The mechanism(s) of language control: Insights from the phonetics of switching and mixing

Maria Fernanda Gavino, mariagavino@u.northwestern.edu
Matthew Goldrick, matt-goldrick@northwestern.edu
Department of Linguistics, Northwestern University

Bilinguals select which language they want to use through language control mechanisms. The inhibitory control model (Green, 1998) suggests that when bilinguals are using one of their languages (e.g. Spanish) they must suppress (i.e., inhibit) representations in their other language (e.g. English). The degree of inhibition changes depending on whether bilinguals are using one language (single context) or both languages (mixed context). Mixed contexts include stay (following word is in same language) and switch (following word is in different language) contexts (Meuter and Allport, 1999). Language selection becomes more difficult when bilinguals are using both of their languages (particularly when switching), requiring greater use of inhibition. Increased language selection difficulty slow down retrieval and increase accentedness (e.g. Amengual, 2016; Balukas and Koops, 2015; Branzi, Martin, Abutalebi, and Costa, 2013; Goldrick, Runnqvist, and Costa, 2014; Kleinman and Gollan, 2018; Olson 2013). Yet, since no study has looked at both reaction times and accentedness, it is unclear if there is a single or there are multiple inhibitory mechanisms driving these effects.

To address this question, we examined reaction times, voice onset time (VOT), and vowel formants in Spanish-English bilinguals. If one control mechanism is responsible for controlling both lexical access and phonetic processing, we expect to see similar sensitivity to the difficulty of language selection across both measures. In other words, increased difficulty in language selection will result in both slower reaction times and more accented productions – i.e. more English-like VOT and vowel quality in Spanish targets and more Spanish-like VOT and vowel quality in English targets. If there are multiple mechanisms responsible for controlling both lexical access and phonetic processing, we expect to see varying sensitivities in the difficultly of language selection. For instance, increased difficulty in language selection could result in slower reaction times, but not an increase in accentedness.

We will present data from analysis currently in process. Eighteen Spanish-English bilinguals from the metro-Chicago area participated in a picture naming task in which they named 32 cognates and 16 non-cognate words in single, switch, and stay contexts in both Spanish and English. Number of participants and items were determined by a power analysis based on previous studies of VOT in Spanish switching (Goldrick et al., 2014). Target words begin with /b p d t/ and are followed by /i e/ in Spanish and /i ɪ e ɛ/ in English. VOT, vowel formants, and reaction times are currently being analyzed using machine learning systems (Shrem, Goldrick, and Keshet, 2019; Dissen and Keshet, 2016).

References

Amengual, M. (2016). Cross-linguistic influence in bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology, 7,* 1-17.

Balukas, C., & Koop, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, *19*(4), 423- 443.

Branzi, F.M., Martin, C. D., Abutalebi, J., & Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia, 52,* 102-116.

Dissen, Y., & Keshet, J. (2016). Formant estimation and tracking using deep learning. In *Proceedings of INTERSPEECH* (pp. 958-962).

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, *25*(4), 1031–1036.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67-81.

Kleinman, D. & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition, 178,* 115-132.

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics, 41*, 407–420.

Shrem, Y., Goldrick, M., & Keshet, J. (2019, in press). Dr.VOT: Measuring positive and negative voice onset time in the wild. In *Proceedings of INTERSPEECH*.

Relationship between brainstem, cortical and behavioral measures to voice onset time: preliminary findings

Jinghua Ou & Alan Yu
Department of Linguistics, University of Chicago
jou@uchicago.edu , aclyu@uchicago.edu

Consonant voicing contrasts are very common in the world's languages [1] with voice onset time (VOT) as the primary acoustic cue for voicing perception [2]. Given its significant function in nearly all languages, it is important to clarify the neural encoding of this acoustic cue. The present study systematically investigated VOT encoding at the brainstem, cortical and behavioral levels. Moreover, we aim to ascertain the interplay between brainstem and cortical levels of processing, as well as the relationship between neural encoding and behavioral measures. Previous neurophysiological studies have mainly examined neural representation of VOT using cortical auditory evoked potentials such as N1 [3,4], and a systematic account of how VOT is encoded along the auditory pathway, i.e. from brainstem to cortical, is still lacking. Unlike their cortical counterpart, auditory brainstem response (ABR) closely mimics the spectrotemporal properties of the original auditory signal [5]. By examining both brainstem and cortical encoding, we can thus delineate how the acoustic-phonetic transformation of VOT is realized within the human brain.

Six native American English speakers participated in this study (age: 23.1 ± 5.63, males = 4). A two-alternative-choice (2AFC) task was used to collect behavioural responses to a synthesized seven-step /ba/-/pa/ continuum differing in VOT. The syllables were held constant at 315 ms with VOT increasing from 14 ms to 98 ms at a step of 14 ms. Following the 2AFC task, electrophysiological (EEG) responses to step 1 (VOT = 14 ms) , step 3 (VOT = 42 ms) and step 7 (VOT = 98 ms) stimulus were collected in a passive listening paradigm, during which each stimulus was repetitively presented for 2000 times. The continuous EEG recordings were differentially bandpass filtered off-line from 1 to 50 Hz and from 80 to 2500 Hz to highlight cortical event-related potentials (ERP) and ABR respectively [6]. For ERP, peak latency and amplitude were measured for the N1 component. For ABR, peak of onset response, and onset of the sustained response were identified. The latency differences between the two onsets were computed to index VOT encoding in ABR.

Speech identification functions are shown in Fig. 1. A logistic GLM revealed that step 3 significantly modulated response (b = -.79, t = -2.40, p = .016) and RT (b = 116.1, t = 3.57, p = .0003). N1 was present for all three stimulus (Fig. 2). Results of GLM revealed that step 7 of VOT significantly modulated the N1 latency (b = 22.6, t = 2.06, p = .05). Moreover, the N1 latency was positively correlated with VOT steps (r = .47, p = .04). For ABR, the differences between the onset of burst and onset of voicing was computed. At the group level, the ABR encoding of VOT for step 1 was 21 ms, 63 ms for step 3, and 106 ms for step 7 (Fig. 3). Significant positive correspondence was found between ABR encoding and VOT steps (r = .83, p < .001). The two neural indices were also correlated (r = .33, p = .05). To ascertain the brain-behavioral relation, neural measures (N1 and ABR) were used to predict perceptual responses. Results showed that both ABR (b = -.007, t = -7.28, p < .001) and N1 latency (b = -.004, t = -1.93, p = .05) could significantly predict behavioural responses (F(3, 14) = 26.2, p < .001, Fig. 4).

