2018 Vol.1

HISPhonCog

Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2018

HIPCS (Hanyang Institute for Phonetics and Cognitive Sciences of Language)
Department of English Language and Literature

Linguistic and cognitive functions of phonetic granularity in speech production and/or perception in L1 and L2

May 18-19, 2018, Hanyang University, Seoul, Korea

Edited by Taehong Cho, Sahyang Kim
Jiyoun Choi, Jonny Jungyun Kim
Say Young Kim, Ki-Jeong Lee

http://site.hanyang.ac.kr/web/hisphoncog 한양 음성·언어 인지과학 연구소

HISPhonCog
Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2018

CORE Initiative for College of Humanities' Research and Education
대학인문역량강화사업

HANYANG UNIVERSITY

PROGRAM AT A GLANCE

Day 1: May 18 (Friday) 2018	
08:00-08:50	Registrations (coffee & some Korean rice cake)
09:00-09:10	Opening (Young Moo Lee, President of Hanyang University)
Oral Session 1: 09:10-10:40 (Moderator: Sahyang Kim, Hongik U.)	
09:10-10:00	➢ Invited Speaker: Elizabeth Johnson (University of Toronto)
10:00-10:40	➢ Paola Escudero, Alba Tuninetti, James Whang ➢ Kristine M. Yu, Sameer ud Dowla Khan, Megha Sundara
10:40-11:00	Coffee Break
Oral Session 2: 11:00-12:20 (Moderator: Jiyoun Choi, Hanyang U.)	
11:00-11:40	➢ Invited Speaker: Aoju Chen (Utrecht University)
11:40-12:20	➢ Hui Zhang, Hongwei Ding, Wai-Sum Lee ➢ Eon-Suk Ko (Chosun U.)
12:20-13:30	Lunch
Oral Session 3: 13:30-15:10 (Moderator: Jonny Jungyun Kim, Hanyang U.)	
13:30-14:10	➢ Invited Speaker: Eva Reinisch (University of Munich)
14:10-15:10	➢ Alba Tuninetti, Natasha Tokowicz ➢ Zhen Qin, Annie Tremblay, Jie Zhang ➢ Gwanhi Yun
Poster Session 1: 15:10-16:40 (with coffee, 15 posters)	
Oral Session 4: 16:40-18:00 (Moderator: Eon-Suk Ko, Chosun U.)	
16:40-18:00	➢ Feng-fan Hsieh, Yueh-chin Chang ➢ Tomas O. Lentz, Hayo R. Terband ➢ Andrea Deme, Marton Bartok, Tekla E. Graczi, Tamas G. Csapo, Alexandra Marko ➢ Luis M. T. Jesus, Maria C. Costa
18:00-20:00	Banquet: HIT, 6th floor, Conference Venue
Day 2: May 19 (Saturday) 2018	
08:30-09:20	Registrations (coffee & some Korean Rice cake)
Oral Session 5: 09:20-10:40 (Moderator: Jongho Jun, Seoul National U.)	
09:20-10:00	➢ Invited Speaker: Holger Mitterer (University of Malta)
10:00-10:40	➢ Stephen Politzer-Ahles, Jueyao Lin, Lei Pan ➢ Sang-Im Lee-Kim
10:40-11:00	Coffee Break
Oral Session 6: 11:00-12:40 (Moderator: Minjung Son, Hannam U.)	
11:00-11:40	➢ Invited Speaker: Annie Tremblay (University of Kansas)
11:40-12:40	➢ Jeremy Steffman ➢ Sahyang Kim, Holger Mitterer, Taehong Cho ➢ Grace Kuo
12:40-13:40	Lunch
Oral Session 7: 13:40-15:00 (Moderator: Hosung Nam, Korea U.)	
13:40-14:20	➢ Invited Speaker: Mirjam Ernestus (Radboud University)
14:20-15:00	➢ Suyeon Im, Jennifer Cole ➢ Stephen Tobin, Marc Hullebus, Adamantios Gafos
Poster Session 2: 15:00-16:20 (with coffee, 15 posters)	
Final Session: 16:20-18:00 (Moderator: Taehong Cho, Haynang U.)	
16:20-17:00	➢ Invited Speaker: Natasha Warner (University of Arizona)
17:00-17:30	➢ Invited Commentator: Anne Cutler (MARCS, and ARC Centre of Excellence)
17:30-18:00	➢ General Discussion (Moderator: Taehong Cho)

Organized by
HIPCS (Hanyang Institute for Phonetics and Cognitive Sciences of Language)
Department of English Language & Literature, Hanyang University

Supported by
CORE (College of Humanities' Research and Education)
Center for the Education of Future Humanities
HIPCS (Hanyang Institute for Phonetics and Cognitive Sciences of Language)

The Organizing Committee

Chair
Taehong Cho (Director of HIPCS, Hanyang University, Seoul)

Committee members
Ki-jeong Lee (Hanyang University, Seoul)
Sahyang Kim (Hongik University, Seoul)
Say Young Kim (Hanyang University, Seoul)
Jiyoun Choi (Hanyang University, Seoul)
Jonny Jungyun Kim (Hanyang University, Seoul)

Graduate Student Staff
Jiyoung Jang (Hanyang University)
Jiyoung Lee (Hanyang University)
Jungyun Seo (Hanyang University)
Hyojin Kim (Hanyang University)
Yewon Hong (Hanyang University)
Hongmei Li (Hanyang University)

TABLE OF CONTENTS

PROGRAM IN DETAIL

Day 1: May 18 (Friday) 2018	
08:00-09:00	Registrations (coffee & some Korean rice cake)
09:00-09:10	Opening (Remark: **Young Moo Lee**, President of Hanyang University)
Oral Session 1: 09:10-10:40 (Moderator: **Sahyang Kim,** Hongik U.)	
9:10-10:00	➢ Invited Speaker: **Elizabeth Johnson (University of Toronto)** *Children's construction of a lexicon from naturally variable input* (p. 9)
10:00-10:40 (2 talks)	➢ **Paola Escudero, Alba Tuninetti, James Whang** (The MARCS Institute, ARC Centre of Excellence) *Effects of cognitive development in speech perception* (p. 25) ➢ **Kristine M. Yu, Sameer ud Dowla Khan, Megha Sundara** (U. of Massachusetts Amherst; Reed College; UCLA) *Implementing finite state intonational grammars to understand gradient prosodic manipulations in infant-directed speech* (p. 27)
10:40-11:00	Coffee Break
Oral Session 2: 11:00-12:20 (Moderator: **Jiyoun Choi,** Hanyang U.)	
11:00-11:40	➢ Invited Speaker: **Aoju Chen (Utrecht University)** *Production and comprehension of prosodic granularity: A developmental perspective* (p. 11)
11:40-12:20 (2 talks)	➢ **Hui Zhang, Hongwei Ding, Wai-Sum Lee** (Shanghai Jiao Tong U.; Shanghai Jiao Tong U.; City U. of Hong Kong) *Effects of precursor context on categorical perception of Mandarin tones in disyllabic words* (p. 29) ➢ **Eon-Suk Ko** (Chosun U.) *Mothers would rather speak clearly than spread innovation: The case of Korean VOT* (p. 31)
12:20-13:30	Lunch
Oral Session 3: 13:30-15:10 (Moderator: **Jonny Jungyun Kim,** Hanyang U.)	
13:30-14:10	➢ Invited Speaker: **Eva Reinisch (University of Munich)** *The relation between the perception and production of second language sound contrasts* (p. 13)
14:10-15:10 (3 talks)	➢ **Alba Tuninetti, Natasha Tokowicz** (The MARCS Institute, Western Sydney U., ARC Centre of Excellence; U. of Pittsburgh) *Constructing L2 phonetic categories: The influence of variability in neural responses during training* (p. 33) ➢ **Zhen Qin, Annie Tremblay, Jie Zhang** (Shanghai Jiao Tong U.; U. of Kansas) *Influence of Fine-Grained Tonal Variability on Native Chinese Listeners and Second-Language Chinese Learners' Word Recognition: An Eye-Tracking Study* (p. 35) ➢ **Gwanhi Yun** (Daegu U.) *Variation in L2 phonological compensation with phonological rules and contexts* (p. 38)

Poster Session 1 (with coffee) 15:10-16:40	
15:10-16:40	Poster (15 posters) + Coffee Break
Oral Session 4: 16:40-18:00 (Moderator: **Eon-Suk Ko,** Chosun U.)	
16:40-18:00 (4 talks)	➢ **Feng-fan Hsieh, Yueh-chin Chang** (National Tsing Hua U.) *Temporal organization of the prenuclear glides in Taiwanese Southern Min* (p. 40) ➢ **Tomas O. Lentz, Hayo R. Terband** (U. of Amsterdam; Utrecht U.) *Articulatory strategies to mark prominence in consonants* (p. 42) ➢ **Andrea Deme, Márton Bartók, Tekla E. Gráczi, Tamás G. Csapó, Alexandra Markó** (Eötvös Loránd U.; Research Institute for Linguistics HAS; Budapest U. of Technology and Economics; MTA-ELTE Lendület Lingual Articulation Research Group) *Context dependent voicing characteristics of the Hungarian /h/* (p. 44) ➢ **Luis M. T. Jesus, Maria C. Costa** (U. of Aveiro) *Aerodynamics and laryngeal features of contrast* (p. 46)
18:00-20:00	**Reception (HIT, 6th floor, Conference Venue)**

Poster Presentation (Day 1, 15:00-16:20, May 18, Friday, 2018)

(P01) **Shen Lue, Stephen Politzer-Ahles** (Hong Kong Polytechnic U.)
Analysis of the Influence of Word Frequency in Auditory Perception (p. 67)

(P02) **Janice Wing-Sze Wong, Jung-Yueh Tu** (Hong Kong Baptist U.; Shanghai Jiao Tong U.)
Native speakers' perception of Mandarin lexical tones in contemporary pop music (p. 69)

(P03) **Jung-Yueh Tu, Janice Wing-Sze Wong, Jih-Ho Cha**
(Jiao Tong U.; Hong Kong Baptist U.; National Tsing Hua U.)
Production of Trisyllabic Third Tone Sandhi in Mandarin by L1 and L2 Speakers (p. 71)

(P04) **Gayeon Son** (U. of Pennsylvania; Kwangwoon U.)
Korean-speaking toddlers' perceptual mapping based on the VOT and F0 dimensions (p. 73)

(P05) **Hyowon Kwon, Vicky Chondrogianni** (Ghent U. Global Campus; The U. of Edinburgh)
The Development of English Tense and Agreement Morphology in Welsh-English Bilingual Children with and without Specific Language Impairment (SLI) (p. 75)

(P06) **Jae-Hyun Sung** (Yonsei U.)
The Three-way Contrast of Conversational Korean Stops (p. 77)

(P07) **Seongjin Park, Natasha Warner** (U. of Arizona)
The role of within-category duration differences in speech perception (p. 79)

(P08) **Luke Horo, Priyankoo Sarmah** (Indian Institute of technology Guwahati)
Role of syllable weight on vowel acoustic space: A case in Assam Sora (p. 81)

(P09) **Daniel Williams, Paola Escudero, Adamantios Gafos**
(Western Sydney U., ARC Centre of Excellence for the Dynamics of Language; U. of Potsdam)
Australian English listeners' cue weighting of spectral change and duration in the categorization of front vowels (p. 83)

(P10) **Seulgi Shin, Annie Tremblay** (U. of Kansas)
Prosodic structure constrains the processing of denasalized nasals in Korean lexical access (p. 85)

(P11) **Seung-ah Hong** (Hankuk U. of Foreign Studies)
Bilingual speakers' perception on non-native segment contrasts: a case study on Maghrebi Arabic speakers' discriminability for Korean vowels (p. 88)

(P12) **Stephen Politzer-Ahles, Katrina Connell, Yu-Yin Hsu** (The Hong Kong Polythechnic U.) *Third-tone sandhi is incompletely neutralizing in perception as well as production: Evidence from visual world eye-tracking* (p. 90)
(P13) **James Whang** (The MARCS Institute, ARC Centre of Excellence) *Syllabification of consonant clusters by L1 Japanese L2 English speakers* (p. 92)
(P14) **Jinyoung Jo, Eon-Suk Ko** (Seoul National U.; Chosun U.) *Sound symbolism in speech directed to children by Korean mothers* (p. 94)
(P15) **Yong-cheol Lee, Dongyoung Kim, and Sunghye Cho** (Cheongju U.; Yonsei U.) *The interaction between prosodic focus and phrasal tone in South Kyungsang Korean* (p. 96)

Day 2: May 19 (Saturday) 2018	
08:30-09:20	Registrations (coffee & some Korean rice cake)
Oral Session 5: 09:20-10:40 (Moderator: **Jongho Jun,** Seoul National U.)	
9:20-10:00	➢ Invited Speaker: **Holger Mitterer (University of Malta)** *The letters of speech: evidence from perceptual learning and selective adaptation* (p. 15)
10:00-10:40 (2 talks)	➢ **Stephen Politzer-Ahles, Jueyao Lin, & Lei Pan** (The Hong Kong Polythechnic U.) *Sensitivity of the N400 to the phonetics but not the phonology of Mandarin tones* (p. 51) ➢ **Sang-Im Lee-Kim** (National Chiao Tung U.) *Contrastive context effects of tone modulated by second language experience: the case of Korean L2 learners of Mandarin Chinese* (p. 53)
10:40-11:00	**Coffee Break**
Oral Session 6: 11:00-12:40 (Moderator: **Minjung Son,** Hannam U .)	
11:00-11:40	➢ Invited Speaker: **Annie Tremblay (University of Kansas)** *The functional weight of suprasegmental cues in the native language predicts spoken word recognition in a second language* (p. 17)
11:40-12:40 (3 talks)	➢ **Jeremy Steffman** (UCLA) *Intonational structure mediates speech rate normalization in the perception of speech segments* (p. 55) ➢ **Sahyang Kim, Holger Mitterer, Taehong Cho** (Hongik U. U. of Malta; Hanyang U.) *Effects of contextual prosodic structural cues on phonological inference: a case of the post-obstruent tensing rule in Korean* (p. 57) ➢ **Grace Kuo** (National Taiwan U.) *Do working memory and autistic traits predict L2 prosody perception?* (p. 59)
12:40-13:40	**Lunch**
Oral Session 7: 13:40-15:00 (Moderator: **Hosung Nam,** Korea U.)	
13:40-14:20	➢ Invited Speaker: **Mirjam Ernestus (Radboud University)** *Does the mental lexicon contain representations for reduced word pronunciation variants? Evidence from native listeners and language learners.* (p. 19)

