty in interpretation is "absorbed" by the ***Other*** response option, the accommodated asking/telling responses that remain will drift **away** from chance (50%) and move closer to ceiling (100% telling) or floor depending on whether the global contour shape is falling or rising, respectively. Moreover, we predict that continuum steps that were closest to chance in our previous experiments would be more likely to receive an ***Other*** response here. We recruited 42 participants from Prolific to judge five repetitions of 34 pitch contours drawn from our previous experiments (170 trials per participant).

Results. Participants were generally willing to accommodate all contours as ***Asking*** or ***Telling***—***Other*** responses ranged from 2-11% across contours. In the aggregate results, the continuum steps that were previously closest to chance (plateaus and mid-level rises) do not map neatly to the rising/falling groups as we predicted. However, participants' **own interpretations** are more consistent: participants will idiosyncratically accommodate these contours as either ***Asking*** or ***Telling***, then continue to give that same response on later repetitions.

The free-text responses show three themes as to why participants selected ***Other*** during the forced choice task. First, some participants may interpret a contour as not **merely *Asking*** or ***Telling***, but rather there is an additional salient (perhaps paralinguistic) meaning: "*The speaker has a little bit of suprise in their tone towards the end"* (for a mid-level rise). Second, some participants interpret some contour as having a **distinct** meaning that lies outside the asking/telling dimension: "*I responded 'other' as it sounds more like the inflection comes from the speaker* ***listing off things****"* (for a plateau-like rise). Finally, some participants report **metalinguistic uncertainty**: *"This sounds like they're telling but might also be asking"* (for a mid-level rise). Ultimately, we find that variation in interpretation comes from a variety of distinct sources at the participant level. Future paradigms would benefit from creative ways to disentangle sources such as uncertainty in a judgment from distinct or secondary meanings.

Speech rate in the encoding of politeness for Spanish offers

Bruno Staszkiewicz
bstaszki@purdue.edu

Whereas seminal works in the fields of phonetics (Ohala, 1984) and pragmatics (Brown & Levinson, 1987) have suggested that higher pitch may be employed to encode politeness, recent investigations have suggested that formal speech correlates with lower pitch (Brown et al., 2014; Hübscher et al., 2017). However, these recent works failed in conducting a systematic analysis of the contextual variables of power, distance, and imposition to frame and analyze politeness. Given the general lack of systematic analysis of context on other suprasegmental features, the current study investigates whether the variables of power, distance, and imposition affect speech rate in Spanish offers. The hypothesis for this experiment is that a lower speech rate will occur in more polite situations.

To answer this research question, 35 Spanish native speakers completed a contextualized sentence-reading task where they read aloud 8 paragraph-length contextualizing situations, followed by an offer. The situations were balanced for two levels of power (high/low), distance (high/low), and imposition (high/low). Participants produced a total of 840 sentences. Analysis using Praat analyzed the syllable duration of the target sentences.

A linear regression model analysis was conducted in R to observe the effect of the contextual variables on the use of speech rate. The results showed that the variables of distance and imposition had a statistical effect ($p < 0.001$) on the use of pitch rate, which increased when the speakers did not know each other (see Figure 1), and when the speaker offered to do something that was of high cost to accomplish (see Figure 2). However, the variable of power was not found to be a significant factor that affects the use of speech rate. The overall findings of this project shed light on the effect of speech rate that can be an indicator of different pragmatic meanings related to politeness.

Figures

Figure 1. Duration per syllable and Distance

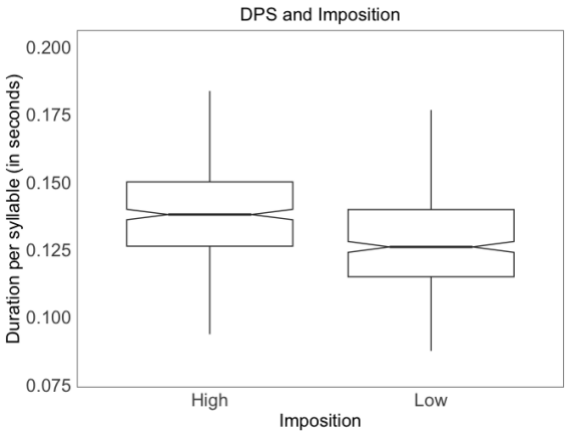


Figure 2. Duration per syllable and Imposition

Selected references

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, *66*, 45–60.

Hübscher, I., Borràs-Comes, J., & Prieto, P. (2017). Prosodic mitigation characterizes Catalan formal speech: The Frequency Code reassessed. *Journal of Phonetics*, *65*, 145–159.

Title: The cost of listening to unfamiliar accents

Authors: Grace E. Teuscher

Washington University in St. Louis, Dept of Psychological & Brain Sciences

g.teuscher@wustl.edu

Drew J. McLaughlin

Basque Center on Cognition, Brain and Language

d.mclaughlin@bcbl@eu

Kristin J. Van Engen

Washington University in St. Louis, Dept of Psychological & Brain Sciences

kvanengen@wustl.edu

Abstract:

The subjective ease of understanding accents that differ from a listener's has typically been assessed using self-reports. This approach, however, relies on metacognitive judgments that are difficult to interpret, inconsistent across participants, and may not converge with objective measures of effort. To address this challenge, this study utilizes effort discounting, a paradigm borrowed from behavioral economics. In the experiment, participants are first familiarized with one L1 English speaker and three L2-accented speakers of varying intelligibilities. After familiarization, participants are given the choice of listening to the L1-accented speaker for a smaller monetary reward or to one of the L2-accented speakers for a larger reward. By varying the reward offered for the easier option based on previous choices, the subjective value of the effort expended for each L2 speaker can be determined. The lowest monetary selection for each L2-accented speaker is obtained, which represents the most amount of money participants are willing to forego in order to not have to expend the extra effort that listening to an unfamiliar accent requires. Data collection is ongoing, and preliminary data will be presented. We expect participants will remain willing to listen to highly intelligible L2-accented speakers. However, as speakers become less intelligible, participants will be less willing to expend the required additional effort and will choose the easier speaker. We also predict that participants who rate L2-accented speakers lower on an affect and attitudes questionnaire will be more likely to discount their reward to avoid the effort required by the L2-accented speech.