To conclude, systematic changes were observed in brainstem and cortical functions in response to stimulus VOT variations. The close correspondence between neural and behavioural data suggest that neural correlates of temporal cues that emerge in early stages of processing in the brainstem may drive the cortical representation of VOT with high fidelity.

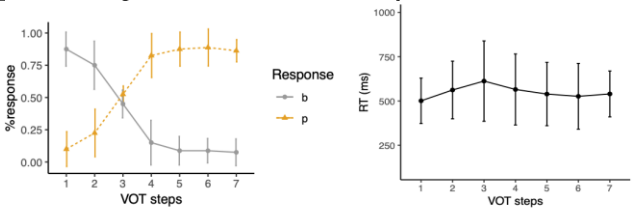


Figure 1. Psychometric identification functions (left); reaction times for speech identification (right).

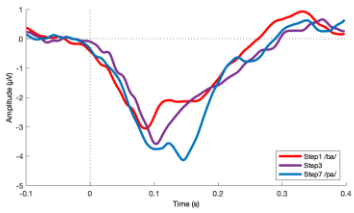


Figure 2. Waveform of N1 across VOT steps.

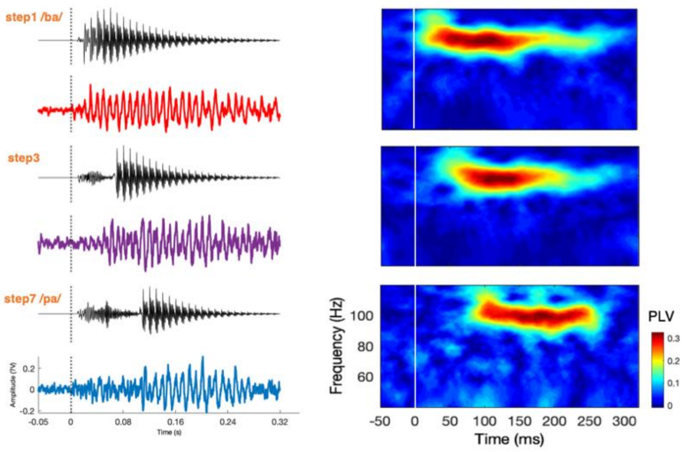


Figure 3. Waveforms of ABRs across steps (left); spectrograms of ABRs across steps (right).

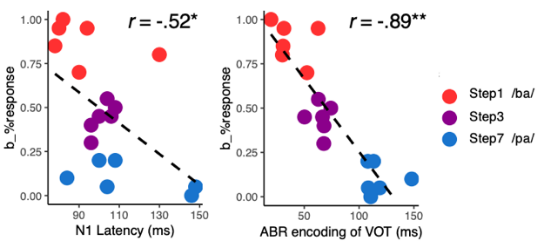


Figure 4. ERP-behavior correlation (left); ABR-behavior correlation (right).

References:

[1] Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages* (Vol. 1012). Oxford: Blackwell.
[2] Lisker, L., & Abramson, A. S. (1964). A cross- language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.
[3] Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*(2), 1078-1083.
[4] Elangovan, S., & Stuart, A. (2011). A cross-linguistic examination of cortical auditory evoked potentials for a categorical voicing contrast. *Neuroscience letters*, *490*(2), 140-144.
[5] Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and hearing*, *31*(3), 302.

[6] Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage, 79*, 201-212.

Individual differences in the success of HPVT: categorical perception and inhibition ability

Jieun Lee (jieunlee@uwm.edu) & Hanyong Park (park27@uwm.edu)
Department of Linguistics, University of Wisconsin-Milwaukee

Considering a wide range of variations in the achievement of nonnative phonological contrast learning, understanding the sources of variations is important. This study investigated the relationship between native English learners' individual differences in categorical perception and inhibition ability and their success at learning the Korean three-way stop contrast in a high phonetic variability training (HPVT) environment. The goal was to examine whether gradiency in speech perception shown in a visual analogue scaling [VAS] task (Figure 1) and inhibition ability measured by a retrieval-induced inhibition [RI] task interact with learners' performances in the process of learning. We hypothesized that learners who are less categorical (more gradient) in perception of their native (L1) categories and/or show higher inhibition ability may perform better at learning the Korean stop contrast and benefit more from training.

The different weight of f0 in the perception of Korean and English stop contrasts (primary cue in Korean, secondary cue in English) may require English native speakers' modification of their cue-weighting strategies while learning Korean stop contrasts (more active use of f0 and less use of native strategies). Here, the VAS and RI tasks were utilized to measure learners' f0 sensitivity and inhibition ability. Inhibition is related to how well learners suppress the influence of their L1 while learning a nonnative language. For the VAS task, an English /ta/ to /da/ continuum varying in VOT and f0 was used. After hearing each English stimulus, participants were asked to click anywhere on the line, based on their judgement of how close what they heard was to either /da/ or /ta/. More gradient response patterns indicated more use of f0 in their native cue-weighting strategies and the stiffness of participants' response patterns was used as numerical indication of gradiency (Kong & Edward, 2011 and 2016). For the RI task, participants memorized eighteen English words and practiced only half the items. Then, they were tested on the recognition of practiced, unpracticed, and control items, measured by reaction time (RT). The longer RTs of unpracticed items were expected by participants with high inhibition ability, since the practice process might cause inhibition of unpracticed items while activating practiced items. After these two tasks, a three-day computer based HPVT training was conducted. Training words were Korean pseudo CV syllables, /p'a/, /pa/, and /p^{h}a/. The baseline token /pa/ was recorded by three Korean native female talkers and we systemically varied both VOT and f0 to create a continuum, from /p'a/ to /p^{h}a/. Training phase started with the pre-AXB test. Every training session ended with an everyday identification (ID) test. After the last session, the post-AXB test, new talker generalization ID test, and new Korean three-way stop contrast generalization AXB test were completed.