14:20-15:00 (2 talks)	➢ **Suyeon Im, Jennifer Cole** (U. of Illinois at Urbana-Champaign; Northwestern U.) *Exemplar encoding of intonation in imitated speech* (p. 61) ➢ **Stephen Tobin, Marc Hullebus, Adamantios Gafos** (U. Potsdam; Haskins Laboratories) *Phonetic convergence in VOT in a cue-distractor paradigm* (p. 63)
Poster Session 2: 15:00-16:20 (with coffee)	
15:00-16:20	Poster (15 posters) + Coffee Break
Final Session: 16:20-18:00 (Moderator: **Taehong Cho,** Hanyang U.)	
16:20-17:00	➢ Invited Speaker: **Natasha Warner (University of Arizona)** *Conversational speech reduction across languages, second languages, and dialects* (p. 21)
17:00-17:30	➢ **Invited Commentator: Anne Cutler** (Western Sydney University, MARCS, and ARC Centre of Excellence)
17:30-18:00	➢ **General Discussion** (Moderator: Taehong Cho, Hanyang University)
Poster Presentation (Day 2, 15:00-16:20, May 19, Saturday, 2018)	
(P01) **Hang Chan** (Hong Kong Baptist U.) *Proposing A Method for Quantifying Speech Prosody: Some Insights from a Singing Workshop* (p. 101)	
(P02) **Jihyo Kim, Eon-Suk Ko** (Chosun U.) *A cross-linguistic comparison of word teaching strategies between Korean- and English-speaking mothers* (p. 103)	
(P03) **Stephen Politzer-Ahles, Lei Pan, Jueyao Lin** (The Hong Kong Polythechnic U.) *Mandarin monosyllables trigger long-lag identity priming but not long-lag morphological priming* (p. 105)	
(P04) **Hankyul Kim** (Cornell U.) *Linguistic and Social Dimensions of Denasalization in Seoul Korean* (p. 107)	
(P05) **Chenhao Chiu** (National Taiwan U.) *The association between electroglottograph amplitude and pitch contour in Taiwan Mandarin tone production* (p. 109)	
(P06) **Jiyoung Lee, Sahyang Kim, Taehong Cho** (Hanyang U.; Hongik U.) *Prosodic strengthening effects on morpheme boundaries in Korean: a preliminary study* (p. 111)	
(P07) **Jungyun Seo, Sahyang Kim, Haruo Kubozono, Taehong Cho** (Hanyang U.; Hongik U.; National Institute for Japanese Language and Linguistics) *A preliminary study on Japanese phrase final lengthening in relation to prosodic structure: mora, lexical pitch accent and focus* (p. 113)	
(P08) **Wendy Lalhminghlui, Priyankoo Sarmah** (Indian Institute of Technology Guwahati) *Acoustic Study of Vowels in Mizo* (p. 115)	
(P09) **Eon-Suk Ko, Kyungwoon On, Jinyoung Jo, Eunsol Kim, Rana Abu-Zhaya, Amanda Seidl** (Chosun U.; Seoul National U.; Purdue U.) *Tactile cues might explain the verb-bias in Korean child-directed speech* (p. 117)	

(P10) **Bei Yang** (U. of Wisconsin, Madison)
Mandarin tone training: considering lexical access (p. 119)

(P11) **Jiyoun Choi, Jiyoung Lee, Sahyang Kim, Taehong Cho** (Hanyang U.; Hanyang U.; Hongik U.; Hanyang U.)
Prosodically-conditioned phonetic cue use in production of Korean aspirated vs. lenis stops (p. 121)

(P12) **Jiyoung Jang, Sahyang Kim, Taehong Cho** (Hanyang U.; Hongik U.)
Effects of prosodic structure on nasal coarticulation in Korean and L2 English (p. 124)

(P13) **Catalina Torres, Janet Fletcher, Gillian Wigglesworth** (The U. of Melbourne, ARC Centre of Excellence for the Dynamics of Language)
Constituent edges and final vowel lengthening in bilingual Drehu (p. 127)

(P14) **Wai-Sum Lee** (City U. of Hong Kong)
Consonant and Vowel Co-articulation in Beijing Mandarin (p. 129)

(P15) **Katalin Tamási, Quin Yow** (Singapore U. of Technology and Design)
The roles of featural distance and type of featural change in word recognition: A pupillometry study (p. 131)

Invited Talks

Children's construction of a lexicon from naturally variable input

Elizabeth K. Johnson

University of Toronto (Canada)
elizabeth.johnson@utoronto.ca

The realization of spoken words is highly variable. For example, the word *thirteen* produced by a male who learned English in Australia would sound quite distinct from the word *thirteen* produced by a female who learned English in Canada. And the production of this same word would vary further depending on a wide range of factors like the talker's mood, the age at which the talker acquired English, the relationship between the talker and the address(s), and phrasal context (e.g., *Thirteen* versus *Hearing thirteen guitars* versus *Hearing thirteen people*). The fact that adult listeners recognize words so efficiently despite all of these many sources of variation in the realization of word forms is one of the more remarkable aspects of human speech perception.

Although mature language listeners readily cope with variation in the production of spoken words, children - who are still acquiring the phonology and lexicon of their native language(s) – have been argued to struggle with many sources of word form variation. For example, 7.5-month-olds familiarized to a word form in a happy tone reportedly fail to recognize that same word in a neutral tone [1]. And 7.5-month-olds familiarized to a word form in a female voice fail to recognize that same word in a male voice [2], or in another female voice with a different accent [3]. These findings have had a strong impact on theoretical models of speech development [4].

Children's difficulty with variation in the realization of word forms is thought to continue well past early infancy. Although by 10.5 months of age infants finally succeed at cross-affect and cross-gender word recognition [1, 2], cross-accent word recognition does not emerge until after the first birthday [3]. Indeed, infants do not readily recognize familiar words in unfamiliar accents until at least 19 months of age [e.g., 5] – and some studies suggest that children continue to struggle to understand accents into early childhood [e.g., 6] and perhaps even adolescence [7].

At the same time, there is growing body evidence that children begin constructing a proto-lexicon far earlier than previously thought. For example, 6-month-olds comprehend not only socially relevant words like *mommy* and *daddy* [8], but also labels for many body parts (e.g., *hand* and *foot*) and other frequently occurring nouns (e.g., *spoon* and *ball*; [9]).

How can we reconcile claims that infants struggle to cope with variation in the realization of words with claims that 6-month-olds already comprehend a number of words, even when produced by an unfamiliar speaker? One possibility is that infants' early lexical development has been over-estimated. One could imagine that this might be particularly true for those infants who have experienced a lot of variation in their input (e.g., infants routinely exposed to multiple accents). Another possibility is that infants may be more competent at handling word form variation than past studies have suggested. Indeed, rather than hindering the ability to cope with variation, extensive exposure to variation may actually boost infants' ability to handle variation in the realization of words. In the remainder of this talk, I will address these possibilities.

How functional is a 6-month-olds' proto-lexicon? And is the speed with which children begin to add words to their lexicon affected by how variable their linguistic input is? In Study 1, we address this first question by testing 6-month-olds' ability to recognize words in an eyetracking paradigm. Our results provide partial support for the notion that 6-month-olds understand some commonly heard words [10]. In Study 2, we address the latter question by examining mono- and multi-accent 6-month-olds' ability to distinguish between their own name and a mispronunciation of their own name (e.g., *Zoey* versus *Doey*). Surprisingly, mono-accented infants listened longer to the correct than incorrect pronunciations of their own name, but multi-accented infants did not. Does this mean that vocabulary development is delayed in multi-accent infants? In Study 3, we address this question in a large-scale study comparing vocabulary growth in mono- versus multi-accent 12- to 33-month-olds (N=1133). We found no difference in vocabulary size growth between the two groups of children. Moreover, we also find no difference between multi-accent children exposed to regional accents (e.g., Australian English) versus multi-accent children exposed to non-native accents (e.g., Spanish-accented English). We conclude that multi-accent input affects how children attend to (or process) speech, but does not affect lexical development.

Study 3 is consistent with the notion that infants are better able to cope with accent variation than past studies have suggested. Can we find additional evidence for this possibility? In Studies 4 and 5, we use a word form recognition task to examine young infants' ability to handle talker-related variation [11, 12]. By making some small changes to the test paradigms used in the past, we show that infants as young as 6 months of age succeed at cross-gender word form recognition. In Study 6, further show that infants as young as 15 months can recognize words produced in unfamiliar accents [13]. Thus, all in all, we find substantial support for the notion that infants are relatively adept at handling talker- and accent-related variation in the realization of words.

In the first part of this talk, I showed that multi-accent infants show no preference for correct over incorrect pronunciations of their own name. But at the same time, vocabulary growth proceeds identically in multi- and mono-accent children, and there is substantial evidence that infants are more adept at handling talker- and accent-related variation than previous research has suggested. How can we reconcile these findings? Was Study 2 a fluke, or can we find additional evidence that multi-accent infants attend to and/or process speech differently than mono-accent infants? In Studies 7 to 9, we show that multi-accent children recognize Canadian accented words more slowly than mono-accent children do [e.g., 14]. However, we find no evidence that multi- and mono-accent children differ in how they process words produced in an unfamiliar accent.

I will end by discussing key questions left unanswered by the research I have presented, and outline some potential plans for building further upon this initial set of findings.

References

[1] Singh, L., Morgan, J.L., White, K.S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, 51, 173-189.

[2] Houston, D.M., & Jusczyk, P.W. (2000). The role of talker-specific information in word segmentation by infants. Journal of Experimental Psychology : *Human Perception and Performance*, 26, 1570-1582.

[3] Schmale, R., & Seidl, A. (2009). Accomodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science, 12*, 583-601.

[4] Johnson, E.K. (2016). Constructing a proto-lexicon : an integrative view of languge development. *Annual Review of Linguistics*, 2, 391-412.

[5] Best, C.T., Tyler, M.D., Gooding, T.N., Orlando, C.B., & Quann, C.A. (2009). Development of phonological constancy : toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, 20, 539-532.

[6] Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents : a preliminary investigation. *Journal of Child Language*, 25, 343-365.

[7] Bent, T. (2018). Development of unfamiliar accent comprehension continues through adolescence. *Journal of Child Language*.

[8] Tincoff, R., & Jusczyk, P.W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10, 172-175.

[9] Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common knowns. *Proceedings of the National Academy of Sciences*, 9, 3253-3258.

[10] Sohail, J., & Johnson, E.K. (2013). The emergence of comprehension in infancy : a longitudinal study. Poster presented at the 54th Annual Meeting of the Psychonomic Society, Toronto, CA.

[11] Van Heugten, M., & Johnson, E.K. (2012). Infants exposed to fluent natural speech succeed at cross-gender word recognition. *Journal of Speech, Language, and Hearing Research*, 55, 554-560.

[12] Johnson, E.K., Seidl, A., & Tyler, M.D. (2014). The edge factor in early word segmentation : utterance-level prosody enables word form extraction by 6-month-olds. *PLOS ONE*, 9, e83546.

[13] Van Heugten, M., & Johnson, E.K. (2014). Learning to contend with accents in infancy : benefits of brief speaker exposure. *Journal of Experimental Psychology : General*, 143, 340-350.

[14] Buckler, H., Siddiqui, N., Orzak-Arsic, S., & Johnson, E.K. (2017). Input matters : speed of word recognition in in toddlers exposed to multiple accents. *Journal of Experimental Child Psychology, 164*, 87-100.

Production and comprehension of prosodic granularity: A developmental perspective

Aoju Chen

Utrecht University (the Netherlands)
aoju.chen@uu.nl

Prosody is used at both the phonological and phonetic level in communication. In line with the autosegmental-metrical framework [1, 2, 3], phonological use of prosody is defined as making perceptually discrete variation in prosody, such as accent placement, choice of accent type, e.g., a falling pitch accent vs. a rising accent, and phrasing – inserting a phrasal boundary or not before and after a word. Phonetic use of prosody is defined as phonetic implementation of a phonological category, e.g. changes in pitch span of a pitch accent or a lexical tone and lengthening at a phrasal boundary. A case in point is prosodic focus marking in languages like Dutch. Focus refers to the predication on a topic in a sentence and typically contains new information to the hearer [4, 5]. In Dutch, focus is typically realised with a falling accent (H*L) whereas post-focus (i.e. constituents following focus) is typically deaccented. Pre-focus (i.e. constituents preceding focus) is usually realised phonologically in a similar way to focus but is phonetically different from focus. [6] showed that H*L-accented subject nouns in SVO sentences are realised with a larger pitch span, a lower pitch minimum after the peak, an earlier peak alignment, and a longer word duration in focus than in pre-focus. [7] reported that adults' comprehension is slowed down when the accent in the subject noun is not realised with more acoustic prominence in the subject-focus condition, similar to when the object noun is not accented in the object-focus condition. These findings are clear manifestations of prosodic granularity in adults' speech production and comprehension.