Phonological Symmetry Does Not Help Generalization of Perceptual Adaptation to Vowels

Züheyra Tokaç, Jennifer Cole (Northwestern University)
zuheyratokac@u.northwestern.edu, jennifer.cole1@northwestern.edu

Lexical context can guide how an ambiguous speech sound is perceived—the same ambiguous stop sound from a /d-t/ continuum may be perceived as /d/ in a *dash-tash* continuum, but as /t/ in a *task-dask* continuum (Ganong, 1980). The lexically guided perceptual adaptation paradigm makes use of this phenomena to show that previously learned phoneme boundaries are not rigid but are subject to change following listening exposure to novel input (Norris, McQueen, and Cutler, 2003). For instance, when exposed to words in which /i/s are replaced by a vowel that is ambiguous between /i/ and /e/ (i.e., words with lowered /i/s), listeners identify more steps of an /i-e/ continuum as /i/ compared to their pre-exposure identification, i.e., lower their /i-e/ crossover boundary (see Figure 1). What is not yet clear is whether perceptual adaptation in vowels is specific to the phonemes listeners were exposed to or if listeners generalize the shift to phonologically related *novel* phoneme contrasts (not presented in exposure trials). Testing whether exposure to lowered or raised front vowels impacts back vowel identification in English and Dutch, Maye, Aslin, & Tanenhaus (2008) and McQueen & Mitterer (2005) find adaptation to front vowels but mixed results for generalization to novel back vowels, whereas Chládková, Podlipský, & Chionidou (2017) find that Greek listeners lower or raise their perceptual boundary between /i/ and /e/ and generalize the shift to novel /u/ and /o/. Chládková et al. argue that generalization is facilitated by the phonologically symmetrical vowel inventory of Greek where phonological similarities across vowels might be more salient than in English or Dutch, whose vowel systems require a greater number of specified features. Alternatively, generalization might have been facilitated by the smaller size of the 5-vowel Greek inventory in contrast to the larger inventories of English and Dutch. The present study replicates Chládková et al. in Turkish, which has a fully symmetrical vowel inventory of 8 vowels. Turkish listeners (N = 20) were exposed to ambiguous front vowels in lexically /i/-biasing or /e/-biasing contexts (i.e., lowered /i/ and raised /e/ conditions), and performed vowel identification tasks on /i-e/ and /u-o/ vowel continua. We predicted the full symmetry of Turkish vowels to facilitate generalization and yield boundary shifts in both the front and back vowel conditions. Although there was a marginally significant interaction of condition and test block suggesting boundary lowering in the lowered /i/ condition (p = 0.06), this result was qualified by a marginally significant interaction with vowel continuum (p = 0.07), and pairwise comparisons revealed that the boundary lowering was significant only in the /i-e/ continuum (see Figure 2). These results do not replicate Chládková et al. for Turkish, suggesting that phonological symmetry might not be a sufficient condition for generalization of perceptual adaptation in vowels. Moreover, the size of the vowel inventory might be a stronger predictor of generalization of perceptual adaptation, with the 5-vowel inventory of Greek facilitating generalization but not the larger vowel inventories of Turkish, English, and Dutch.

Figure 1: Boundary shift

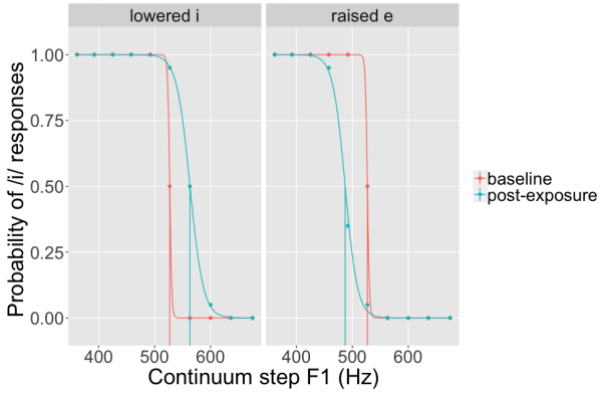
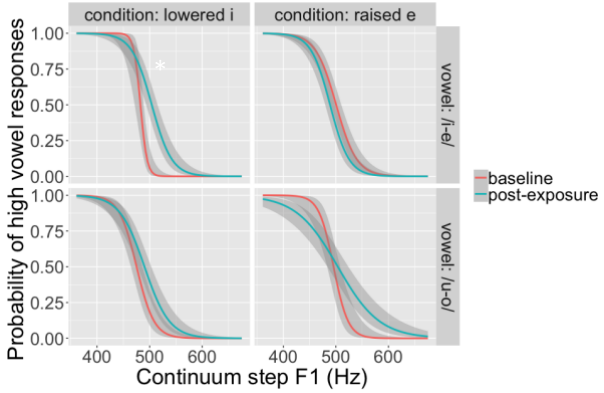


Figure 2: Vowel identification results



Airstream Mechanism Compensation in Speakers with Laryngeal Dystonia

Tzu-Pei Tsai[1], Chang-Jung Lu[2], Shyue-Yih Chang[3], Tessa Bent[1]

[1]Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN.; [2]Department of Clinical Speech and Language Studies, Trinity College Dublin, Dublin, Ireland; [3]Voice Center, Cheng Hsin General Hospital, Taipei, Taiwan.