Participants were divided into two groups based on how they performed on the VAS task (gradient vs. categorical). Overall, the gradient group outperformed the categorical group in all types of tests. In particular, the gradient group showed their ability to generalize to novel and untrained stimuli. Simple linear regression showed that individuals' quantified gradiency predicted individuals' scores on the Day 3 ID test and two generalization tests. To examine how participants changed their cue-weighting strategies in mapping training stimuli to Korean stop categories, individuals' responses for the Day 3 ID test were compared to the ones for Day 1. Only the gradient group showed more nativelike reliance on f0 cues in perception of Korean lenis and aspirated contrast for the Day 3 ID test. Thus, we preliminarily concluded that the gradient group's high f0 sensitivity in L1 category perception might help them acquire how to use f0, modify their cue-weighting strategies based on that feature, and consequently result in a more nativelike perceptual pattern. Inhibition ability failed to successfully predict learning outcomes in this study.

In sum, the results of this study showed the VAS task's predictiveness of nonnative contrast learning outcomes. Also, they showed that the reliance on multiple cues in the perception of native language may transfer and be advantageous in the course of nonnative contrast learning.

References
Kong, E., & Edwards, J. (2011, August). Individual Differences in Speech Perception: Evidence from Visual Analogue Scaling and Eye-Tracking. In *ICPhS*(pp. 1126-1129).
Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of phonetics*, *59*, 40-57.

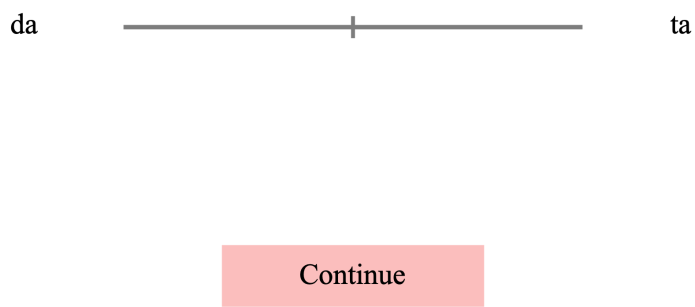


Figure 1. Illustration of visual analogue scaling (VAS) task

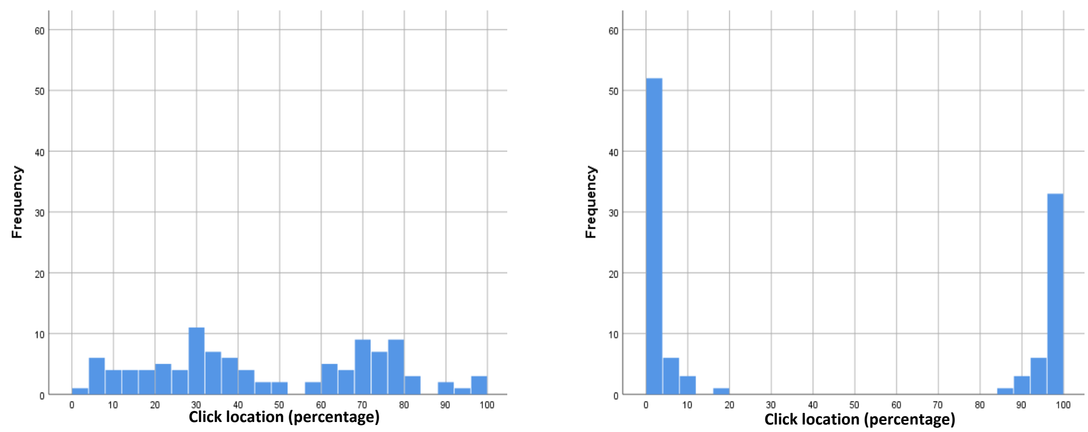


Figure 2. Examples of response patterns for the VAS task: gradient (left) and categorical (right)

Title	Speech complexity is not signal complexity
Authors	Daniel R. Turner*†, Ann R. Bradlow*, Jennifer S. Cole*
Correspondence	dturner@u.northwestern.edu†
Affiliations	Northwestern University*

Abstract

Nonspeech analogs that do not resemble speech, such as pulse trains and pure tones, are often used as points of reference for research on the psychoacoustics of speech perception, but recent findings suggest that tracking pitch in natural spoken language is more difficult than with nonspeech signals (Turner et al. 2019). Our goal was to test whether differences in pitch perception in speech and nonspeech can be traced to differences in signal complexity, or if additional cognitive processing is implicated in the processing of pitch in speech signals. To accomplish this, our participants (n=24) listened to 5 repetitions of 30 pitch excursions in nonspeech that varied by complexity (simple, complex), pitch direction (rising, falling, flat), excursion size (0—50Hz), and onset frequency (250Hz, 300Hz). Their task was to judge the pitch direction of each stimulus as rising, falling, or flat in a speeded manner. ***Stimuli.*** All stimuli were 1100ms in duration and contained a single pitch movement on a linear ramp. "Simple" signals consisted of sawtooth wave pulse trains, a common research analog for speech, while "complex" signals consisted of temporally scrambled speech (TSS). We classify TSS as nonspeech because it lacks phonetic structure, though it was derived from speech stimuli used in an earlier pitch perception study comparing speech to pulse trains. This parallel allowed us to directly compare our previous speech stimuli with TSS and pulse trains. To create TSS, we divided the natural speech recordings of phonetically controlled CVCVCV nonsense words into 100ms intervals, concatenated them in a random order, and resynthesized them using the PSOLA method. Here, we define complexity in terms of variation in the distribution of energy (SPL) across the spectrum. The simple signals (sawtooth pulse trains) are uniform in the distribution of their energy across the frequency spectrum while the complex signals (TTS and speech) show greater variance in the distribution of energy across frequencies. Asymmetries in the spectra of our simple and complex stimuli were confirmed in an analysis of the long term average spectrum of the stimuli. The analysis shows that TSS closely approximates speech, and both of these stimulus types are more complex than pulse trains (see Figure 1). For each trial we recorded the response accuracy and latency, which were analyzed using generalized linear mixed effects models. ***Results.*** While the accuracy of TSS responses approximated pulse trains for smaller pitch excursions (15—20Hz), it otherwise patterned between pulse trains and speech. This also holds true for response time: for smaller excursions TSS tracks with pulse trains, but as excursions become larger TSS response times are consistently between those of speech and pulse trains (see Figure 2). Our statistical model results echo this interpretation: for accuracy, pulse trains > TSS ($\beta = 0.710$; SE = 0.085; $p < 0.001$) and pulse trains > speech ($\beta = 1.012$; SE = 0.078; $p < 0.001$) as they interact with pitch excursion size. For reaction time, TSS > pulse trains ($\beta = 0.025$; SE = 0.007; $p < 0.001$) and speech > pulse trains ($\beta = 0.050$; SE = 0.005; $p < 0.001$) as they interact with pitch excursion size. From these results we conclude that speech-like complexity does not fully account for differences in accuracy or response time between speech and nonspeech. We suggest that the observed differences between common speech analogs might stem from additional cognitive processes that are brought online to parse and process speech. This accords with previous findings that stipulate an acoustic processing model that includes a 'speech mode'. More broadly, our results motivate an effort to revisit earlier assumptions about the informativity of mapping landmarks from the nonspeech domain to natural language.