In this talk, I will discuss the production and comprehension of prosodic granularity in children. In the light of data on monolingual Dutch-speaking children's prosodic realisation of focus (Study 2) and processing of the focus-to-prosody mapping (Study 1), I suggest that children acquire phonetic use of prosody later than phonological use of prosody in both production and comprehension and that phonetic use of prosody occurs earlier in production than in comprehension.

Study 1: Naturally produced SVO declarative sentences were elicited from four to eight year olds (N = 23) in an interactive setting via a picture-matching game (Example 1). Each subject noun and object noun occurred in both the focus condition and the non-focus condition. We analysed the prosody in the subject and object nouns regarding accentuation, type of accent and the prosody of the subject nouns regarding phonetic implementation of H*L. Our statistical analyses showed that the children were adult-like in the use of accentuation at the age of four or five and in choice of accent type at the age seven or eight but could vary the phonetic realisation of H*L only in pitch scaling in the subject nouns at the age of seven and eight [6, 8]. In a subsequent perception test, we presented adult native speakers of Dutch (N = 15) with pairs of subject noun phrases selected from the children's answer sentences (e.g. The cleaner) and asked them to determine for each pair which sound was the beginning of the answer to a who-question. The adult listeners' perception was at chance level in the pairs selected from the four- to five-year-olds but was accurate in over 65% of the cases in the pairs form the seven- to eight-year-olds and over 75% of the cases in the pairs from adult speakers. These results are compatible with the success in which children and adults use phonetic implementation of H*L in subject nouns.

Study 2: Children aged four to seven (N = 71) listened to short question-answer dialogues between a boy and his pets on a set of pictures that were presented on the computer screen (Example 2). Their task was to judge for each dialogue whether the pet gave a correct answer by pressing the corresponding button of the button box. The focus was on the subject noun in half of the experimental trials and on the object noun in the other half of the experimental trials. The prosody was contextually appropriate in half of the dialogues and contextually inappropriate in the other half of the dialogues in each focus condition (subject-focus, object-focus). The answer sentences were segmentally and lexically correct in all experimental trails. Multivariate modelling on log-transformed reaction times on the experimental trials in which the answers were judged to be correct showed that children in this age range responded faster in the inappropriate prosody condition than in the appropriate prosody condition when the subject was the focus, unlike adults, but they responded slower in the inappropriate condition than in the appropriate condition when the object was the focus, like adults.

(Example 1)

Subject-focus:

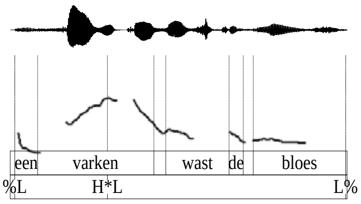
Experimenter: Look! A beet. It seems that someone is eating the beet. Who is eating the beet?
Participant: The cleaner is eating the beet.

Object-focus:

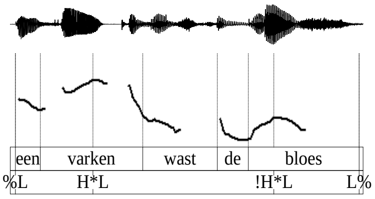
Experimenter: Look! A cleaner. It seems that she is picking up something. What is the cleaner picking up?
Participant: The cleaner is picking up a vase.

(Example 2)

Who is washing the blouse?

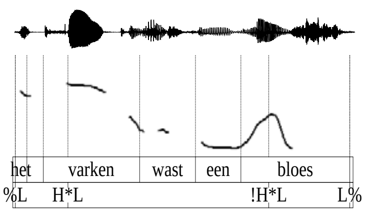


Initial focus-appropriate

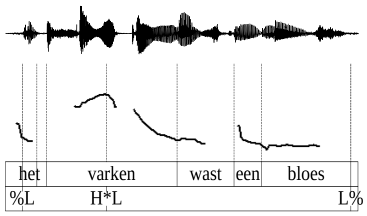


Initial focus-inappropriate

What is the pig washing?



Final focus-appropriate



Final focus-inappropriate

References

[1] Pierrehumbert, J. B. (1980). The phonology and phonetics of English intonation. PhD dissertation, MIT. New York: Garland Press.

[2] Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: CUP.

[3] Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: CUP.

[4] Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the representations of discourse referents.* Cambridge: CUP.

[5] Vallduví, E. & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, 34(3), 459-520.

[6] Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. Vigário, S. Frota, & M. J. Freitas (Eds.), *Phonetics and Phonology: Interactions and interrelations* (pp. 91-106). Amsterdam: John Benjamins.

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

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

The relation between the perception and production of second language sound contrasts

Eva Reinisch[1]

[1]*Ludwig-Maximilians University of Munich (Germany)*
evarei@phonetik.uni-muenchen.de

Learning to perceive and produce the sounds of a second language (L2) is not always easy, especially later in life and in an environment in which the L2 is not frequently used. Learning an L2 entails the establishment of phonetic categories and contrasts that are not part of the native language (L1) sound inventory. According to models of L2 sound learning [1,2] learners perceive L2 sounds by mapping them onto their L1. Importantly, difficulties in the perception of an L2 contrast are typically mirrored in production. If learners fail to produce two L2 sounds as sufficiently distinct they are hard to understand and are perceived as speaking with a foreign accent [3]. However, the characterization of the relation between L2 perception and production has been up for debate. Specifically, it is unclear whether well-differentiated L2 categories in perception are a prerequisite for well-differentiated categories in production. In my talk I will present two studies that address the L2 perception-production relationship from two different perspectives, using two different L1-L2 language pairs and L2 sound contrasts: the English vowel contrast /ɛ/-/æ/ for native speakers of German and the German contrast /ʔ/-/h/ for native speakers of Italian.

The English vowel contrast /ɛ/-/æ/ has been shown to be challenging for learners of German because the two sounds tend to be mapped onto a single native category /ɛ/ (e.g., [4,5]). In our study [6] we asked whether there is a relation between the magnitude of acoustic differences that German learners produce to distinguish between word pairs differing in this vowel contrast and their ability to use these cues in perception. Specifically, we tested how learners recognize the intended words of the minimal pairs as produced by other learners. This allowed us to assess the perception of a range of vowel productions from poorly to well differentiated and ask whether good producers would be better perceivers even if only weak acoustic cues were available (i.e., in the poorly differentiated productions). Overall, learners who produced large acoustic differences between the vowels were better at identifying minimal pairs produced by other learners than those learners who themselves produced small acoustic differences. Interestingly, this was the case when identifying good as well as poor productions, but with a larger benefit for the better productions. For a sound contrast in which two L2 sounds are mapped onto a single L1 sound it hence appears that better production is indeed related to better abilities in using available acoustic cues to vowel identification. However, the direction of causality and how learners are able to improve remains for discussion.

The second study concerns a type of L2 contrast in which neither of the sounds has a clear counterpart in the learners' L1. This type of contrast has received relatively less attention than the one discussed above. The case in point here is the German contrast /ʔ/-/h/ for native speakers of Italian. In most dialects of Italian these sounds only occur as phrase markers or in paralinguistic function. Importantly, anecdotal evidence suggests that Italian learners of German tend to drop /h/ in production in words that actually are /h/-initial; but at the same time, they produce /h/ in words that in German canonically start in a glottal stop (i.e., words that are orthographically vowel-initial). Studying learners' production and perception of these sounds is of interest since, firstly, /h/ is a well-known problem for Italian learners whereas the existence of /ʔ/ is not explicitly known even to many native speakers of German. Secondly, the presence of /h/ is indicated by a grapheme while that of German /ʔ/ is not. We can hence test whether learners' awareness of difficulties with a given L2 sound impacts learning.

In a production study we elicited /ʔ/- and /h/-initial words by asking learners to construct sentences from pictures. Target words were always preceded by words ending in a vowel or nasal to facilitate the decision whether a glottal stop or glottalization had been produced. We found that learners produced both /h/ and /ʔ/ correctly about 70% of the time with errors consisting of both deletions and substitutions in both directions.

Two experiments, each consisting of a visual-world eye-tracking task and a goodness judgment task were then conducted to test the perceptual representation of these sounds. One experiment tested reactions to sound substitutions and the other one to sound deletions. A previous study had shown that native speakers of German recognize words more slowly if either of the two sounds is deleted and only show stronger penalties for the deletion of the orthographically represented /h/ in an explicit goodness rating task but not in an implicit word recognition task [7]. Therefore, the combination of implicit and explicit perception tasks allowed us to test possible effects of learners' awareness of the difficulty for /h/ but not for /ʔ/. Analyses of Italian learners' target fixations in the eye-tracking task revealed that word recognition in the correct vs. substituted condition was not different for either target sound. However, in the correct vs. deleted condition learners fixated on the target less if the initial sound was deleted, but this effect was subtle. The deletion effect was also numerically larger in size for /h/ than for /ʔ/. Assuming that eye-tracking reveals the degree of match between acoustic input and lexical representations that are being accessed [8], these results suggest that the Italian learners have established a fuzzy representation for a glottal sound for (otherwise) vowel-initial words (i.e., effect of deletion), which might be stronger for /h/-initial words. However, the nature of the sound as /ʔ/ vs. /h/ is not well defined (i.e., no effect of substitution). In contrast, results from the explicit goodness rating task showed effects of substitutions and deletions for both, /ʔ/ and /h/, suggesting that learners can acoustically differentiate between the two sounds, presumably because they are known as paralinguistic sounds in the L1. In the explicit task, the deletion of /h/ was rated worse than the deletion of /ʔ/, mirroring the Germans' results [7]: the awareness of difficulties with an L2 sound mainly appears to influence explicit perception.

Relating the results from the perception tasks back to production we find a dissociation. While 70% correct production is far from ceiling, it is more than one might expect given the perception data from the eye-tracking task. These suggest that the actual representation of the two sounds may not be fine-grained enough for a differentiation between the two sounds in online processing.

Taken together the two studies suggest that the perception and production of L2 sound contrasts may be related such that better producers are those who are better at detecting and using the relevant acoustic cues in perception. However, if neither of the sounds of an L2 contrast is found in the learners' L1 sound inventory, then good production abilities may even precede perception. Issues of task differences and attention during processing will be discussed.

References

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

[2] 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.

[3] Eger, N. A., & Reinisch, E. (in press). The role of acoustic cues and listener proficiency in the perception of accent in non-native sounds. *Studies in Second Language Acquisition, online-first view*.

[4] Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

[5] Llompart & Reinisch, (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 1040- 1056.

[6] Eger, N. A., & Reinisch, E. (in press). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

[7] Mitterer, H. & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language, 85,* 116-134.

[8] Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

The letters of speech: evidence from perceptual learning and selective adaptation

Holger Mitterer[1]

[1] *University of Malta ((Malta)*
Holger.mitterer@um.edu.mt

While every model of visual-word recognition for alphabetic scripts assumes that letters play an important role in mediating between the sensory input and lexical representations, no such clear consensus exists for spoken-word recognition. In this talk, I will provide an overview of recent developments in this unit-of-perception debate. Partly based on results from a perceptual learning paradigm [1], there is at least a consensus that some form of intermediate unit is involved. This paradigm makes use of a two-part procedure. In the first part of an experiment, participants are exposed to ambiguous phones (e.g., a fricative that could be interpreted as either /s/ or /f/) in disambiguating context. Such a context could, for instance, be "platypu[$^{s}/_{f}$]" where ambiguous [$^{s}/_{f}$] can be interpreted as /s/ because platypus is a word but platypuf is not. This exposure has repercussions for other words, even if the fricative now occurs in an ambiguous position (e.g., [lai$^{s}/_{f}$], where it could be interpreted as /s/ → *lice* or /f/ → *life*). This form of generalization shows that listeners make use of sublexical units in speech perception [2]

However, the form of this unit is still under debate, with phonological features [3], articulatory features [4], allophones [5], and phonemes [6] as most prominent contenders. Mitterer et al. [5] had argued that the same perceptual learning paradigm that led to the consensus that some form of intermediate unit is used [1] can also be used to delineate the form of these units based on patterns of generalization or non-generalization of learning. They showed that learning about the /l/-/r/ contrast in Dutch, which displays a lot of allophonic variability, is restricted to the allophones used during exposure. This finding is in line with the assumption that the units are either allophones or articulatory gestures—since the articulatory gestures differ between allophones—but is problematic for the assumption of phonological features or abstract phonemes.

Further work with this paradigm investigated the potential role of articulatory features. An articulatory-feature account predicts that learning about a given articulatory gesture—such as a labial closure—should generalize over manner of articulation. That is, when one learns that a given speaker does not produce a clear /b/, learning should generalize to other segments that require the same articulatory gesture (such as /p/ and /m/). This prediction was, however, not confirmed [7]–[9]. These studies also showed that phonetic form but not phonological specification seem to be pivotal for recalibration. When the phonetically similar surface forms with different underlying specifications are used (e.g., phonetically unvoiced stops in word-final position arising from underlying unvoiced or underlying voiced stops that undergo devoicing), only recalibration for voiceless stops is found, and participants learn about phonetically unvoiced stops just as well from devoiced as from underlyingly voiceless stops.

This line of research was then questioned by Bowers, Kazanina, and Andermanse [6] on two grounds. First, they argued that the finding of spatial selectivity in perceptual learning [10] undermines that the perceptual learning paradigm reveals units of perception. Secondly, they used the selective-adaptation paradigm to show that selective adaptation generalizes across position for stops, questioning the role of position-specific allophones.