tsaitz@iu.edu; luc3@tcd.ie; hydiung@gmail.com; tbent@iu.edu

Adductor laryngeal dystonia (AdLD) is a neurological condition that affects the vocal folds with uncontrollable closing movements during speech (Simonyan et al., 2021). This uncontrollable laryngeal activity can interrupt the coordination of airstream mechanisms for speech sounds. AdLD mostly impacts stability, loudness, and quality of vowels and voiced consonants. However, it can also affect the timing of consonantal gestures (Yanagida et al., 2015). To date, no study has documented obstruent substitution in this population. Thus, we investigated phonetic alteration of airstream mechanisms for obstruents from clinical records. Sound recordings were retrieved from a database of routine clinical speech-language assessment at voice clinic at Cheng Hsin General Hospital, Taipei, Taiwan. Recordings included 21 clients before and 15 after botulinum toxin injection, which can provide temporary relief from AdLD symptoms. A sound sample from each client included 53 obstruents for analysis. Speech tasks in the sample included counting from 1-10 and reciting of a phonetically balanced article consisting of 78 syllables. The sounds with changes in airstream mechanisms were annotated and transcribed by two researchers. Transcription consensus was reached through discussion. Consonantal deviations were frequent. Three instances of clicks in total from 3 clients were found for /p/ and /t/. Ejectives as substitutions for stops, fricatives, and affricates were observed in 6 clients for a total of 18 instances. Certain places and manners of articulation and phonetic environments were associated with more frequent changes in airstream mechanism. These velaric and glottalic airstream mechanisms appeared as compensation for an interrupted pulmonic airstream and were only found in recordings taken before botulinum toxin injection. Ejective and click substitutions have been found in speech in patients with other speech and hearing related disorders (e.g., Ball & Müller, 2007). Our findings provided further evidence of these processes stemming from neurological and motor disturbances in speech. Future work should consider the implications of these findings for phonological theories and models for speech production.

References

Ball, M. J., & Müller, N. (2007). Non-pulmonic-egressive speech in clinical data: A brief review. *Clinical Linguistics & Phonetics*, *21*(11–12), 869–874.

Simonyan, K., Barkmeier-Kraemer, J., Blitzer, A., Hallett, M., Houde, J. F., Kimberley, T. J., … Sharma, N. (2021). Laryngeal dystonia: Multidisciplinary update on terminology, pathophysiology, and research priorities. *Neurology*, *96*(21), 989–1001.

Yanagida, S., Nishizawa, N., Mizoguchi, K., Hatakeyama, H., & Fukuda, S. (2015). Voice onset time for the word-initial voiceless consonant /t/ in Japanese spasmodic dysphonia—A comparison with normal controls. *Journal of Voice*, *29*(4), 450–454. https://doi.org/10.1016/j.jvoice.2014.09.028

Intonation meets emotion: Evidence from production for robust phonological contrasts in F0

Daniel R. Turner (dturner@u.northwestern.edu) & Jennifer S. Cole (jennifer.cole1@northwestern.edu), Northwestern Univ.

Intonation serves to encode cues to linguistic meaning as well as the speaker's emotional state. Experimental paradigms used in intonation research are typically designed with emotionally neutral contexts, presumably to avoid introducing linguistically irrelevant acoustic variation. The current study takes a different approach by jointly modeling the effects of pragmatic meaning and speaker emotion on F0 in Mainstream American English (MAE) in order to explore how emotional variation influences the production of linguistically meaningful intonation. Our focus is the "nuclear tune", phonologically defined by the sequence of prosodic phrase-final pitch accent, phrase accent and boundary tone according to the Autosegmental-Metrical (AM) model [1]. In their simplest form (setting aside bitonal accents and downstepped High), tunes consist of a sequence of three High or Low tone features, yielding an inventory of 8 'basic' forms (HHH, HHL… LLL). Each tune is predicted to yield a distinct phrase-final F0 trajectory [2], yet recent work shows that certain tunes are not reliably distinguished in production or perception [3], especially tunes whose F0 trajectories are of similar shape and which differ only slightly in the magnitude or alignment of an F0 rise or fall. This finding can be viewed as evidence for a smaller inventory of intonational tunes in MAE, but another hypothesis is that the neutral emotional context is damping the contrastive phonetic implementation of tunes. We seek to answer two questions: (1) Does the emotional context modulate the predicted phonological distinctions among tunes in speech production? If so, is there a specific emotional context in which speakers reliably implement all of the predicted tune contrasts? (2) Are the effects of a given emotion on F0 consistent across tunes? If so, listeners may be able to parse effects on F0 due to emotional context to aid in identifying the phonologically specified and linguistically meaningful intonation. We investigate these questions through acoustic analysis of F0 in the production of nuclear tunes under specific enacted emotions. We test the 'basic' tunes from the AM model of MAE across four emotions that differ along the psychometric dimensions of Valence (V) and Potency (P) [4]. According to the emotion model, V and P affect the acoustic expression of speech along acoustic dimensions shared by intonation, making them appropriate for our experimental manipulation, which was operationalized using counterbalanced emotions [4]: love [+V, -P], pride [+V, +P], shame [-V, -P], and anger [-V, +P]. **Methods.** Tunes were elicited using the paradigm of imitated intonation from [3], where participants hear a tune presented in an auditory model sentence (*Her name is Marilyn*) and reproduce the tune over new text (*From Madelyn*). Tunes were produced on the final word, always a name. Unlike [3], the intended emotional state was identified for each target sentence, which was presented in a brief discourse context congruent with the emotion (Fig. A). Participants were recruited in two groups: 13 voice actors ('VA'; 7 female, 6 male; avg. age 26.5 years) from an industry crowdsourcing platform and 19 untrained speakers from a subject pool ('SP'; 17 female, 2 male; avg. age 19.5 years). Prior work has shown voice actors are a reliable source for authentic-sounding vocal cues of emotion [5]. VAs produced 200 target sentences (8 tunes x 5 emotions x 5 reps) while SPs produced 120 (8 tunes x 5 emotions x 3 reps), for a total of 4,880 sentences. The final dataset excludes 116 recordings with audio artifacts or background noise, leaving 4,764 recordings total. We extracted the time-normalized F0 trajectory from each tune-bearing word and submitted them to k-means clustering and a GAMM based on [3]. The clustering algorithm groups trajectories based on shape, while the GAMM uses tune and emotion as separate predictors to shed light on how tunes are encoded given a particular emotion, and to gauge the generalizability of emotion's effect on the F0 dynamics of tunes. We hypothesized that accounting for emotional variation would lead to stronger observed tune contrasts and emotions would affect tunes in a consistent manner. **Results.** The clustering model showed 5 emergent groups (Fig. B), similar to the results of [3], suggesting emotion played a secondary role in determining the final F0 trajectory. Comparing our tune distribution across clusters (Fig. C) to [3], we found similar patterns of converging tunes produced in emotional contexts as [3] found in neutral ones. In the GAMM, emotions structure within-tune variation in similar ways, meaning that tune shape was generally preserved (Fig. D), which accords with the clustering results and [3]. There was no single emotion that supported an 8-way tune distinction, but each pair of confusable tunes (based on [3]) were significantly distinct in at least one emotion according to a difference GAMM analysis. The effect of V on tunes was more transparent than P. **Conclusion.** Intonational contrasts, as represented by tunes, are generally preserved despite speaker emotion, but there remain problematic gaps in the set of contrasts speakers produce.