Figures

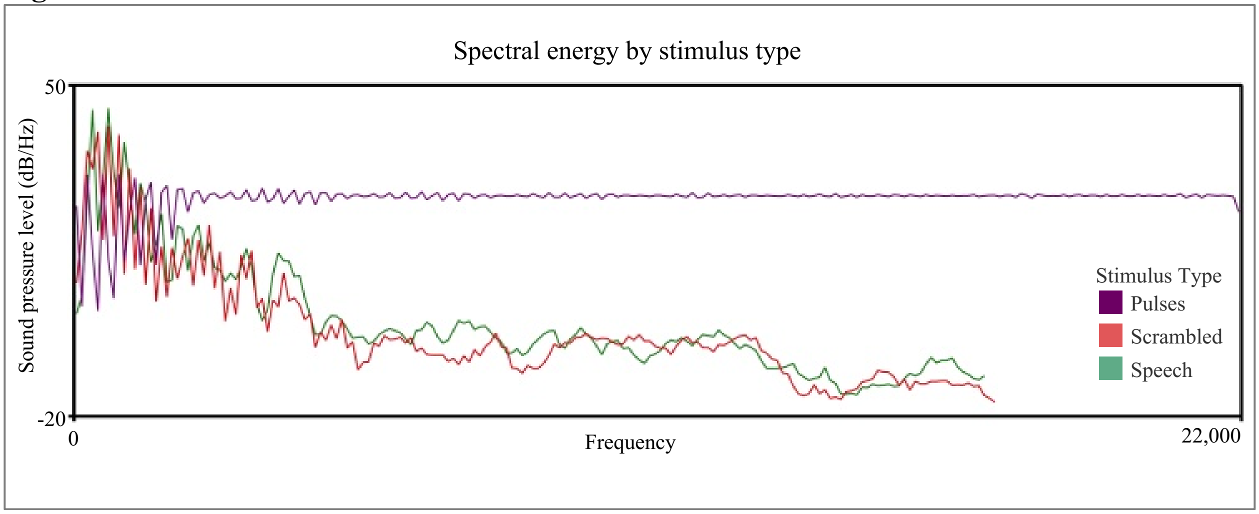


Figure 1. Spectral energy from 0—22,000Hz for our three stimulus types.

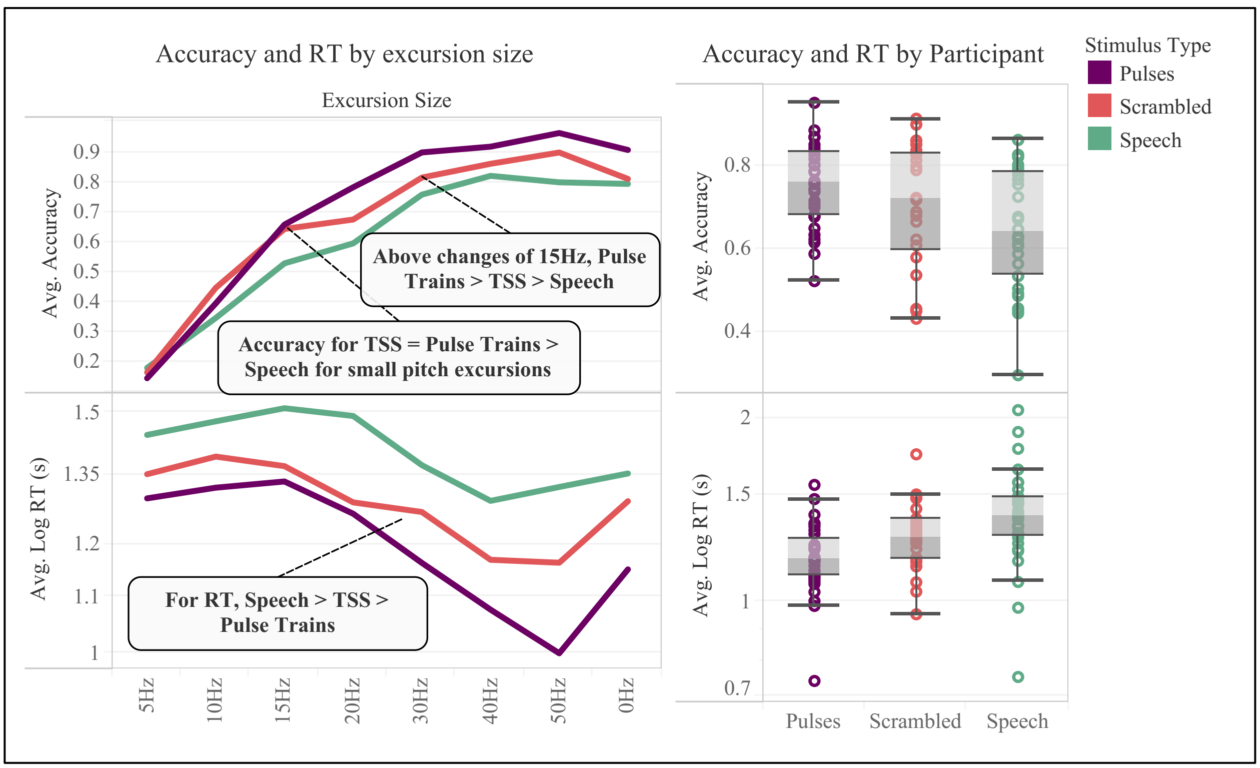


Figure 2. The accuracy and response time for TSS was more like pulse trains, especially for smaller excursion sizes. Speech stimuli elicited substantially longer RTs on average.