The first argument is based on evidence in which exposure and test trials are separated by only a few seconds. I will present data that no spatial selectivity is observed if exposure and test trials are separated by more than one minute. This hence indicates that data from paradigms with repeated short sequences of exposure and test trials (as in [9]) have to be treated with caution.

The second argument by Bowers et al. [6] was countered by Mitterer, Reinisch, and McQueen [11] who showed that selective adaptation does not reliably occur even if adaptor and test stimuli share the same phoneme but are different allophones. They made use of the same /r/-/l/ allophony in Dutch as [5] and additionally used the German back fricatives [ç] and [x], which made it possible to manipulate phonemic and allophonic overlap independently. This was possible because the fricative not only has two allophones (velar and palatal fricative) but [ç] also arises as a possible pronunciation of word final /g/ in some variants of German.

In summary, these lines of research suggest that the allophone is an important unit in mediating between input and lexical representations. However, it may not be the only relevant unit, because some evidence indicates that listeners can learn about units larger than one segment, (e.g., syllables [12]) or parts of phones, such as release bursts, which may occur in different allophones [8]. This suggests that the units of speech perception may be driven by what type of patterns reliably occur in the input.

References

[1] D. Norris, J. M. McQueen, and A. Cutler, "Perceptual learning in speech," *Cognit. Psychol.*, vol. 47, pp. 204–238, 2003.

[2] A. Cutler, F. Eisner, J. M. McQueen, and D. Norris, "How abstract phonemic categories are necessary for coping with speaker-related variation," in *Laboratory phonology 10*, C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée, Eds. Berlin: de Gruyter, 2010, pp. 91–111.

[3] A. Lahiri and H. Reetz, "Underspecified recognition," in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds. Berlin: Mouton de Gruyter, 2002, pp. 637–676.

[4] C. A. Fowler, "The reality of phonological forms: a reply to Port.," *Lang. Sci. Oxf. Engl.*, vol. 32, no. 1, pp. 56–59, 2010.

[5] H. Mitterer, O. Scharenborg, and J. M. McQueen, "Phonological abstraction without phonemes in speech perception," *Cognition*, vol. 129, pp. 356–361, 2013.

[6] J. S. Bowers, N. Kazanina, and N. Andermane, "Spoken word identification involves accessing position invariant phoneme representations," *J. Mem. Lang.*, vol. 87, pp. 71–83, Apr. 2016.

[7] H. Mitterer and E. Reinisch, "Surface forms trump underlying representations in functional generalisations in speech perception: the case of German devoiced stops," *Lang. Cogn. Neurosci.*, vol. 32, no. 9, pp. 1133–1147, Oct. 2017.

[8] H. Mitterer, T. Cho, and S. Kim, "What are the letters of speech? Testing the role of phonological specification and phonetic similarity in perceptual learning," *J. Phon.*, vol. 56, pp. 110–123, May 2016.

[9] E. Reinisch and H. Mitterer, "Exposure modality, input variability and the categories of perceptual recalibration," *J. Phon.*, vol. 55, pp. 96–108, Mar. 2016.

[10] M. Keetels, J. J. Stekelenburg, and J. Vroomen, "A spatial gradient in phonetic recalibration by lipread speech," *J. Phon.*, vol. 56, pp. 124–130, May 2016.

[11] H. Mitterer, E. Reinisch, and J. M. McQueen, "Allophones, not phonemes in spoken-word recognition," *J. Mem. Lang.*, vol. 98, pp. 77–92, Feb. 2018.

[12] K. Poellmann, H. R. Bosker, J. M. McQueen, and H. Mitterer, "Perceptual adaptation to segmental and syllabic reductions in continuous spoken Dutch," *J. Phon.*, vol. 46, pp. 101–127, Sep. 2014.

The Functional Weight of Suprasegmental Cues in the Native Language Predicts Spoken Word Recognition in a Second Language

Annie Tremblay

University of Kansas
atrembla@ku.edu

Traditional approaches to the study of phenomena that are predominantly suprasegmental (e.g., lexical stress, sentence-level prosody) in non-native speech perception and spoken word recognition have assumed that the more similar the first- (i.e., native-) language (L1) and second-language (L2) systems, the easier it is for learners to perceive suprasegmental information and use it in word recognition. For example, listeners whose L1 does not have lexical stress (e.g., [Metropolitan] French, [Seoul] Korean) have been predicted (and found) to have difficulty perceiving stress and using it in word recognition [1,2,3,4,5,6].

In this presentation, I will argue that it is not the similarities between the L1 and the L2 at the level of the phonological system that yield successful use of suprasegmental information in non-native speech perception and word recognition, but rather *the weight of particular suprasegmental (and segmental) cues for signaling lexical identity in the L1*. I will present the results of two experiments, one on the processing of lexical stress in English (Exp. 1) and one on the processing of sentence-level prosody in French (Exp. 2), that provide empirical support for this proposal.

English and Dutch both have lexical stress but differ in how stress is realized, with unstressed vowels having a more centralized place of articulation (i.e., being more reduced) in English than in Dutch [7]. Because both segmental (i.e., vowel reduction) and suprasegmental (i.e., fundamental frequency [F0], duration, intensity) cues signal lexical stress in English, the relative weight of suprasegmental cues for signaling lexical identity is weaker in English compared to Dutch. Like English and Dutch, (Beijing) Mandarin has lexical stress in some words [8,9]. More importantly, lexical identity in Mandarin depends on the lexical tone that the word carries [8,9]. The functional weight of suprasegmental cues to lexical identity is thus greater in Mandarin than in English or Dutch. In contrast to English, Dutch, and Mandarin, neither French nor Korean has lexical stress (or lexical tones), but intonational cues can signal word boundaries in both languages [10,11]. Lexical identity in French and Korean thus depends much less on suprasegmental cues compared to Mandarin, English, and Dutch.

If the weight of cues for signaling lexical identity in the L1 predicts L2 learners' successful use of suprasegmental information in word recognition, Mandarin L2 learners of English listeners should make greater use of suprasegmental cues to English stress compared Korean L2 learners of English, but they would not necessarily make greater use of segmental cues to English stress (since full and reduced English vowels can be assimilated to different vowels in both Korean and Chinese). Similarly, Dutch L2 learners of French should make greater use of suprasegmental cues to word-final boundaries in French compared to English L2 learners of French.

In Experiment 1 [12], Mandarin and Korean L2 learners of English (matched in their English proficiency and experience) and native English listeners completed a visual-world eye-tracking experiment with printed words adapted from Cooper et al. [13]. In each trial, participants heard an auditory word in the carrier sentence *Click on* _____ and saw four orthographic words on the screen, including the target word and a competitor word. The first variable was whether the target and competitor words differed in their stress pattern. In the stress-mismatch (experimental) condition, the target and competitor words overlapped in their first syllable but differed in their stress pattern (e.g., *CARpet* as target and *carTOON* as competitor; *PArrot* as target and *paRADE* as competitor); in the stress-match (control) condition, the target and competitor words overlapped in their first syllable and had the same stress pattern (e.g., *CARpet* and *CARton; PArrot* and *PArish*). The second variable was whether, in the stress-mismatch condition, the first syllable of each the target and competitor words had the same vowel. In the condition without vowel reduction, the vowel in the target and competitor words was the same (e.g., *CARpet* as target and *carTOON* as competitor); in the condition with vowel reduction, the vowels in the target and competitor words differed, with the target containing a full vowel and the competitor containing a reduced vowel (e.g., *PArrot* as target and *paRADE* as competitor). The results showed that English listeners used both suprasegmental cues alone and segmental and suprasegmental cues together to recognize English words, with the effect of stress being greater for combined cues. Additionally, as predicted, Mandarin L2 learners of English outperformed Korean L2 learners of English in the use of suprasegmental cues alone but not in the use of segmental and suprasegmental cues combined. Importantly, Korean L2 learners of English were able to use suprasegmental cues to English stress only when these cues co-occurred with segmental cues, suggesting that this segmental information gave them a means to process English stress.

In Experiment 2 [14], English and Dutch L2 learners of French (matched in their French proficiency and experience) and native French listeners completed a visual-world eye-tracking experiment with printed words adapted from Christophe et al. [15] and Tremblay et al. [16]. They heard sentences containing a monosyllabic noun and a multisyllabic adjective, where the noun and the first syllable of the adjective could temporarily be segmented as a disyllabic word (e.g., *chat lépreux* 'leprous cat,' which could be segmented as *chalet* 'cabin' before *–preux* is heard). The visual display contained the monosyllabic target noun and the disyllabic competitor word, together with two distracter words. The results showed that, as predicted, Dutch L2 learners of French made greater use of the F0 rise to locate word-final boundaries compared to English L2 learners of French, and they even made greater use of this F0 rise compared to French listeners.

These results provide empirical support for the proposal that the weight of suprasegmental (and segmental) cues for signaling lexical identity in the L1 predicts the use of the corresponding information in L2 spoken word recognition.

References

[1] Peperkamp, S., & Dupoux, E. (2002). A typological study of stress deafness. In C. Gussenhoven (Ed.), *Proceedings of Laboratory Phonology 7* (pp. 203–240). Berlin: Mouton de Gruyter.

[2] Lin, C. Y., Wang, M. I. N., Idsardi, W. J., & Xu, Y. I. (2014). Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition, 17*, 316-346.

[3] Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics, 29*, 553-584.

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

[5] Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness". *The Journal of the Acoustical Society of America, 110*, 1606-1618.

[6] Tremblay, A. (2009). Phonetic variability and the variable perception of L2 word stress by French Canadian listeners. *International Journal of Bilingualism, 13*, 35-62.

[7] Sluijter, A. M. C., & van Heuven, V. J. (2016). Acoustic correlates of linguistic stress and accent in Dutch and American English *Proceedings of the International Congress of Spoken Language Processing*. Newark: University of Delaware.

[8] Chao, Y.-R. (1968). *A grammar of spoken Chinese.* Oakland, CA: University of California Press.

[9] Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford: Oxford University Press.

[10] Jun, S.-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology, 15*, 189-226.

[11] Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus, 14*, 147-172.

[12] Connell, K., Hüls, S., Martínez-García, M. T., Qin, Z., Shin, S., Yan, H., & Tremblay, A. (in press). English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study. *Language Learning*.

[13] Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45*, 207-228.

[14] Tremblay, A., Broersma, M., & Coughlin, C. E. (in press). The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition*.

[15] Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language, 51*, 523-547.

[16] Tremblay, A., Coughlin, C. E., Bahler, C., & Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Laboratory Phonology, 3*, 385-423.

Does the mental lexicon contain representations for reduced word pronunciation variants? Evidence from native listeners and language learners

Sophie Brand[1] & Mirjam Ernestus[1,2]

[1]*Centre for Language Studies, Radboud University (The Netherlands)*

[2]*Max Planck Institute for Psycholinguistics (The Netherlands)*

swmbrand@gmail.com, m.ernestus@let.ru.nl

In casual conversations, words can occur in pronunciation variants that are reduced compared to their full, citation variants. For instance, the English word *police* /pə'li:s/ may sound like /p'li:s/. Previous research has shown that listeners rely on the sentence's grammatical structure and on fine phonetic detail to understand reduced word pronunciation variants (e.g., [1,2]). This suggests that listeners reconstruct the words' full variants on the basis of different types of information. Possibly, listeners may also recognize reduced word pronunciation variants because they have stored these in their mental lexicons. This talk discusses studies investigating the potential storage of word pronunciation variants and its consequences.

Ranbom and Connine [3] were the first to investigate the lexical storage of word pronunciation variants. They argued that if word pronunciation variants are lexically stored, they should show the same processing as the words' full variants. That is, these variants should be recognized more quickly the higher their frequencies of occurrence. Ranbom and Connine showed that this is true for the flap variants of American English words like *winter*, which can be pronounced with /nt/ (the full variant) and with a flap.

We further investigated the possibility of the lexical storage of reduced word pronunciation variants in two studies. Our first study [4] aimed at excluding alternative explanations for Ranbom and Connine's results [3]. Listeners heard words, which they had to classify as real French words or pseudo words (lexical decision). In the experimental real words, a schwa was present in the first syllable (the full variant of a word, e.g. /pəluz/ *pelouse* 'lawn') or absent (the reduced variant, e.g. /pluz/ for /pəluz/). Afterwards, the participants rated the frequency of occurrence of each word's full variant relative to the frequency of occurrence of its reduced variant (and vice versa). Importantly, the ratings provided by the native listeners differed from those provided by the Dutch advanced learners of French. This difference reflects the difference in experience the two listener groups have with French reduced word pronunciation variants. In line with Ranbom & Connine results [3], our lexical decision experiment showed that both the native listeners and advanced learners of French reacted more quickly to a word pronunciation variant the higher its relative frequency. Importantly, a group's reaction times only correlated with the relative frequencies provided by that very same group: the native listeners' reaction times did not correlate with the advanced learners' relative frequencies, or vice versa. This supports the hypothesis that it is the listeners' experience with a word variant that determines their processing of the variant. Listeners have stored the (relative) frequencies of at least some word pronunciation variants.

Our second study on the lexical storage of reduced word pronunciation variants investigated what inhibits the word recognition process more: the low frequency of occurrence of a variant (which inhibits lexical access) or a larger deviation from the full variant (which may inhibit reconstruction of the full variant). We focussed on French words ending in an obstruent-liquid-schwa cluster, like /ministʁə/ *ministre* 'minister'. Our corpus study [5] has shown that this cluster is mostly produced without schwa (e.g. as in /ministʁ/, in 37% of tokens, henceforth OL variant), and, moreover, that it is more often completely absent (e.g. as in /minis/, 16%; henceforth X variant) than that only the liquid surfaces (e.g. as in /minisʁ/, 1%, henceforth L variant). Native listeners and Dutch advanced learners of French heard prime words in the middle of sentences. Right after a prime word, they saw a target word on the screen, which they had to classify as a real French word or as a pseudoword. Each experimental target word (e.g. <ministre>) had as its prime word the same word in its OL variant (e.g. /ministʁ/), in its L variant (e.g. /minisʁ/), or its X-variant (e.g. /minis/), or an unrelated word. Participants classified the experimental target words most slowly when the prime words were unrelated to the target words. Importantly, they reacted more quickly when the obstruent-liquid-schwa cluster of the prime word was presented in its (highly reduced and highly frequent) X variant than in its (mildly reduced and rare) L variant. This shows that the frequency of the variant affects processing speed more than degree of reduction does.