Fig. A

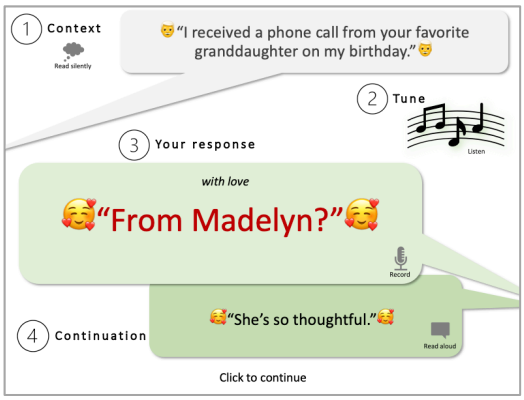
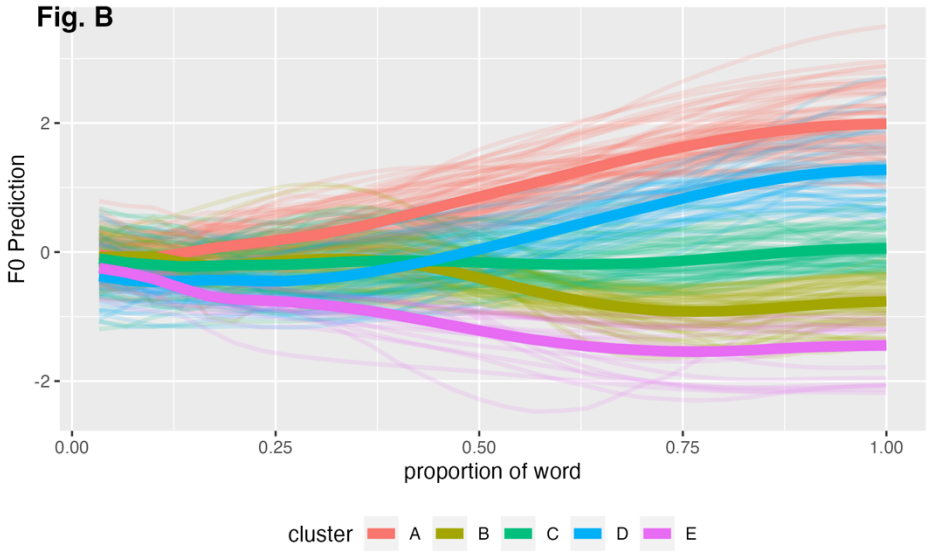