References

Turner, Daniel R., Ann R. Bradlow, Jennifer S. Cole. "Perception of Pitch Contours in Speech and Nonspeech", ISCA Interspeech 2019, Graz, to appear.

The Role of Phonological Neighborhood Density in Naming Images

Sabine Heuer, University of Wisconsin, Milwaukee, heuer@uwm.edu
Naomi Hashimoto, Eastern Michigan University, nhashimo@emich.edu
Anne Pycha, University of Wisconsin, Milwaukee, pycha@uwm.edu

Naming an everyday object requires both semantic and phonological knowledge (Dell & O'Seaghdha, 1992; Levelt, Roelofs, & Meyer, 1999). Upon seeing an image of a canine, for example, we must activate semantic information (*furry, common pet*) as well as a phonological code ([dɔg]) before we can produce the appropriate name, *dog.* In the literature, however, the theoretical models of these activation patterns remain underspecified, particularly with regard to the role of phonological neighbors. Neighbors are words that differ from the target by a single phoneme: for example, *net* has neighbors *pet, not, neck,* etc. (Goldinger, Luce, & Pisoni, 1989). While previous studies have examined the impact of neighbors on naming (e.g., by pairing spoken [nɛk] with image NET), their results are decidedly mixed. While some studies found facilitatory effects on reaction times (Middleton & Schwartz, 2010; Vitevitch, 2002), others found inhibitory effects (Chan & Vitevitch, 2010; Sadat et al.; 2014) or mixed effects (Gordon & Kurczek, 2014). Regardless of the findings, investigators have tended to explain their findings using the interactive spreading (IA) model, by which events at any stage of processing can affect activation on the target (Gordon & Kurczek, 2014). In the current study, which is the first of several planned to clarify the organization of phonological activation during naming, we examined the effects of neighbors on naming RTs for high- versus low-density images.

Forty-one adult English-speaking participants saw ninety-six black-and-white line drawings, and were asked to name them as quickly as possible. A spoken word (either a neighbor or an unrelated distractor word, which served as a baseline) was presented 300 ms after image onset. The key manipulation was the phonological neighborhood density (PND) of the target image, which was either low or high. For example, the word *screen* has three neighbors and its PND is therefore classified as low, while *net* has twenty-one neighbors and its PND is classified as high. Half of the images corresponded to monosyllabic low PND words (M = 9.43, SD = 4.55), and half corresponded to monosyllabic high PND words (M = 24.71, SD = 7.44). We controlled for word frequency and also for image characteristics of size, orientation, and visual complexity.

We predicted that presentation of a spoken neighbor would have a greater facilitative effect for low PND images (e.g., spoken [skɹitʃ] paired with image of SCREEN) compared to high PND images (e.g., spoken [nɛk] paired with image of NET). This prediction was based on the IA model: For an image corresponding to a high PND word such as NET, the neighbor *neck* is only one of many activated neighbors and we therefore expect its effect to be small. In contrast, for an image corresponding to a low PND word such as SCREEN, the neighbor *screech* is one of very few activated neighbors, and we therefore expect its effect to be greater.

Results are in Table 1. We conducted a 2 x2 ANOVA with RT as the outcome variable, using the factors Image PND (high vs. low) and Distractor Type (neighbor vs. unrelated). There was a significant effect of Image PND $F(1, 188) = 19.93$, $p < .001$, but no effect of Distractor Type. Thus, participants took longer to name low-density images, a finding that replicates previous studies (Middleton & Schwartz, 2010; Vitevich, 2002), but neighbors had no effect. One possibility is that, by the time neighbors were presented, naming was already complete (Damian & Martin, 1999); future work with shorter SOAs (e.g., 150ms) will eludicate this issue. A more interesting possibility, which also awaits confirmation, is that the IA model requires further refinement that may include limits on its interactivity (Levelt et al., 1999).

	Phonological neighbor	**Unrelated distractor**
High PND image	860.87 (258.2)	858.61 (268.85)
Low PND image	913.23 (254.9)	914.02 (274.91)

Table 1. Reaction times to naming images, in milliseconds.

References

Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive science, 34*, 685-697.

Damian, M.F., & Martin, R.C. (1999). Semantic codes interact in single word production. *Journal of Experimental Psychology: Language, Memory & Cognition, 25*, 345-361.

Dell, G.S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production, *Cognition, 42*, 287-314.

Goldinger, S. D., Luce, P. A., Pisoni, D.B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*, 501–518.

Gordon, J. K., & Kurczek, J. C. (2014). The ageing neighbourhood: Phonological density in naming. *Language, Cognition and Neuroscience, 29*, 326-344.

Indefrey, P. & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101-144.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-38.

Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology, 68*, 33-58.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 735-747.

Middleton, E. L., & Schwartz, M. F. (2010). Density pervades: An analysis of phonological neighbourhood density effects in aphasic speakers with different types of naming impairment. *Cognitive Neuropsychology*, *27*, 401-427.

The Gradient Acceptability in Mandarin Nonword Judgment

Shuxiao Gong gong@ku.edu
Jie Zhang zhang@ku.edu
University of Kansas

Syllable well-formedness judgment experiments reveal that speakers exhibit gradient judgment on novel words, and the gradience has been attributed to both grammatical factors and lexical statistics (e.g., Coetzee, 2008). This study investigates gradient phonotactics stemming from the violations of four types of grammatical constraints in Mandarin Chinese: 1) principled phonotactic constraints, 2) accidental phonotactic constraints, 3) allophonic restrictions, and 4) segmental-tonal cooccurrence restrictions. A syllable well-formedness judgment experiment was conducted with native Mandarin speakers to examine how the grammatical and lexical statistics factors contribute to the variation in phonotactic acceptability judgment.