Together these experimental results show that a word pronunciation variant’s frequency of occurrence plays a major role in speech processing. One might hypothesize that listeners store the (relative) frequencies of the variants with the relevant reconstruction rules (e.g. a schwa insertion rule, reconstructing /pəluz/ from the input /pluz/), and that a reconstruction rule applies more quickly (or its output is a more probable pre-lexical representation) if the resulting word pronunciation variant is more frequent. This account, however, does not work, because the frequency of the reduced word pronunciation variant can only become available after the full variant has been reconstructed and recognized. This frequency then cannot affect the recognition process.

In contrast, the alternative hypothesis that word pronunciation variants are lexically stored together with their frequencies of occurrence can account for our results. We therefore argue that our results strongly suggest that word pronunciation variants can be stored. This hypothesis has the additional advantage that word pronunciation variants do not differ from full variants in their processing, which keeps the word recognition model as simple as possible. However, this hypothesis raises many other questions, the most important ones being “Which word pronunciation variants are stored?” and “Is there also lexical competition between pronunciation variants of the same word?”.

References

[1] Viebahn, M., M. Ernestus, & J. McQueen (2015). Syntactic predictability in the recognition of carefully and casually produced speech. Journal of Experimental Psychology: Learning, Memory, and Cognition 41, 1684-1702.

[2] Ernestus, M., H. Kouwenhoven, & M. van Mulken (2017). The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics* 62, 50-64.

[3] Ranbom, L.J. & C.M. Connine (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language* 57, 273-298.

[4] Brand, S., & M. Ernestus, (2017). Listeners’ processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: evidence from native listeners and learners of French. *Quarterly Journal of Experimental Psychology*. doi.org/10.1080/17470218.2017.1313282

[5] Brand, S. & M. Ernestus (submitted).

Conversational Speech Reduction across Languages, Second Languages, and Dialects

Natasha Warner, Seongjin Park

University of Arizona (USA)
nwarner@email.arizona.edu, seongjinpark@email.arizona.edu

In natural speech, speakers do not produce all the sounds one would expect in a word based on its lexical form [1, 2, 3]. For example, a speaker saying "pretty good" might reduce the first word to [p^{h}ɹɪi], with only a small dip in amplitude or no trace of the /t/. This is especially common in spontaneous casual conversation between interlocutors who know each other well. An example from one of our recordings is "And I didn't really know" produced as [ʌ̃ɛ̃ːɚ ɹl̩ĩ no^{u}], with deletion and reduction of all consonants of the first three words. In the current study, we examine whether the reduction shown in conversational speech differs in American English vs. Japanese, and whether it differs in three dialects of English. That is, is conversational speech reduction language specific? Past literature has shown that similar types of reduction occur in many languages [3, 4]. For example, a wide range of languages contain tokens of voiced stop phonemes realized as approximants rather than stops. However, direct quantitative comparison of reduction in multiple languages or dialects has not been common. We further examine native Japanese speakers' production of conversational speech reduction in their L2, English.

As part of a larger project, we recorded 8-13 native speakers each of American English (mostly from Arizona), New Zealand English, Canadian English (Edmonton), and Japanese. The Japanese speakers were recorded both in their native Japanese and their L2 English. Results on only the Arizona English data were previously reported in [5]. Each speaker was recorded talking on the phone with a friend or family member for 10-15 minutes, and reading a word list. The Japanese speakers were living in the U.S. or Canada, and performed both tasks in both languages. All recordings were made in sound-treated booths with a head-mounted microphone, and the telephone was used only to allow the speakers to converse comfortably with a well-known interlocutor while seated in the booth. The English wordlist contained tokens of /p, t, k, b, d, g/ in intervocalic position before unstressed vowels, either at word boundary (e.g. would a, work again), or word-medially (e.g. baby, credit, broccoli). The Japanese wordlist contained only word-medial stops (e.g. /daigaku/ 'university,' /apaato/ 'apartment'), since word-final stops are impossible in Japanese. For the conversational data, researchers listened to each conversation and identified which words' lexical forms contain a stop phoneme in the corresponding environments (e.g. "writing," "yogurt," "it if," /dakedo/ 'however,' /hayaku/ 'quickly').

Several acoustic features of each stop (or location where a stop would be expected lexically) were measured. For the current paper, we focus on the measurement of drop in intensity from the intensity peaks of the preceding and following vowel to the minimum during the consonant, however it was realized in a given token. If a stop phoneme is realized as a voiceless stop, the intensity drop from the surrounding vowel peaks is large. If the stop phoneme is realized as an approximant, the intensity drop is small. If the stop phoneme is deleted, there will be an intensity drop because the measurement method identifies the maxima of surrounding vowels and minimum between them, but such an intensity "drop" will be extremely small. Thus, a larger value of intensity drop indicates a more stop-like, consonantal production, and a smaller value indicates a realization of the stop phoneme that is more like an approximant or deletion.

The results show several patterns. Overall, stop phonemes in conversational speech are realized as more approximant-like than those in wordlist reading, indicating reduction in casual conversation. We examine /t, d/ separately from other stops, since /t, d/ in this environment are expected to be realized as flaps in American English, so they demonstrate effects of the phonological flapping process as well as of conversational speech reduction. For non-alveolars (not subject to flapping), Japanese speech shows a greater difference between conversation and word-list than either Japanese speakers' L2 English or American English, which are comparable (Fig. 1). The difference in the size of the speech style effect primarily reflects more stop-like tokens in Japanese read speech, while the conversation tokens are similar for Japanese, L2 English, and American English, except perhaps for greater reduction of Japanese /g/ due to nasalization of this phoneme. This suggests that Japanese uses clearer stop productions in careful (word-list reading) speech, while all three varieties show a similar amount of reduction of stops in conversational speech. The clearer stop productions in Japanese may reflect mora-rhythm (clear CV syllables in read speech, while conversational Japanese shows reductions, deletions, and elisions just as any other language does [6]). The /t, d/ phonemes, where flapping is expected in English, show related patterns (Fig. 2). However, they show the L2 English produced by Japanese speakers as intermediate between their L1 Japanese (where /t, d/ are fully distinct and /t/ shows more reduction with conversation) and American English (where /t, d/ do not differ, and are so approximant-like even in read speech that they do not show substantial further reduction in conversation). Overall, these results show the Japanese speakers acquiring speech-style-appropriate reduction in their L2 English, and partially acquiring the phonological flapping pattern. We also investigated word frequency effects in this data, and found only a few significant effects of word frequency, consistent with L2 speakers producing clearer stops in less frequent words. We further compare related effects in American vs. New Zealand vs. Canadian English.

Overall, these results show that there is a language-specific attribute to conversational speech reduction. Furthermore, the results show native Japanese speakers successfully acquiring the phonetics of speech style effects in their L2 English, even though they have only partially acquired the phonological flapping pattern.

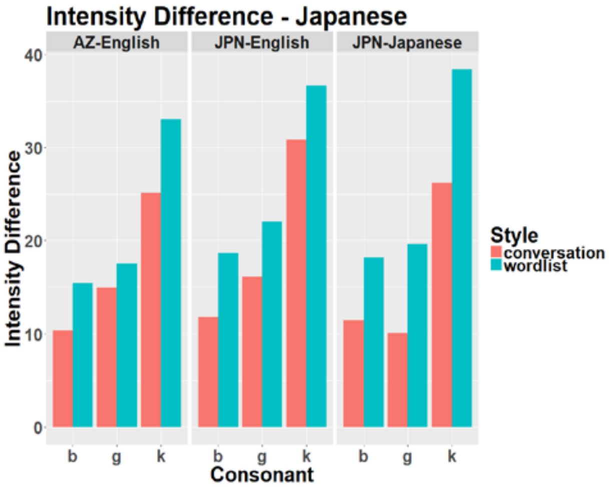


Fig.1 Intensity drop from neighboring vowel peaks to consonant minimum, for non-alveolar consonants with sufficient tokens in conversation. JPN-English is Japanese speakers' L2 English.

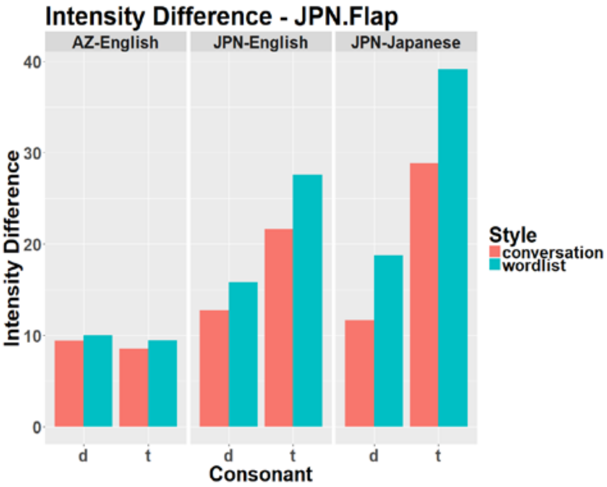


Fig.2 Intensity drop from neighboring vowel peaks to consonant minimum, for alveolar (flapping) consonants. JPN-English is Japanese speakers' L2 English.

References

[1] Ernestus, M. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface.* Utrecht: LOT.

[2] Keune, K., Ernestus, M., Van Hout, R., & Baayen, R.H. (2005). Social, geographical, and register variation in Dutch: From written 'mogelijk' to spoken 'mok.' *Corpus Linguistics and Linguistic Theory, 2,* 183-223.

[3] Ernestus, M., and Warner, N. 2011. An introduction to reduced pronunciation variants. Guest editors' introduction to the special issue of *Journal of Phonetics* on speech reduction. *39(3*), 253-260.

[4] Barry, W. & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association, 31,* 51-66.

[5] Warner, N., and Tucker, B.V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America 130,* 1606-1617.

[6] Arai, T. (1999). A case study of spontaneous speech in Japanese. *Proc. of the International Congress of Phonetic Sciences (ICPhS), San Francisco, 1,* 615-618.

Oral Presentations

(Day 1)

Effects of Cognitive Development in Speech Perception

Paola Escudero[1,2], Alba Tuninetti[1,2], James Whang[1,2]

[1]*MARCS Institute, Western Sydney University, Sydney, NSW, Australia*
[2]*ARC Centre of Excellence for the Dynamics of Language, Canberra, ACT, Australia*
paola.escudero@westernsydney.edu.au, a.tuninetti@westernsydney.edu.au, research@jameswhang.net

In order to handle various sources of information during speech perception, listeners must be able to perceive, categorise, and abstract. Specifically, listeners must be able to ignore irrelevant, non-linguistic (indexical) information (e.g., speaker identity) in the speech signal to arrive at the underlying linguistic message (i.e., the lexical and semantic tokens). However, this process is predicated on the ability to discriminate indexical and linguistic cues. The current study presents new electroencephalography (EEG) data on infant speech perception and discusses their significance in the context of recent related behavioural, neural, and modelling studies on how adults and infants process the two types of cues. We also propose a framework to explain dissociations between neural and behavioural data, as well as developmental changes in speech perception from infancy to adulthood.

Behavioural studies show that adults are skilled at ignoring indexical information to arrive at the intended linguistic information, adapting within very short time frames to different speakers [1,2] and even accents [3,4]. However, adult listeners seem to have trouble ignoring accentedness when presented with natural isolated vowel tokens and asked to generalise across four different speech changes (three indexical: speaker, sex, accent; one linguistic: vowel; [5]). Participants were able to generalise new tokens to existing categories in all conditions except those that were accented; they were only able to generalise accent when presented concurrently with feedback. Tuninetti and colleagues [6] used the same stimuli from Kriengwatana et al. [5] in an EEG experiment and found that while adults are indeed most sensitive to accent changes in a preattentive paradigm examining MMN responses, they are also similarly sensitive to changes in sex. These two changes had the largest voice quality difference (F0) from the standard stimulus, demonstrating that the MMN response reflected the auditory difference between these stimuli. The neural sensitivity and behavioural insensitivity to change in sex in particular reveal a dissociation, such that adults are able to ignore certain indexical information when tasked with adapting across those differences, but they are actually sensitive to them at a neural level.

Unlike with adults, behavioural studies of infants suggest equal sensitivity to linguistic and indexical information. Mulak and colleagues [7] presented 12-month-olds with the same stimuli as Kriengwatana et al. [5] and Tuninetti et al. [6] in an eye-tracking paradigm and found that after familiarization, looking times were not significantly different for changes in speaker, accent, or vowel identity. Rather, looking times increased overall suggesting that while 12-month-olds are able to notice speech changes, they do not discriminate indexical and linguistic information. To explore possible behavioural and neural dissociations as those seen in adult studies, we conducted a new EEG study with fifty-six 12-month-old infants, using the same stimuli and experimental paradigm as Tuninetti et al. [6]. We found that the infants show positive mismatch responses (MMR) to sex, vowel, and accent changes in the speech stream, which were not significantly different ($F < 1$), supporting the behavioural results. However, the infants also showed an MMN response specifically to speaker changes. The MMN response is thought to be a more mature index of processing, usually associated with more cognitive processing [8] and suggests particular salience above other cues of speaker change at the neural, preattentive level, a distinction that was lost in behavioural studies. However, as these are the same stimuli as those in earlier studies, the preattentive responses do not reflect voice quality changes as they did with adult participants. Instead, it is important to note that the stimuli were all female tokens; this suggests that the heightened sensitivity to the speaker change in infants may reflect an inherent bias to prefer a familiar female speaker over an unfamiliar female speaker. We argue that this inherent bias may come from the known preference for a mother's voice compared to unfamiliar female speakers (e.g., [9]), but future work will test to see if they show the same sensitivity to speaker changes with male tokens.