Fig. B



C

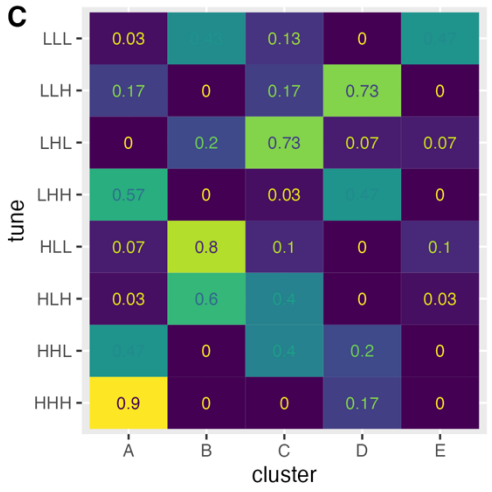
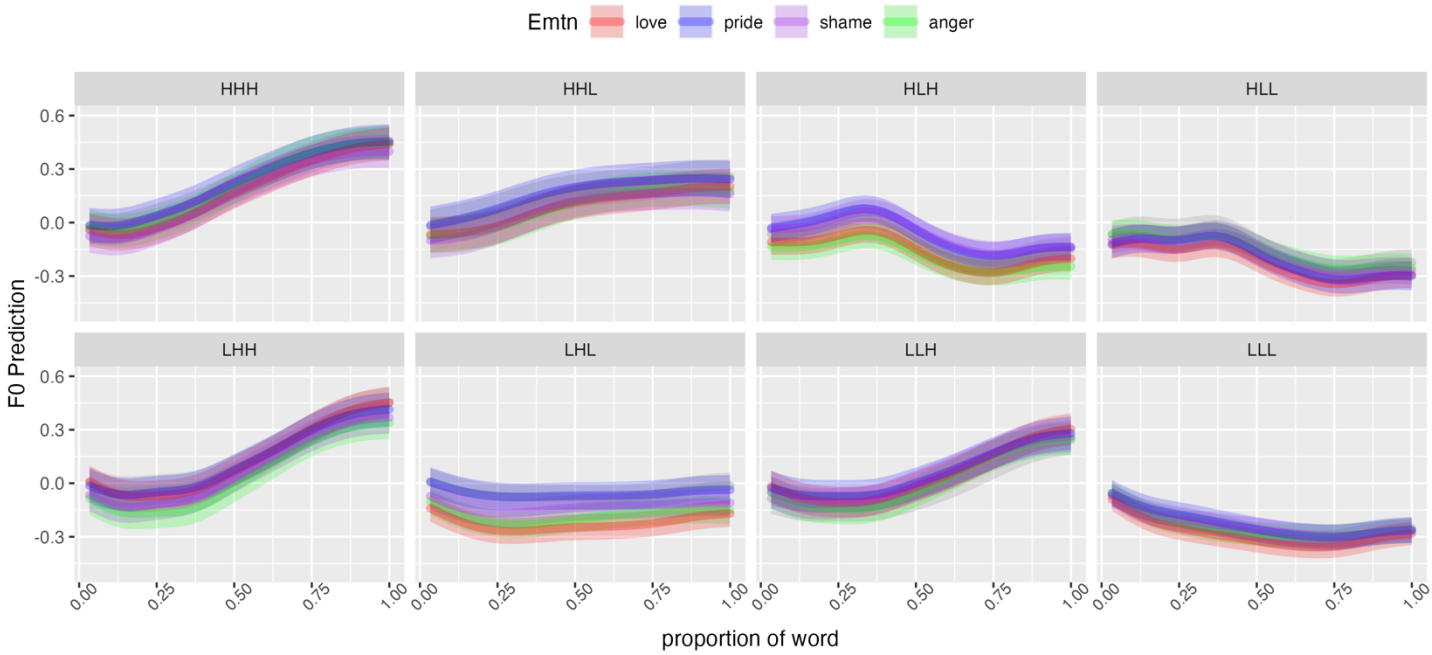


Fig. D



References. [1] Ladd, D. R. (2008). *Intonational phonology* (2nd ed). Cambridge Univ. Press. [2] Pierrehumbert, J. B. (1980). [PhD Thesis]. MIT. [3] Cole et al. (2023). *Laboratory Phonology*: 1-51. [4] Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). *Psychological Science,* 18 (12), 1050-1057. [5] Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). *Emotion*, 12(5), 1161–1179.

Investigating the Variability of External Sandhi Application
– A Case Study of Morpho-syntactically Complex Mandarin Tone 3 Sandhi

Chang Wang and Jie Zhang, University of Kansas (changwang@ku.edu, zhang@ku.edu)

External sandhi refers to phonological alternation across word boundaries. Previous literature documented much variability in its application. The Production Planning Hypothesis (PPH) accounts for the variability by proposing that the realization of external sandhi is constrained by the availability of phonological information across words during online production planning [1].

In this study, we investigated how two specific factors, syntactic boundary and speech rate, influence the application of morpho-syntactically complex Mandarin Tone 3 sandhi (T3S). In Mandarin, Tone 3 (T3) surfaces as Tone 2 (T2) when it immediately precedes another T3 syllable. Its application is complex and variable at the phrasal level [2]. According to PPH, the phonological context embedded in the forthcoming word is less likely to be available across a larger syntactic boundary during the current word production. Consequently, there is a decrease in the likelihood of T3S application across a bigger boundary. A slower speech rate is found to be associated with a more incremental planning strategy and a narrower planning window. As a result, it's less likely for the upcoming word to be available at a slower speech rate, leading to a lower likelihood of T3S application.

We conducted a production experiment of Mandarin T3 sequences that manipulated the syntactic boundary in the stimuli (see an example in Table 1) and speech rate using blocked-design SOAs. Syllables that immediately precede the manipulated syntactic boundaries were directly compared on their T3S application.

Results showed that syntactic boundary significantly predicts T3S application when lexical frequency is comparable. The predicted probability of T3S application for the larger boundary Verb-Object is marginally lower than the smaller boundary Modifier-Noun (estimate = -1.328, p = .064). No significant differences were found for Verb-Object vs. Subject-Predicate (p = .229) or Modifier-Noun vs. Subject-Predicate (p = .413). Speech rate and any interaction terms involving it do not significantly contribute to the model, indicating that no speech rate effect was observed. The syntactic boundary effect Verb-Object < Modifier-Noun aligns with the prediction of PPH that external sandhi application is less likely to apply across a larger syntactic boundary. The unexpectedly high application likelihood of the Subject-Predicate might be caused by its significantly higher trigger word frequency than Modifier-Noun and Verb-Object (p < .001 for both comparisons). The lack of speech rate effect suggests that speech rate alone likely cannot influence external sandhi application or production planning in general.

Table 1. An example of stimuli with three different syntactic structures. Only the T3S application of the first syllable was compared across the syntactic boundary condition.

[σ[σσ]]	**Modifier-Noun**	**Verb-Object**	**Subject-Predicate**
Example	nü3. sheng3.zhang3	zhu3.shui3.jiao3	ma3. xiang3.pao3
Gloss	woman governor	boil dumpling	horse want run
Meaning	'female governor'	'boil dumpling'	'The horse wants to run.'

[1] Wagner, M. (2012). Locality in Phonology and Production Planning. In J. Loughran & A. McKillen (Eds.), *Proceedings of phonology in the 21 century: Papers in honour of Glyne Piggott* (Vol. 22, pp. 1–18).

[2] Lin, Y.-H. (2007). *The sounds of Chinese*. Cambridge University Press.