Mandarin provides an excellent test case for gradient phonotactics. There are relatively clear principled phonotactic constraints in Mandarin, as proposed by Duanmu & Yi (2015) below. These constraints can all be considered as variants of the Obligatory Contour Principle, which finds wide typological support and is rooted in perception and processing (e.g., Frisch et al. 2004).

a. *HH: The vowel feature [+high] cannot occur in succession (e.g., *[m**ui**] *[t**yu**]).
b. *[Cor]_[Cor]: [Cor] cannot occur in both G and X (e.g., *[**j**a**i**] *[p**j**e**i**]).
c. *[Lab]_[Lab]: [Lab] cannot occur in both G and X (e.g., *[**w**o**u**] *[n**w**a**u**]).
d. Identical articulators cannot occur in succession in C and G (e.g., *[**tʂj**an] *[**pw**aŋ]).

Mandarin also has a rich set of allophonic variations in vowels, where variants [ə e o] belong to the mid vowel /ə/, and [a ɑ ɐ] to the low vowel /a/, as shown below.

a. ə → ə / _ n, ŋ, #   ə → e / j, ɥ _ #, or _ i   ə → o / w _ #, or _ u
b. a → a / _ i, n, #   a → ɑ / _ u, ŋ   a → ɐ / j, ɥ _

Finally, there are also segmental-tonal cooccurrence restrictions in Mandarin, as there are tonal gaps with existing syllables. For example, [nei] does not occur with the high-level Tone 1.

To construct the stimuli for the acceptability rating experiment, all theoretically possible syllables under the CGVX structure were first enumerated, using only the high-level tone. These syllables were grouped into five types: **systematic gaps** (missing syllables that violate some principled phonotactic constraint), **segmental accidental gaps** (missing syllables without any principled phonotactic violations), **allophonic gaps** (missing syllables that only violate allophonic restrictions), **tonal accidental gaps** (existing syllables that cannot bear the high-level tone), and **real words**. Forty syllables were then randomly drawn for each type to form 200 test stimuli. The 200 syllables were recorded by a native Beijing Mandarin speaker with phonetic training and normalized for intensity. Thirty native Mandarin speaker participants were asked to judge the acceptability of these auditory stimuli on a scale of 1 (bad) to 7 (good).

The data were fitted with a linear mixed-effects regression model, using z-score transformed rating scores as the dependent variable and stimulus type, neighborhood density, and stimulus duration as independent variables. Stimulus duration was included as a nuisance variable as the stimuli in different stimulus types had different durations. Step-wise likelihood ratio tests suggested that the best model is *rating* ~ *type* + *duration* + *ND* + *type:duration* + *type:ND* + (1|*item*). The ANOVA table of the model is shown in Table 1. Crucially, although neighborhood density and stimulus duration both significantly affect the rating, stimulus type not only significantly improves the model independently, but also accounts for the largest amount of variance in the rating result.

The mean z-scored ratings for each stimulus type are shown in Figure 1, which indicates that there is a sharp contrast between the acceptability of real words and non-words, but the speakers' phonotactic judgment is clearly gradient. Figure 2 shows that, overall, neighborhood density is positively correlated with ratings, but the effect in real words is weaker than in other stimulus types.

We conclude that phonotactic acceptability is gradient in Mandarin nonce syllables. Although part of the gradience can be accounted for by neighborhood density and stimulus duration, the greatest contributor to the gradience is the grammatical factor of stimulus type, with the violations of principled constraints leading to the lowest rating and the violations of tonal and allophonic gaps leading to the highest rating.

	Sum Sq.	Mean Sq.	NumDF	DenDF	F value	p value
Type	51.284	12.8211	4	185	28.5454	<.0001
Duration	6.905	6.9049	1	185	15.3733	.0001
Neighborhood Density	4.029	4.0291	1	185	8.9706	.0031
Type:Duration	4.929	1.2323	4	185	2.7437	.0299
Type:Neighborhood Density	3.455	0.8636	4	185	1.9228	.1084

Table 1 Type III analysis of variance table of the ratings model

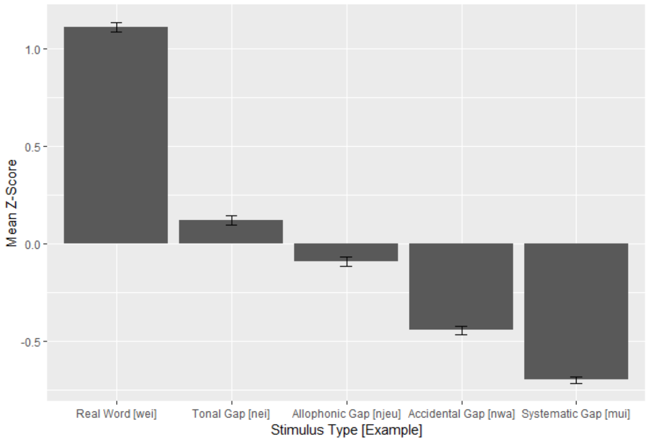


Figure 1 Mean z-scored ratings by stimulus type

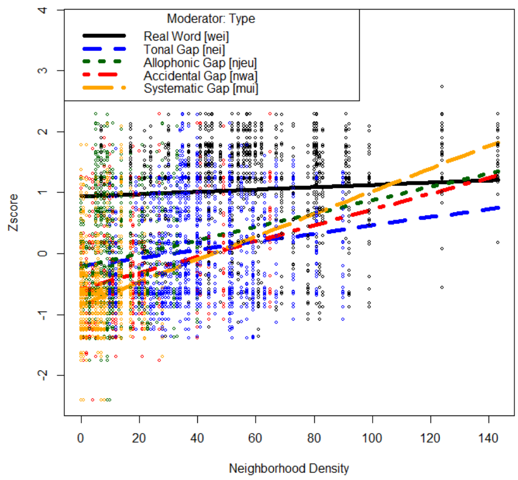


Figure 2 Effects of ND on ratings

References Grammaticality and Ungrammaticality in Phonology. Language, 84(2), pp.218–257. † Duanmu, S. (2007). The Phonology of Standard Chinese (2nd ed.). Oxford; New York: Oxford University Press. † Duanmu, S., & Yi, L. (2015). Phonemes, Features, and Syllables: Converting Onset and Rime Inventories to Consonants and Vowels. Language and Linguistics, 16(6), 819–842. † Dupoux, E., Hirose, Y., Kakehi, K., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*. † Jaeger, J. J. (1980). Testing the Psychological Reality of Phonemes. Language and Speech, 23(3), 233–253.