Additionally, the studies above show together a clear difference between infants and adults in how they perceive and process both linguistic and indexical sources of information. Recent models of speech perception typically focus either on adults [10,11] or infants [12], and although informative, they reveal little about how the human speech perception system develops from infant-like perception and processing of both indexical and linguistic cues to adult-like focus on linguistic cues. We propose a framework that utilises the belief-updating mechanism of an ideal adapter [10] but also integrates insight from the cognitive development literature to introduce limits on information-processing capacity, which increases with age [13]. Capacity limits make cue prioritisation imperative, and thus a goal-based prioritisation mechanism (e.g., identify mother, access lexicon, etc.) is necessary to explain how infants develop from not discriminating indexical from linguistic cues to eventually learning to ignore irrelevant indexical information as adults do.

References

[1] Clarke, C. M. & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, 116(6), 3647-58.

[2] Kraljic, T. & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56, 1-15.

[3] Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32, 543-62.

[4] Scharinger, M., Monahan, P. J., & Idsardi, W.J. (2011). You had me at "Hello": Rapid extraction of dialect information from spoken words. *NeuroImage*, 56(4), 2329-38.

[5] Kriengwatana, B., Terry, J., Chládková, K., & Escudero, P. (2016). Speaker and accent variation are handled differently: Evidence in native and non-native listeners. *PloS one*, 11(6), e0156870.

[6] Tuninetti, A., Chládková, K., Varghese, P., Schiller, N. O., & Escudero, P. (2017). When speaker identity is unavoidable: Neural processing of speaker identity cues in natural speech. *Brain and Language*, 174, 42-49.

[7] Mulak, K. E., Bonn, C. D., Chládková, K., Aslin, R. N., & Escudero, P. (2017). Indexical and linguistic processing by 12-month-olds: Discrimination of speaker, accent and vowel differences. *PloS one*, 12(5), e0176762.

[8] He, C., Hotson, L., & Trainor, L. J. (2007). Mismatch responses to pitch changes in early infancy. *Jour nal of Cognitive Neuroscience*, *19*(5), 878-892.

[9] Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martinez- Montes, E., Belin, P., Beland, R., Francoeur, D., Carceller, A-M., Wallois, F., & Lassonde, M. (2011). Mother and stranger: An electrophysiological study of voice processing in newborns. *Cerebral Cortex*, *21*(8), 1705–1711. http://doi.org/10.1093/cercor/bhq242

[10] Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148-203.

[11] Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39, 456-466.

[12] Werker, J. F. & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1(2), 197-234.

[13] Halford, G. S. & Andrews, G. (2002). Information-processing models of cognitive development. *The Wiley-Blackwell Handbook of Childhood Cognitive Development, Second edition*, 697-722.

Implementing Finite State Intonational Grammars to Understand Gradient Prosodic Manipulations in Infant-directed Speech

Kristine M. Yu[1], Sameer ud Dowla Khan[2] & Megha Sundara[3]

[1]University of Massachusetts Amherst (USA), [2]Reed College (USA), [3]University of California Los Angeles (USA)
krisyu@linguist.umass.edu, skhan@reed.edu, megha.sundara@humnet.ucla.edu

The linguistic function of fundamental frequency (f0) variation in an utterance has long been understood from the perspective that the f0 variation arises as the realization of a well-formed sequence of discrete tonal elements. Which tonal sequences are well-formed (and thus, what range of f0 variation occurs) is determined by the language-specific intonational grammar. Intonational grammars—as well as the linguistic meanings of tonal elements and well-formed tonal sequences—have been proposed for diverse languages of the world ([1-3], i.a). Although this body of intonational phonology literature provides a foundation for understanding linguistic functions of gradient modulations in the f0 contour, working towards this understanding runs up against a number of methodological challenges: (1) the entanglement of extra-linguistic and linguistic factors in conditioning f0 variation, (2) evolving hypotheses about what the tonal elements and licit sequences of tonal elements in the intonational grammar are, (3) the generalizability of proposed grammars to a wider range of speech styles and contexts, and (4) the analysis of tonal elements in the context of the tonal sequences they occur in. This paper presents a strategy for confronting these challenges: *implementing proposed intonational grammars as finite state machines (FSMs)*, see Figure 1. Here, we use a classic example of gradient prosodic manipulation—infant-directed speech (IDS) or "motherese" (a re-analysis of [4])—as a case study to show how implementing intonational grammars as FSMs can help address the methodological challenges enumerated above.

Relative to adult-directed speech, IDS has been shown to exhibit higher mean f0, higher maximum f0, an expanded f0 range, and greater f0 variability across a variety of languages ([5, 6], et seq.). The function of these f0 manipulations in IDS has largely been characterized as extra-linguistic: motivated by the need to regulate infant attention and affect (and unaffected by intonational grammar). But the first detailed analysis of IDS using intonational grammars on Tokyo Japanese [7] presented a striking result: previously, it had been thought that IDS in Tokyo Japanese did not exhibit an expanded f0 range like in other languages. [7] showed that in fact, Tokyo Japanese IDS did show an expanded f0 range, but only in the realization of boundary tones. This work thus characterized one way in which extra-linguistic and linguistic factors jointly condition f0 modulation: the phonetic "exaggeration" (e.g., a wider pitch range) of specific tones. [4] found evidence for this kind of joint conditioning in Bengali and English, too, as well as characteristic gradient aspects of f0 variation in IDS effected via categorical changes (e.g, more boundary tones of a certain type).

A problem with the analysis in [4] (and many other extant intonational analyses, see [8], et seq.) is that tonal elements were treated as independent events, e.g., we compared the frequency of HLH% boundary tones in Bengali between recorded corpora of IDS and non-IDS speech without considering the sequences in which HLH% tones actually arose. However, of course, changes in choices of intonational elements in an utterance are systemic and correlated—tonal elements occur as part of a sequence. To take this into account in our re-analysis, we defined intonational grammars for Bengali and English developed for non-IDS as high-level phonotactic restrictions in *xfst* [9], software written by linguists that compiles high-level regular expressions (e.g., SPE rules) into FSMs.[1] We then examined how well these models generalized to IDS and used the distribution of intonational elements in the recorded corpora to estimate the probabilities of different tonal choices within the FSMs. Crucially, the computed probability to reach a state corresponding to a particular

[1] Figure 1 is misleading in that it makes it seem like we might be able to define FSMs for intonational grammars and compute with them by hand on a piece of paper. But in fact that the FSM shown for Bengali in Figure 1 only generates the most common sequences; the FSM compiled by xfst for Bengali has over a hundred arcs between states.

tonal choice was dependent on the probabilities of different paths through the FSM that could reach that state. Finally, we used the computed probabilistic finite state machines to quantify the effect of speech style on the probability of particular tonal events within each language.

Implementing the FSMs allowed us to analyze the entanglement of extra-linguistic and linguistic factors in conditioning f0 variation in terms of categorical changes in tone choices between IDS and non-IDS, with an analysis considering tonal elements in the context of the tonal sequences they occur in. Moreover, implementing the intonational grammars allowed us to write python software to parse all of the intonational transcriptions to confirm that they were accepted by the grammars. This allowed us to check the generalizability of the proposed grammars to a wider range of speech styles and contexts than they were originally developed for. In particular, we discovered the current proposal on restrictions in deaccenting in Bengali did not admit some of the transcribed data. Given that proposed intonational grammars are hypotheses and thus subject to revision, working with the FSMs proved to be valuable because it allowed us to go ahead and work with the current hypothesis, while also helping us consider how it might be revised in the future.

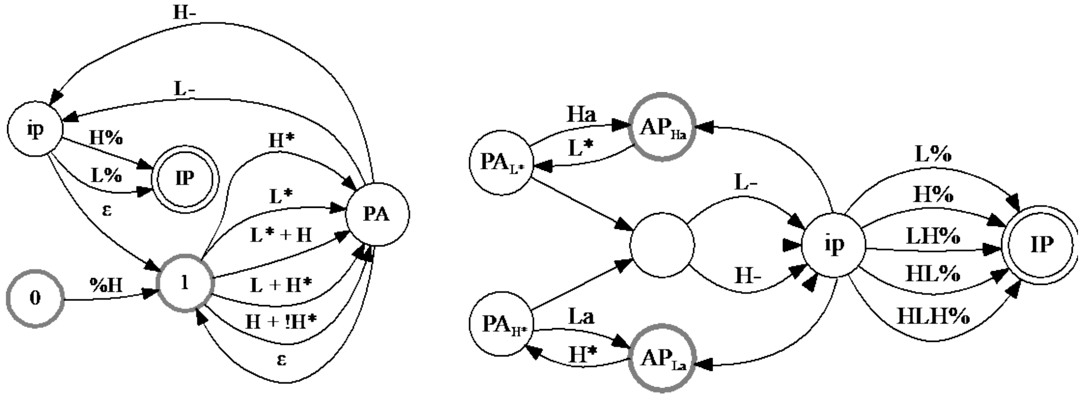


Fig. 1 Finite state machines (FSMs) generating (some) licit intonational tunes in Mainstream American English (left; after [10], [11], i.a.) and in Standard Bangladeshi Bengali (right; after [12], [13]).

References

[1] Jun, S.A., ed. (2005). *Prosodic typology*. Oxford: Oxford University Press.

[2] Jun, S.A., ed. (2014). *Prosodic typology II.* Oxford: Oxford University Press.

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

[4] Yu, K.M., Khan, S.D, & Sundara, M. (2014). Intonational phonology in Bengali and English infant-directed speech**.** *Speech Prosody 2014.* 1130-1133.

[5] Stern, D.N., Spieker S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology, 18(5),* 727–735.

[6] Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang., 16,* 477–501.

[7] Igarashi, Y., Nishikawa K., Tanaka K., & Mazuka R. (2013). Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *J. Acoust. Soc. Am., 134(2),* 1283–1294.

[8] Dainora, A. (2001). An empirical based probabilistic model of intonation in American English. Ph.D. dissertation. University of Chicago, Chicago, IL.

[9] Beesley, K. R., & Karttunen, L. (2003). *Finite state morphology*. CLSI: Stanford.

[10] Pierrehumbert, J.B. (1980). The phonology and phonetics of English intonation. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

[11] Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. *The original ToBI system and the evolution of the ToBI framework*. (2005). In S. A. Jun (Ed.), *Prosodic typology,* Ch. 2: 9-54.

[12] Khan, S.D. (2008). Intonational phonology and focus prosody of Bengali. Ph.D. dissertation, University of California, Los Angeles, CA.

[13] Khan, S.D. (2014). The intonational phonology of Bangladeshi Standard Bengali. In S.-A. Jun (Ed.), *Prosodic typology II.* Oxford, UK: Oxford University Press.

Effects of Precursor Context on Categorical Perception of Mandarin Tones in Disyllabic Words

Hui Zhang[1], Hongwei Ding[2] & Wai-Sum Lee [3]

[12]*Shanghai Jiao Tong University (China), [3]City University of Hong Kong (HK)*
Zhanghui_Helen@126.com, hwding@sjtu.edu.cn, w.s.lee@cityu.edu.hk

This study explores whether Mandarin listeners show normalization for F0 variation caused by coarticulation in disyllabic word.

Mandarin Chinese, a language with four tones, uses F0 to discriminate lexical meaning. The citation form of the four tones, in Chao's [1] tone notation system, is tone 1 (level tone: 55), tone 2 (rising tone: 35), tone 3 (low tone: 213) and tone 4 (falling tone: 51).

F0 realization of tones in the second syllable of Mandarin disyllabic words, however, deviates from its citation form in conflicting phonetic context. The deviation can even result in one tone resembling the citation form of another. In spite of deviation, native listeners are not bothered by the F0 variation in the second syllable. This is possibly because the listeners are able to rescale tone category in the second syllable relative to its preceding tone context.

The process, that listeners rely on extrinsic phonetic context for tone perception, is called tonal normalization. Abundant research has shown that a preceding context with different mean F0 or different F0 range, caused by inter-speaker and intra-speaker difference, shifted the perception of the following tone in a contrastive manner [2, 3, 4, 5].

Previous studies used a sentence as precursor context and the context was separated from the target sound by an interval of silence. That is to say, listeners are able to normalize the F0 variations caused by inter-speaker and intra-speaker difference and rescale what the citation form of a tone should be like for a particular context. What remains unknown is whether listeners are capable of normalizing the F0 variation resulting from coarticulation in conflicting phonetic context in disyllabic words. Although Rong [6] summarized the tendency of categorical perception of tones preceded by different tones in disyllabic words, segmental structures in the target sound are inconsistent and it is also in lack of statistical evidence. The first purpose of the study is to explore the effect of tone context on the categorical perception of low tone and falling tone contrasts in the second syllable of Mandarin disyllabic words.

In addition, there is discrepancy on whether normalization happens with speech mechanism or general auditory mechanism. Zhang et al. [4], Chen and Peng [5] and Zhang, Wang and Peng [7] are in favor of speech mechanism while Huang and Holt [2,3] suggested a general auditory mechanism. The second purpose is to examine whether tone normalization in this study happens with a speech mechanism or a general auditory mechanism.