Glides in Belizean Kriol: [j] from /i/ after syllable-initial oral stops

Jarod Warner, Indiana University

jacwarne@iu.edu

In Belizean Kriol (an English lexifier creole) syllable-initial oral stop consonants may be followed by a glide prevocalically. This presentation examines the syllable structure of Belizean Kriol (BK) to determine specifically if the palatal glide [j] is in the underlying representation and if not, how it surfaces. The study applies primarily to basilectal BK, or the variety that is furthest away phonologically and syntactically from the lexifier, English. I hypothesize that the glide is part of the onset as a result of the surfacing of underlying /i/ as [j]. Importantly, the phonotactics of syllable structure in BK will not permit the sequence [iɑ], a high diphthongal onglide. This will be critical into advancing an explanation of the surfacing of the front high vowel as a glide. As a further elaboration of this theory, I will show that while syllable-initial oral stop-glide sequences occur before [ɑ] in BK, not all oral stops followed by [ɑ] will have a prevocalic glide. This indicates that it is not merely the presence of [ɑ] which produces the preceding glide.
In addition, the theory proposed opens a door to an account of affrication of syllable-initial alveolar oral stops with /j/ in some BK words. The analysis proposes that an underlying glide undergoes hardening into a fricative, with the stop and subsequent fricative becoming an affricate. This is another indication that underlying /i/ results in a surface palatal glide whereas underlying /j/ has a different surface representation.

Here the analysis is represented in Optimality Theory. This approach was chosen to account for the fact that Jamaican Creole (JC), with very similar lexifier and substrate language makeup, shows rising sonority in diphthongs while BK does not. As will be seen, BK has constraints against metathesis (LINEARITY) as well as sonority rise in the nucleus. That is, the segments [iɑ] in the nucleus exhibit sonority rise, which is a violation of the constraint SONORITY FALL. Thus, these segments cannot be a diphthongal segment. If the segments are reversed, sonority fall is satisfied, but the constraint against metathesis is violated. Metathesis is disallowed in BK as otherwise the JC diphthongs would surface yet BK sees glide-vowel sequences instead. JC on the other hand, has a different ranking of constraints, to be contrasted with BK in tableaus.

Representing Exceptions to Turkish Vowel Harmony

Jae Weller & Eric Raimy
jdweller@wisc.edu, raimy@wisc.edu
University of Wisconsin – Madison

Turkish is well-known for vowel harmony wherein vowels match the leftmost vowel of a word in terms of rounding (high vowels only) and backness [GK], (1). Words can be disharmonic in two ways. First is morpheme-internal disharmony (MID), where vowels within a morpheme do not harmonize with preceding vowels in rounding/backness, (2), but some speakers regularize these forms [CS, P]. MID can be observed in 'exceptional' suffixes with non-alternating disharmonic vowels (bolded), (3). The second disharmony type is root-suffix disharmony (RSD) where roots take disharmonic suffixes, (4).

We provide an account of Turkish vowel harmony, including both disharmony types, based on Precedence and Relation Oriented Phonology, PROP [Pa], and Modified Contrastive Specification, MCS, [D, PRS]. PROP posits directed graphs to represent precedence relations, with parallel paths encoding autosegmental information [Pa, G]. MCS provides a theory of which distinctive features are phonologically active with PRS developing a ternary distinction among *marked*, *unmarked*, and *unspecified* features. These proposals represent the Turkish vowel system as two archi-segments [CH] contrasting in height (TH 'high' vs. DOR^{TH} 'non-high' [PRS]) with backness (TT 'front' and DOR^{TT} 'non-front') and rounding (LAB 'round' vs. OP^{LAB} 'non-round') as parallel autosegmental features for each morpheme. Vowel harmony and disharmony is based on scope and layering of the parallel backness and rounding features following PROP.

(5-6) present PROP representations for a harmonic root, *œvyntʃ* 'pride', and the suffix *-lAr* PLURAL. (5) is harmonic for rounding and backness because the features LAB and TT are in a parallel precedence stream with the whole representation. Height specification for each individual vowel has been deautosegmentalized [G] to combine with the vowel feature to form an archi-segment (vowel + TH = high, vowel + DOR^{TH} = non-high). (6) is *unspecified* for any backness feature so will harmonize for backness but it is *unmarked* for roundness due to the presence of the parallel OP^{LAB} feature. This structure will harmonize following vowels for frontness. (7) demonstrates how a PROP representation produces both harmony and disharmony effects from a single principle. The vowels in (7) are archi-segments that contrast for height but then harmonize for backness and roundness. The TT 'round' feature is parallel across the whole representation producing 'front harmony'. The *lar* suffix adds a 'non-back' (OP^{LAB}) feature only parallel to the suffixes (material in the two boxes) which blocks the LAB feature from the root from the suffixes. The nested relationship between LAB (outside) and OP^{LAB} (inside) produces a disharmonic structure. (8) presents the disharmonic root, MID, which simply has nested specifications for backness. (9) shows an RSD root, *kalp* 'heart', which encodes its suffixes as 'front' through the TT feature at the end. (10) shows that suffixes added to (9) nest into the TT feature causing the disharmony.

PROP and PRS provide representations that encode both vowel harmony and disharmony with a single set of representations/principles based on required phonological material. No exception diacritics are required. Other examples of disharmony in Turkish will be discussed with ramifications about 'exceptions' in harmony systems being explored.