Ander Beristain (anderb2@illinois.edu)
University of Illinois at Urbana-Champaign

American learners' production and perception of Basque lamino-alveolar, apico-alveolar, and pre-palatal voiceless sibilants: A developmental study

Introduction

Basque is a rich language regarding the number of voiceless fricative sibilants it has, these are the lamino-alveolar <z> (/s̻/), apico-alveolar <s> (/s̺/), and pre-palatal <x> (/ʃ/) sounds. Only one sound is shared with English, i.e., /ʃ/. However, /s̻/ and /s̺/ do not appear in English. Yet, whereas the Basque lamino-alveolar phoneme <z> resembles English /s/, Basque <s> is acoustically intermediate between English /s/ and /ʃ/. This study analyzes the development of the production and perception of the three Basque fricative sibilants over a semester long course.

Theoretical framework

The Perceptual Assimilation Model (Best, 1995) or the Speech Learning Model (Flege, 1995) explain that students will map the production and perception of L2 sounds to (similar) sounds in their native languages. Afterwards, they will create new categories considering the phonetic proximity of the new sounds and those in their native language.

Research questions

1) Will L1 English native speakers be able to produce and perceive the three fricative sibilants in Basque?
2) Will a semester-long course be enough for students to progress in producing and perceiving these sounds?

Methodology

PARTICIPANTS: 11 students enrolled in the *Beginners' Basque* course at the University of Illinois at Urbana-Champaign, as well as 2 native speakers of Basque who served as control group undertook a production experiment followed by a perceptual one two times during the semester (Week 5 vs. Week 15).

PRODUCTION: a read-aloud task was conducted. Students read "I read TARGET" carrier phrases in Basque. Target words were sibilant initial followed by a corner vowel. 108 tokens were extracted per participant (3 sibilants x 3 vowels x 3 words per combination x 2 repetitions x 2 sessions). Measurements of the Center of Gravity were used to describe the places of articulation of the phonemes (Jongman et al., 2000). Voicing percentage was also analyzed.

PERCEPTION: an ABX discrimination task was undertaken. AB stimuli were produced by one female speaker whereas X was produced by a second female speaker. Students took 36 critical trials (and 72 filler trials). Critical pairs included an initial sibilant sound followed by a corner vowel. The level of accuracy in responses and Reaction Times were extracted.

Results

Linear mixed-effect models were run for COG, voicing, and RT, and a binomial logistic regression for response accuracy.

PRODUCTION: during Session 1, learners tended to merge the two alveolar sibilants into an English /s/-like sound. During Session 2, students start to map two different sound categories for Basque alveolar sibilants, with varying results (Fig. 1). Regarding voicing, the percentage of /s̻/ decreases significantly from Session 1 to 2.

PERCEPTION: accuracy scores between the 2 alveolar sibilants were significantly lower, especially when followed by the vowel /u/ (Fig. 2). No significant effect was found from Session 1 to 2. Regarding RT, learners did not show any significant effect.

We thus conclude that, in this case, perception does not precede production.

Ander Beristain (anderb2@illinois.edu)
University of Illinois at Urbana-Champaign

Figures

Figure 1

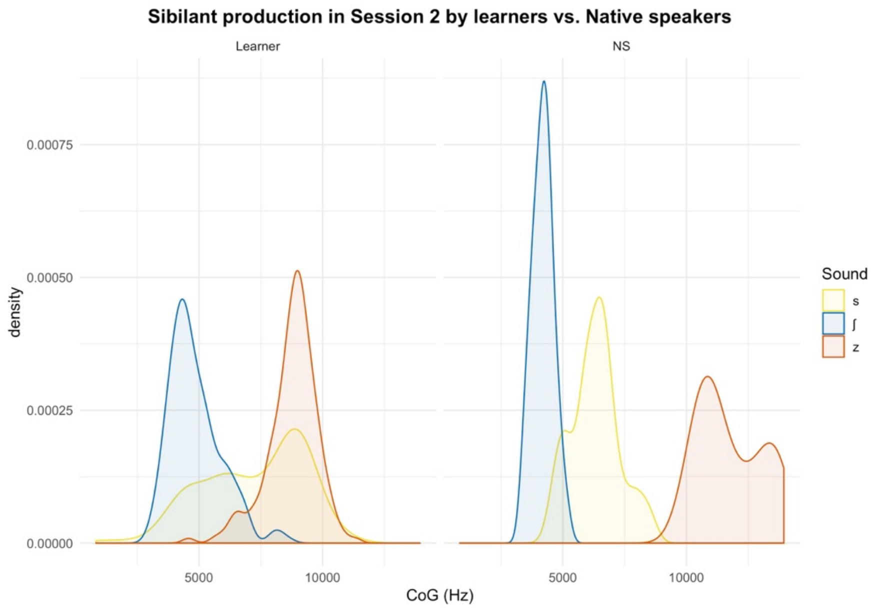
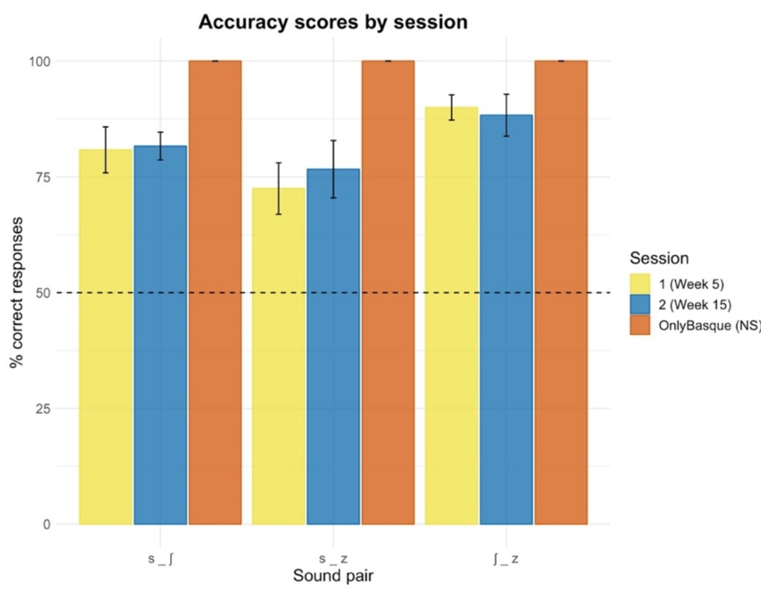