The present study takes the Mandarin syllable [huo] continuum with 14 steps as target sound. One endpoint is low tone meaning "fire" and the other is falling tone meaning "goods". The target sound is preceded by three types of speech context, [fa] with level tone, [ʂa] with rising tone and [ɕia] with falling tone respectively. All the six resulting disyllabic words are frequently used words in Mandarin. Nonspeech context is monotone synthesized by Praat [8]. Acoustics of the three types of nonspeech context corresponds to that of three types of speech context. Sixteen subjects did this experiment on Qualtrics. Each subject listened to the same stimuli for a repetition of 7 times. In block one, participants were asked to listen to target sound in monosyllable and determine whether it is 豁 (huō) (level), 活 (huó) (rsing), 火 (huǒ) (low) or 货 (huò) (falling). In block two, they were asked to listen to the target sound immediately preceded by a nonspeech context and do the same task as block 1. In block three, they were asked to listen to the target sound preceded by three types of speech context and do the same task with block 1. The order of stimuli in each block was randomized whereas the order of the three blocks was fixed.

The average percentage for falling tone identification was counted across seven times repetition. Boundary was determined by Probit analysis at the identification of 50%. The results of two-way repeated measures ANOVA (2 speech types × 3 tone types) show that boundary of the target sound is only influenced by speech context. When preceded by speech context, boundary is highest when the first syllable is rising, second highest when the first syllable is high level, second lowest in isolation and lowest when the first syllable is falling. Post hoc analysis reaches significance for each comparison. This indicates that listeners show tone normalization for F0 variations of the second syllable in disyllabic words and normalization is a speech mechanism. Normalization is contrastive in that the higher offset will lead listeners to show a bias towards a lower tone in perception. What is interesting is that although rising tone and level tone have the same offset, the boundary of target sound following rising tone is significantly higher than that of level tone. One possible reason is that the perceived height of offset for rising tone is higher than that for level tone.

References

[1] Chao, Y.-R. (1930). A system of tone-letters. *Le Maître Phonétique*, 45,24–27.

[2] Huang, J., & Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *The Journal of the Acoustical Society of America*, 125(6), 3983-3994.

[3] Huang, J., & Holt, L. L. (2011). Evidence for the central origin of lexical tone normalization (L). *The Journal of the Acoustical Society of America*, 129(3), 1145-1148.

[4] Zhang, C. C., Peng, G., & Wang, W. S. Y. (2012). Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. *The Journal of the Acoustical Society of America*, 132, 1088–1099.

[5] Chen, F., & Peng, G. (2016). Context effect in the categorical perception of mandarin tones. *Journal of Signal Processing Systems*, 82, 253–261.

[6] Rong, R. (2013). *Tone Perception Pattern in Mandarin [汉语普通话声调的听感格局]* (Doctoral dissertation, Nankai University).

[7] Zhang, K., Wang, X., & Peng, G. (2017). Normalization of lexical tones and nonlinguistic pitch contours: Implications for speech-specific processing mechanism. *The Journal of the Acoustical Society of America*, 141(1), 38-49.

[8] Boersma, P., & Weenink, D. (2014). Praat: Doing Phonetics by Computer [Computer software]. Version 5.3. 84.

Mothers Would Rather Speak Clearly than Spread Innovation: The Case of Korean VOT

Eon-Suk Ko

Chosun University
eonsuk@gmail.com

The effects of voice onset time (VOT) and the fundamental frequency (f0) of the following vowel in distinguishing lenis vs. aspirated stops in Korean have attracted much attention due to the change in the role each of these cues has played as a primary cue in production [1]. This paper tests the tonogenesis hypothesis in Korean by investigating child-directed speech (CDS), which might serve as the source of sound change. Results of logistic regression analyses show that VOT, often thought to have completely neutralized, still plays a significant role in discriminating the lax from the aspirate series in CDS. We suggest that mothers might provide enhanced VOT to infants because it plays a primary role in perception.

Earlier acoustic studies [2, 3] suggest that the three-way laryngeal distinction in Korean stops was mainly achieved by the differences in VOT. Recent consensus, however, is that the lenis and the aspirate series have merged in VOT and are now solely distinguished by the difference in f0. According to an influential theory of sound change [4], sound change can occur from listener failing to reconstruct the target representation of the speaker. Applied to the context of Korean tonogensis, this would mean that mothers fail to deliver the VOT cue and children treat the f0 as the primary cue for distinguishing lenis from aspirate.

The mothers participating in our study belong to the group of young female speakers that are reported to adopt the most advanced form of the historical change. We analyzed 4,695 tokens of phrase-initial stops produced in the CDS of 35 mother-child pairs engaged in spontaneous interactions divided in three age groups: preverbal (6-9 months old, 12 dyads), early speech (12-16 mo, 11 dyads), and multi-word (25-28 mo, 12 dyads) stage. Each recording lasted 50 minutes, including 10 minutes of adult-directed speech (ADS) from the same mothers.

We constructed a mixed effects logistic regression model for each age group to investigate the relative contribution of each predictor in determining the dependent variable (lenis vs. aspirated). We constructed the models based on standardized VOT and pitch values. The models included random intercepts for subjects. We also included random slopes for the two fixed effects since likelihood tests showed the inclusion of these to be significant (χ^2=19.297, p<0.001, df=2 for pitch, and 12.387 p=0.002, df=2 for VOT). Assuming that each of these fixed effects and the intercept are independent from each other, we used the following formula:

dv ~ scaled.vot+scaled.f0+(1|subject)+(scaled.vot+0|subject)+(scaled.f0+0|subject).

The results show that, in ADS, consistent with the claims in previous research, only f0 plays a significant role in distinguishing lenis from aspirate stops. However, in CDS, VOT played a significant role in distinguishing the two stop categories (Table 1). Further, we found that the f0 difference is most enhanced when the child is at the stage of vocabulary spurt (Figure 1).

It thus seems that mothers provide an enhanced secondary cue for children to help distinguish lenis and aspirate categories. It might be that the primacy of the cues in perception is different from production, and CDS provides enhancement for the cue known to play an important role in perception [5]. That is, Korean mothers' primary interest might be in facilitating the perceptual development of the infants rather than spreading the innovation in sound change. The results also suggest that the role of VOT as an acoustic cue for the lenis vs. aspirate distinction in Korean consonants is not to be underestimated, and that it is premature to interpret the emerging role of f0 in Korean phonology as tonogenesis.

Table 1 Coefficients in mixed effects logistic regression models using lme4 package of R (model with random intercept for subjects, and random slopes for duration and pitch)

	Age 0	Age 1	Age 2	ADS
F0	1.54 ***	1.53 ***	2.56 ***	2.75 ***
VOT	0.52 **	0.59 *	0.82 ***	-0.10

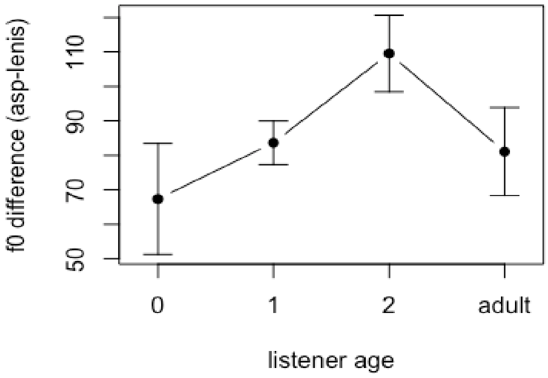


Fig. 1 F0 differences between aspirated and lenis consonants as a function of listener age (0=preverbal, 1=early speech, 2=multi-word stage)

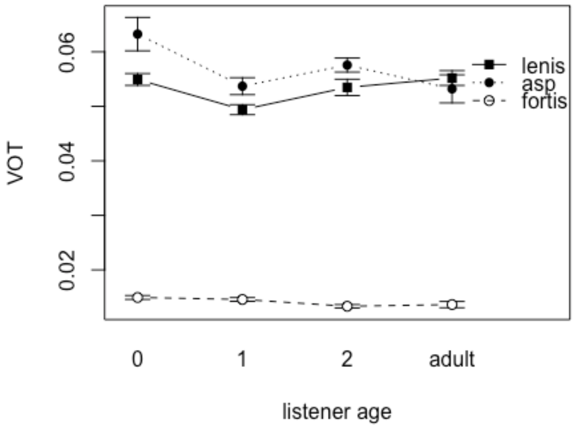


Fig. 2 VOT values as a function of listener age (0=preverbal, 1=early speech, 2=multi-word stage)

References

[1] Silva, David James. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23.2. 287–308.

[2] Lisker, Leigh, and Arthur S. Abramson. (1964). A cross-linguistic study of voicing in initial stops: Acoustic measurements. *Word* 20.384-422.

[3] Han, Mieko and Weitzman. (1970). Acoustic feature of Korean /P,T,K/, /p,t,k/, and /ph,th,kh/. *Phonetica* 22:112-128.

[4] Ohala, J. John. (1981). The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Ling. Soc. 178 - 203.

[5] Kong, E. & Lee, H. (2016). Attentional modulation and individual differences in explaining the changing role of f0 in the Korean laryngeal stop perception, poster presented at Laboratory Phonology Conference 2016.

Constructing L2 Phonetic Categories: The Influence of Variability in Neural Responses during Training

Alba Tuninetti[1,2] & Natasha Tokowicz[3]

[1] MARCS Institute, Western Sydney University, Penrith, NSW, Australia, [2] ARC Centre of Excellence for the Dynamics of Language, Canberra, ACT, Australia, [3] Learning Research & Development Center, University of Pittsburgh, Pittsburgh, PA, USA

a.tuninetti@westernsydney.edu.au, tokowicz@pitt.edu

Even with years of practice, adult learners tend to need more focused and targeted input to achieve native-like perception and production of second language (L2) sounds compared to children. The present study aims to clarify the neural mechanisms through which L2 perception is influenced by variability in first language (L1) sounds. Native English and native Spanish speakers completed a five-day training paradigm during which they learned to discriminate nonnative Hindi sounds: /b/, /p/, and aspirated /p/, or /p^h/. In particular the /p-p^h/ contrast was of interest because the Spanish language does not use aspiration as an acoustic feature, but English does. However, it is not a meaningful distinction in English because it does not, e.g., form minimal pairs as it does in Hindi. Therefore, we predicted that there would be differences between the two language groups in their learning and perception of the nonnative Hindi contrasts based on their existing (or non-existing) L1 sounds. Specifically, it may be easier for native Spanish speakers to learn to construct a new category (aspirated /p/) than for native English speakers to “split” an existing category (/p/ into unaspirated /p/ and aspirated /p/).

Participants underwent electroencephalogram (EEG) recordings from the scalp, baseline discrimination tasks, and training. We expected that the L1 would modulate the EEG waveform known as the mismatch negativity (MMN) after sound onset elicited in an oddball paradigm. This measure indexes early phonetic learning and previous research has shown that the waveform’s amplitude can change or shift with new phonetic learning, indicating a reorganization of early acoustic and phonetic processing (see Näätänen, 2001 [1], for a review). Importantly, we manipulated the oddball paradigm such that the frequent stimuli were variable; stimuli varied from each other in increments of 10 ms of aspiration each. Therefore, participants had to construct and use phonetic categories from varying stimuli (see Näätänen, Pakarinen, Rinne, & Takegata [2], 2004). Unlike previous studies that use the same stimulus repeated as the frequent stimuli (e.g., Tuninetti & Tokowicz, 2018 [3]), varying the standard stimulus requires more naturalistic processing for using existing and constructing new phonetic categories. The training consisted of XAB tasks over three days, with pre- and post-test EEG. In the MMN paradigm, the /p/ was always the standard and the /b/ and /p^h/ phones were always the deviants.

Training responses during the XAB task improved over the three days, even as the task became more difficult (adaptive training). For EEG, our results demonstrated that both learner groups showed a modulation in the MMN waveform after training, but the change was eclipsed by the native contrast that was tested as a control. Native Hindi speakers showed an MMN response to both deviants but there was no difference between the deviants ($F < 1$), demonstrating that both deviants were perceived similarly by the native speakers, as expected. Native Spanish and native English speakers showed larger MMN responses to the /b/-deviant, as expected because this is a native contrast in both of those languages. Pre-test EEG showed no response to the /p^h/-deviant before training for either language group, but did show a significant MMN response after training. Importantly, native English speakers showed a larger MMN response to the /p^h/-deviant after training compared to native Spanish speakers, which can be explained by the fact that the existence of aspiration as a feature in English /p/ allows them to better use the variability in the frequent stimuli to shift their perceptual representation. Native Spanish speakers must build up a new category with varying stimuli which does not provide reliable cues to aspiration as an important feature. The response to the /p^h/-deviant was eclipsed by the response to the native contrast, such that the main effects and interactions showed significant results only for the /b/-deviant when compared to the other deviant. These results suggest that testing non-native contrast perception with variable standards in an MMN paradigm is feasible, but care needs to be taken in stimulus creation and presentation to highlight the effects of variability and not lose sensitivity to new speech categories.

Due to the inherently variable nature of speech, examining how the brain responds to natural speech variation for phonetic categorization allows us to more confidently examine how the MMN response indexes the underlying phonetic representation in the L1 and how that changes with exposure to L2 phonetic representations. Future researchers may wish to examine how varying standards within an oddball paradigm can more effectively mimic the natural variations within the speech signal. Although it is more difficult to control for acoustic effects, employing varying standards illustrates that the MMN is an effective objective measure of phonetic differences between categories. Our study demonstrates that the same type of MMN response can be elicited with natural speech variation, at least for contrasts that vary with aspiration and voicing, lending itself to a more ecologically valid model of speech perception.