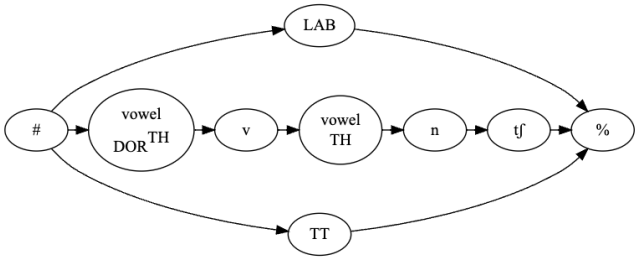
(1) sarɯmsak ‘garlic’ hedije ‘gift’ œvyntʃ ‘pride’

(2) anne ‘mother’ elma ‘apple’ menkɯbe ‘legend’
manikyr ‘manicure’ blendɯr ‘blender’ ʃofœr ‘chauffeur’

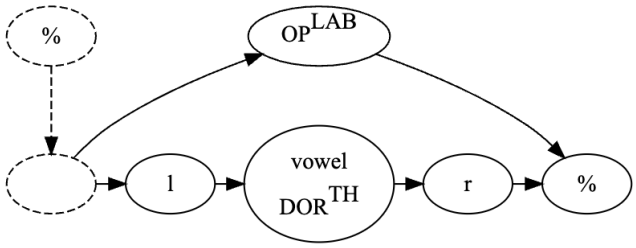
(3) Imperfective: gel-ij**o**r ‘he/she/it is coming’
Converbial marker: otur-k**e**n ‘while sitting’
Ability: oku-yab**ili**r ‘he/she/it is able to read’

(4) onun saat-**i** ‘his/her/its hour’ onun petrol-**y** ‘his/her/its oil’
kalp-l**e**r ‘hearts’ normal-d**e** ‘normally’

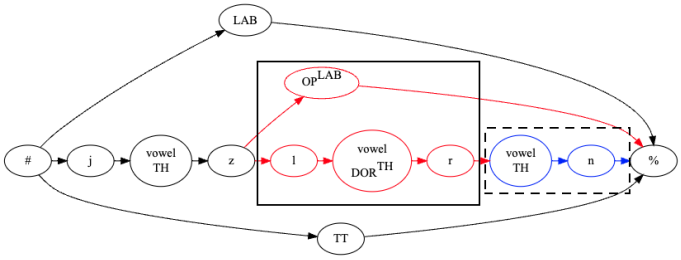
(5) Harmonic root *œvyntʃ* ‘pride’



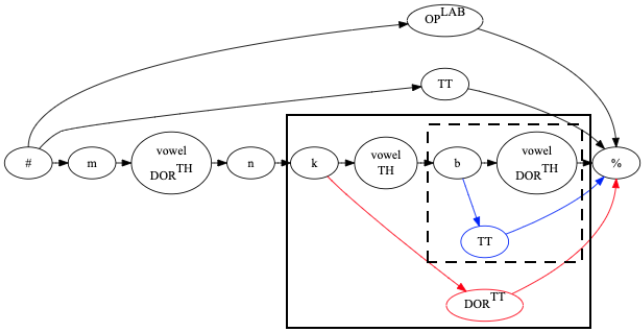
(6) *-lAr* PLURAL suffix



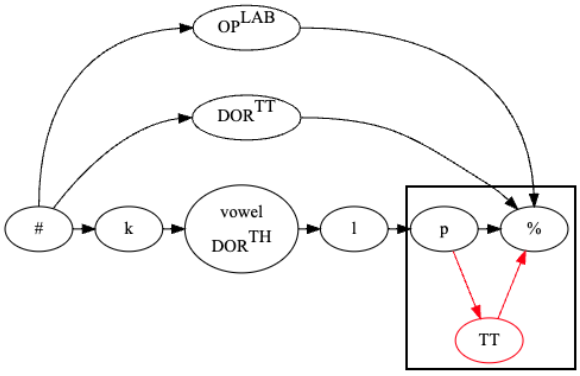
(7) Harmonic *jyz-lar-in* 'your faces'



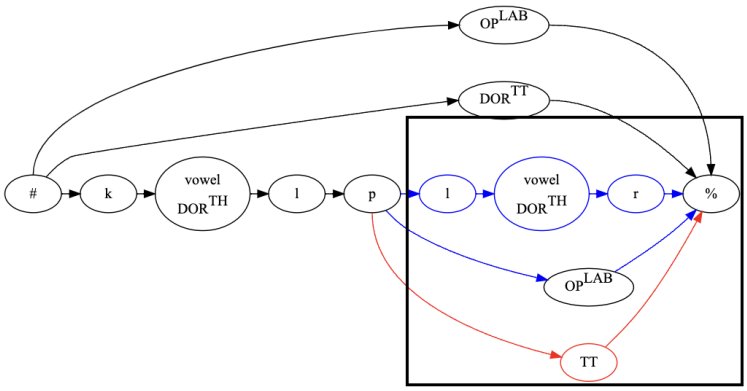
(8) Disharmonic root *menkube* ‘legend’



(9) Post disharmonic root *kalp* 'heart'



(10) Disharmonic *kalp-ler* 'hearts



[CH] Chomsky, N & M. Halle. (1968). *The sound pattern of English*. Harper & Row.

[CS] Clements, G.N & E. Sezer. (1982). Vowel and Consonant Disharmony in Turkish. In H. van der Hulst and N. Smith (eds.) *The structure of Phonological Representations (Part II)*. Dordrecht, Foris publications. pp. 213 - 255.

[D] Dresher, B. E. (2009). *The contrastive hierarchy in phonology*. CUP.

[G] Goldsmith, J. (1976). *Autosegmental phonology.* PhD Dissertation, MIT.

[GK] Göksel, A. & C. Kerslake. (2005) *Turkish: A Comprehensive Grammar*. London: Routledge.

[P] Pöchtrager, M. A. (2010). Does Turkish Diss Harmony? *Acta Linguistica Hungarica*, 57(4), 458–473. http://www.jstor.org/stable/26190972

[Pa] Papillon, M. (2020) *Precedence and the lack thereof: Precedence-relation-oriented phonology*. PhD Dissertation, University of Maryland

[PRS] Purnell, T.C., Raimy, E., & Salmons, J. (2019). Old English vowels: Diachrony, privativity, and phonological representations. *Language* 95(4), e447-e473. doi:10.1353/lan.2019.0083.