Figure 2



References

Best, C.T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research* (pp. 167-200). Baltimore: York Press.

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

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustic Society of America*, *108*(3), 1252-1263.

A comparison of the production of alveolar stop allophones in child-directed speech versus adult-directed speech

Robin Fritche, University of Wisconsin – Milwaukee
rfritche@uwm.edu
Jae Yung Song, University of Wisconsin – Milwaukee
songjy@uwm.edu

Abstract
Studies have shown that child-directed speech (or CDS) is 'clearer' speech as compared to adult-directed speech (ADS), but this has mostly been studied at the phonemic level. Therefore, there is lack of understanding about how caregivers produce non-contrastive sound variants (also known as allophones), which do not result in a change of meaning, but do affect how 'native' or 'natural' a speaker sounds. To fill in this gap, the present study compared mothers' production of three types of allophonic variants of alveolar stop /t/: flapped (as in *butter*), glottalized (as in *button*) and unreleased (as in *got*). In light of the previous findings on CDS, it was hypothesized that mothers would produce fewer variants of /t/ in CDS than in ADS. It was further hypothesized that mothers of younger children would also produce fewer variants of /t/ in CDS than mothers of older children. To test this hypothesis, ten native American English-speaking mothers of children aged 10-37 months were recorded in two story-telling sessions, one with their child and one with an adult researcher. For analysis, we used mixed effects regression models with two fixed factors, listener (ADS vs. CDS) and child age group (mothers with children aged 10-14 months vs. mothers with children aged 19-37 months), and two random effects, participants and words. Results showed that while there is very little difference between ADS and CDS in the use of the flap variant of /t/, there were significant differences for other variants. As expected, mothers released word-final /t/ (therefore not producing the 'unrelease' variant) more often in CDS than in ADS, though only in utterance-final position ($p<.001$). However, mothers of older children also released word-final /t/ more often than mothers of younger children, which was unexpected ($p<.05$). Concerning word-medial glottalized /t/, there was a significant interaction between listener and age group in both utterance-medial ($p<.01$) and utterance-final positions ($p<.05$). While both groups of mothers did less often produce glottalized /t/ in CDS than in ADS, the difference between CDS and ADS was larger for mothers of older children, which was also unexpected. These results show that mothers do indeed produce less phonetic variants in CDS, and that this effect is more robust for mothers of older children. These results suggest that around two years of age, when children's vocabularies expand rapidly, mothers produce less phonetic variants in comparison with mothers of younger children who have only just begun to speak.

Typologically rare phonetic and phonological phenomena in undocumented languages in…Indiana?

Kelly H. Berkson, Samson Lotven, Stefon Flego – Indiana University, Bloomington

In this talk, we present an overview of a developing collaboration between IU linguists and members of Indiana's Burmese refugee community. Indiana is currently home to more than 25,000 Burmese refugees, more than 19,000 of whom live in Indianapolis just an hour north of the IU Bloomington campus. Most Indianapolis community members are originally from Chin State in western Myanmar and speak under- and un-documented Tibeto-Burman languages from the Kuki-Chin branch. For the most widely spoken of these—Hakha Chin, also known as Laiholh or Hakha Lai—syntactic and morphological work exists but phonetic and phonological work is minimal. Other languages, such as Zophei and Lutuv, are completely undocumented—their names have been mentioned by previous scholars, but no linguistic work apart from that underway in Indiana exists. Yet Zophei, spoken by ~20,000 people worldwide, is spoken by ~4,000 people in Indiana, including more than a dozen students on the IU campus. Lutuv, with ~15,000 speakers worldwide, is spoken by ~700 people in Indiana and at least 2 students on the IU campus.

These realities mean that IU linguists have an invaluable opportunity to do intensive fieldwork with many speakers of typologically rich undocumented languages right here on campus or, at most, an hour's drive away. The community also has practical needs related to language and language work: like any other refugee community, they face communication challenges in both urgent (e.g. emergency rooms) and daily (e.g., navigating cell phone contracts) settings that linguists and community scholars can help address via documentation, translation, and resource creation. In this talk we briefly describe the ways in which our team of IU faculty and students, student members of the Indianapolis Chin community, and community partners from churches, social organizations, and local businesses is working to engage in scholarship <u>and</u> address practical needs *in tandem*.

After presenting a big-picture overview of our developing work, we turn to a discussion of ongoing phonetic and phonological investigations. As noted, Chin languages exhibit a number of interesting typologically rare phenomena: they generally contain laterally-released obstruents (/tl, tl^h/) and a suite of non-modal sonorants (most commonly /m̥ n̥ ŋ̊ l̥ r̥/). Hakha Chin also contains a 5-way coronal stop contrast involving dental and alveolar obstruents (/t̪ t̪^{h} d t t^h/). We review our ongoing articulatory and acoustic studies of these and other phenomena. With regards to phonology, syllable structure simplification across Kuki-Chin has yielded a synchronic situation in which individual languages are spread along a cline ranging from more conservative (e.g. complex onsets, vowel length distinctions) to more innovative (e.g. reduced onset contrasts, no coda consonants at all). We conclude our talk by discussing the ways in which data generated via fieldwork conducted in Indiana can help clarify issues in Kuki-Chin phonology raised by previous researchers.