On Factors that Trigger an Assimilative Parsing over a Dissimilative Parsing

Yuka Tashiro | Indiana University | yutash@iu.edu

The lack of invariance problem (Liberman et al., 1954) refers to the question of how listeners parse coarticulated speech to retrieve intended phonemes in speech perception. How do listeners parse coarticulated speech so successfully? There are two major types of parsing mechanisms proposed in the literature, which often make opposite predictions. I refer to these parsing mechanisms as *dissimilative parsing* (e.g., Mann, 1980) and *assimilative parsing* (e.g., Fujimura et al.). In dissimilative parsing, listeners parse two successive segments to identify one as more different from the other, while in an assimilative parsing, listeners parse two successive segments in a way that identifies one as more similar to the other. For example, an ambiguous fricative noise between [s] and [ʃ] (target) may be identified as [s] more often when its phonetic context is [u] than [i] for a dissimilative parsing, but the target sound may be identified as [ʃ] more often when its phonetic context is [u] than [i] for an assimilative parsing. The previous literature seems to agree that the default parsing mechanism is dissimilative. What factors cause an assimilative parsing then? More recent literature (Rysling et al., 2019) suggests that listeners exhibit an assimilative parsing over a dissimilative parsing when 1) an ambiguous target is immediately followed by its phonetic context; 2) the transition between an ambiguous target and its phonetic context is continuous; and 3) the information the preceding target is not informative enough.

The present study explores factors that may trigger an assimilative parsing over a dissimilative parsing. In particular, the present study investigates the identification of ambiguous fricative noise between [s] and [ʃ], and effect of various factors on it: syllable position (CV vs. VC), phonological category ([i] vs [u]), and whether the vowel is a canonical example of its phonological category (canonical, non-canonical). The third factor is based on the findings in my preliminary study, which suggested that non-canonical vowels as phonetic contexts may trigger an assimilative parsing (but the effect is within phonological category). However, the vowel stimuli in my preliminary study did not have formant transitions. Hence, the present study also investigates the roles of formant transitions (transitions, flat formants).

For the CV stimuli, the results show that listeners exhibit a dissimilative parsing for phonological categories ([i] vs [u]). The existence of formant transitions did not change this trend. Within vowel category, however, the listeners exhibited an assimilative parsing for [u]s with and without transitions, while the listeners exhibited a dissimilative parsing for [i]s without transitions. There was no significant phonetic boundary shift for [i]s with transitions. As for the VC stimuli, the listeners exhibited a dissimilative parsing for phonological categories ([i] vs [u]) without formant transitions. However, the listeners exhibited an assimilative parsing for phonological categories ([i] vs [u]) with formant transitions. In terms of the roles of canonical and non-canonical vowels, the influence was not significant for the VC stimuli. The results of the VC stimuli are particularly interesting as they are similar to what Fujimura et al. (1978) have reported. However, the target sounds followed their phonetic contexts ($V_{context}C_{target}$) in the present study. This does not meet one of the conditions for an assimilative parsing that Rysling et al. (2019) proposed, which states that the target sound needs to be preceding its phonetic context ($V_{target}C_{context}$) to trigger an assimilatory parsing. The conditions that cause an assimilative parsing will be further discussed.

Beyond the Lab: Acoustic Analysis of Speech Data from Smartphones and Traditional Recordings

Ke Zhang (kezhang@uwm.edu) & Hanyong Park (park27@uwm.edu)
Department of Linguistics, University of Wisconsin-Milwaukee

Traditionally, researchers often record participants' speech in a lab setting, utilizing equipment such as a computer connected to a microphone through an audio interface, to ensure high-quality recordings suitable for acoustic analysis. However, this approach has become increasingly challenging due to various factors, including constraints posed by COVID-19, and scheduling difficulties. In response, we investigated whether smartphone recordings could serve as a viable alternative for collecting speech data for acoustic analysis. We conducted a comparative analysis of common acoustic measures for vowels and fricatives obtained from two recording methods, aiming to discern potential differences in speech data between these methods.

Our study included twenty-three speakers (16 females, 7 males), encompassing diverse native language backgrounds (English = 11, Arabic = 6, Chinese = 1, Persian=1, Japanese=1, Spanish=1, Czech=1, Dutch=1). These participants were asked to read "The North Wind and the Sun" in English twice: first using their own smartphones, either iPhone or Android devices, with the lossless mobile phone application *Awesome Voice Recorder*, and second using lab recording equipment (iMac, Earthworks M30 microphone, Sound Devices USBPre 2 Audio interface) and Praat. Both recording sessions took place in a sound booth, with the same recording settings maintained for both conditions (44,100 kHz sampling rate, Mono). To evaluate whether the phone recordings could provide speech data comparable to the lab recordings for acoustic analysis, we examined several acoustic properties across the two recording methods: vowel duration, F0, F1, F2, and F3 at the midpoint of monophthongs; the same properties at one third, mid, and two thirds of diphthongs; and fricative duration, mean intensity, and the first four spectral moments (center of gravity, variance, skewness, and kurtosis) for fricatives. For vowel analysis, we employed the Online Forced Aligner, a tool based on the Penn Phonetics Lab Forced Aligner for English. For fricative analysis, we utilized the Praat script by DiCanio (2021).

Our findings indicated that the acoustic measures obtained from phone recordings closely aligned with those from lab recordings, with the exception of fricative intensity. Phone recordings consistently exhibited higher intensity levels compared to lab recordings. In light of these findings, our results suggest that utilizing participants' own smartphones for recording, particularly in a quiet environment with a lossless mobile phone application, can yield speech data suitable for acoustic analysis across the majority of acoustic measures.