Furthermore, the differences between the native English and native Spanish speakers suggest that variability can affect how L2 learners construct new perceptual categories. In a traditional MMN paradigm, native Spanish speakers showed a larger MMN response to the aspirated /p/ than native English speakers (Tuninetti & Tokowicz, 2018), but this was not the case for the variably-presented standards, as detailed here. This suggests that variability can hinder new category formation (as native Spanish speakers would have to do) but help highlight existing features in a category (as native English speakers have). These results are examined in light of the Perceptual Assimilation Model (PAM; Best, 1991 [4], 1995 [5]), the Speech Learning Model (SLM; Flege, 1995 [6]), the Native Language Magnet model (NLM; Kuhl et al., 2008 [7]), and the Unified Competition Model (UCM; MacWhinney, 2005 [8]), examining similarity between L1s, neural hardwiring in the brain, and competition between phonetic contrasts.

References

[1] Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent MMNm. *Psychophysiolog*y; *38*:1–21.

[2] Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards an optimal paradigm. *Clinical Neurophysiology, 115*, 140-144.

[3] Tuninetti, A., & Tokowicz, N. (2018). The influence of a first language: training nonnative listeners on voicing contrasts. *Language, Cognition, & Neuroscience,* 1-19. https://doi.org/10.1080/23273798.2017.1421318

[4] Best, C. T. (1991). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *Haskins Laboratories Status Report on Speech Research, SR-107/108*, 1-30.

[5] Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 171-204). Timonium, MD: York Press.

[6] 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*, Timonium, MD: York Press.

[7] Kuhl, P., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philosophical Transactions of the Royal Society B, 363, 979-1000. doi:10.1098/rstb.2007.2154)

[8] MacWhinney, B. (2005). A unified model of language acquisition. In J.F. Kroll & A.M.B DeGroot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49-67): Oxford University Press.

Influence of Fine-Grained Tonal Variability on Native Chinese Listeners and Second-Language Chinese Learners' Word Recognition: An Eye-Tracking Study

Zhen Qin[1], Annie Tremblay[2], Jie Zhang[2]

[1]*Shanghai Jiao Tong University, China,* [2]*University of Kansas, United States*

qinzhenquentin@yahoo.com, atrembla@ku.edu, zhang@ku.edu

Research on spoken word recognition that uses the visual-world eye-tracking paradigm has shown that native listeners are sensitive to within-category phonetic variability in the segmental domain (e.g., the fine-grained variability of VOT), with this information modulating target and competitor word activation (as indexed by listeners' fixations to target and competitor words) [1, 2]. What is unclear, however, is whether listeners are sensitive to within-category phonetic variability also in the suprasegmental domain, specifically with lexical tones. Tone differs from most segmental contrasts in that it is dynamic over time, and it shows a great deal of variability, within and across both talkers and tonal categories.

The present study investigates how fine-grained tonal variability influences native and non-native (Mandarin) Chinese listeners' word recognition. The aims of the study are twofold: (i) to test whether, and if so, how, the fine-grained, within-category phonetic variability of lexical tones modulates native Chinese listeners' word recognition as the speech signal unfolds; and (ii) to determine whether English-speaking second-language (L2) learners of Chinese differ from Chinese listeners in how the within-category variability of lexical tones modulates word recognition.

We hypothesize that the effect within-category phonetic variability will depend on the lexical tone heard and will differ between Chinese and English listeners. The present study focuses on the Tone 1 (T1, a high-level tone) – Tone 2 (T2, a rising tone) pair. Native Chinese listeners perceive the tonal boundary in the T1-T2 continuum as closer to T2 than to T1 [3]. Hence, we predict that Chinese listeners' lexical activation will be more strongly influenced by the same acoustic change towards (or away from) the competitor tone when T2 is the target than when T1 is the target. Since English listeners attend more to pitch height than to pitch contour differences in their tone perception [4]—a finding attributed to their use of pitch height in the perception of lexical stress [5], we predict that English listeners will be more sensitive to acoustic change towards (or away from) the competitor tone when T1 is target than when T2 is the target, as this acoustic change is more likely to be perceived as a pitch-height difference when T1 is the target than when T2 is the target.

A total of 36 native Chinese listeners and 26 proficient adult English-speaking Chinese learners were tested in a visual-world eye-tracking experiment. All Chinese words in the experiment were imageable monosyllabic nouns. The target word contained T1 (e.g., /jā/ 'duck') and the competitor word contained T2 (e.g., /já/ 'tooth'), or vice versa; when the target and competitor words carried T1 and T2, the two distracter words carried T3 (e.g., /tɕǐŋ/ 'well') and T4 (e.g., /tɕìŋ/ 'mirror') (Fig. 1). The T1-T2 auditory stimuli were manipulated such that the early pitch of the target tone was either canonical (standard condition, T1-S and T2-S) or phonetically more distant from (distant condition, T1-D and T2-D) or closer to (close condition, T1-C and T2-C) the competitor (Fig. 2).

Growth-curve analyses [6] were conducted on the differences between listeners' proportions of target and competitor fixations from the onset to the offset of the target word, with a 200-ms delay (from 200 to 654 ms; Fig. 3).

As can be seen in the results (Fig. 3), Chinese listeners' recognition of words containing T2 showed a decrease in tonal competition in the distant condition compared to the standard condition and an increase in tonal competition in the close condition compared to the standard condition; their recognition of words containing T1 did not show such effects. Conversely, English listeners' word recognition was disadvantaged in the distant and close conditions compared with the standard condition when the target contained T1 but similar when the target contained T2.

These results suggest that fine-grained tonal variability influenced native and non-native listeners' word recognition but did so differently for the two groups. Native Chinese listeners' sensitivity to within-category phonetic variability with T2 may be due to its limited room for within-category phonetic variability [3]. As a result, an equal decrease/increase in Hz in T1 targets may not have the same impact on competition as the same decrease/increase in Hz for T2 targets. Further research is needed to determine whether T1 stimuli with a greater difference in early pitch between the canonical and non-canonical tokens would modulate Chinese listeners' online word recognition. By contrast, English listeners' greater sensitivity to within-category phonetic variability with T1 than with T2 may be due to their using pitch height differences to perceive non-native tones [4]: A small change in the onset of T1, a level tone, may be more likely to be perceived as a pitch height difference, and thus to modulate lexical competition, than the same change in the onset of T2. These findings suggest that the native language has an important impact on the way in which within-category tonal variability modulates lexical activation in native and non-native online word recognition.

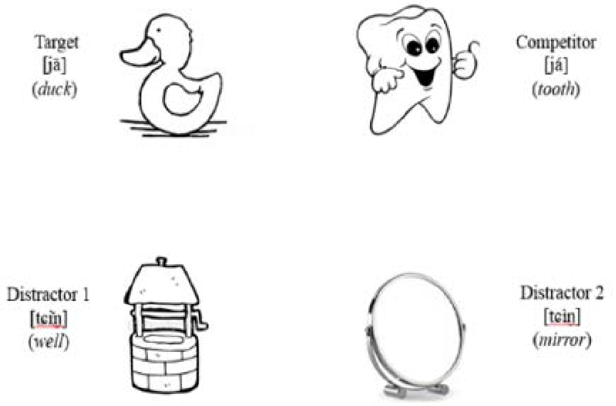


Fig. 1 A visual display of a T1-T2 trial used in the visual world paradigm (the orthographic transcriptions were not presented in the actual experiment)

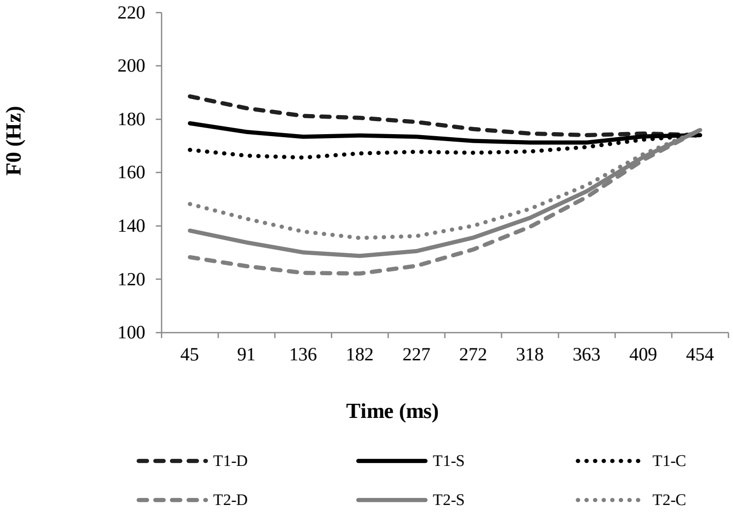


Fig. 2 T1-T2 continuum used in the test trials

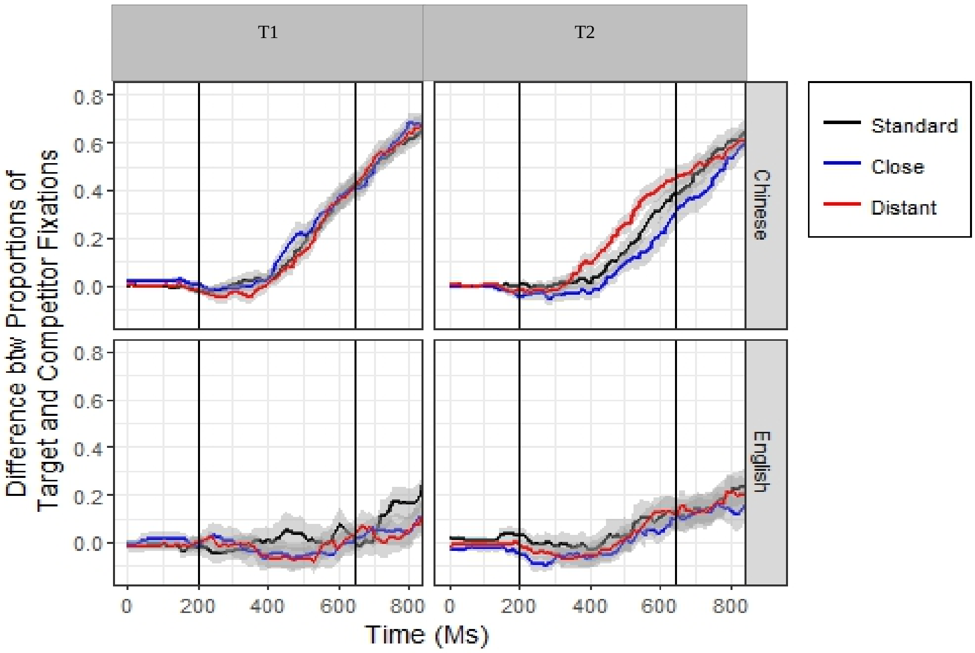


Fig. 3 Chinese (top) and English (bottom) listeners' differential proportions of fixations in the standard, close, and distant conditions for T1 and T2; the shaded area represents one standard error above and below the mean; the vertical bars represent the beginning and end of the time window used for the statistical analyses

References

[1] McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Gradient effects of within–category phonetic variation on lexical access, *Cognition*, *86*, B33–B42.

[2] McMurray, B., Tanenhaus, M., & Aslin, R. (2009). Within–category VOT affects recovery from "lexical" garden paths: Evidence against phoneme–level inhibition. *Journal of Memory and Language, 60*, 65–91.

[3] Xu, Y. S., Gandour, J., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of F0 direction. *Journal of the Acoustical Society of America*, *120*, 1063–1074.

[4] Gandour, J. T. (1983). Tone perception in far Eastern languages. *Journal of Phonetics*, *11*, 149–175.

[5] Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustic Society of America*, *32*, 451–454.

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

Variation in L2 phonological compensation with phonological rules and contexts

Gwanhi Yun

Daegu University
ghyun@daegu.ac.kr

It has been pointed out that L2 listeners have difficulty restoring the underlying phonemes and speakers' intended words because of the twisted surface forms caused by the application of phonological rules. Many previous studies showed that L2 proficiency, the incomplete application of phonological rules, or the presence of similar phonological rules in L1 affects the degree of phonological compensation [1, 2, 3].

Given this background, the present study aims to explore (i) to what extent the context where words that have undergone phonological rules influence L2 phonological rules, (ii) the effect of wordhood, (iii) the effect of absence/presence of phonological rules in L1, and (iv) the effect of homophony [4, 5].

Word identification tests were conducted with English speakers of L2 Korean and three Korean phonological rules (palatalization, nasalization, and lateralization). First, in order to tease apart potential influences of the trace of the underlying phonemes, words that have undergone complete phonological rules were employed as listening stimuli. Second, to see the effect of L1 phonological rules, two types of Korean rules were chosen: palatalization rule existing both in L1 English and L2 Korean vs. lateralization and nasalization rules existing only in L2. Finally, to exclude the effect of L2 proficiency, only beginner-level L2 learners with 1 or 2 years of formal learning participated in the experiment. Three nine L2 Korean learners who are native speakers of English participated in the identification task where they were asked to choose what word they heard between words with underlying phonemes and words with twisted phonemes (e.g., listen to [cilli] and choose one between /cinli/ and /cilli/). Listening stimuli were presented with and without contexts, and the choice items were given in different orthography. The rates of responses for the word with underlying phonemes were subject to linear regression mixed effects of ANOVA with subjects as random effects and contexts and wordhood as fixed effects.

First, contextual effects on phonological compensation emerged differently depending on individual type of L2 phonological rules as evident in (2). Specifically, contexts where palatalized words are placed significantly enhanced phonological compensation whereas those containing nasalized and lateralized words did not.

Second, the effect of wordhood as well occurred differently in accordance with phonological rule type as seen in (3). Perceptual recoverability of underlying phonemes substantially increased for real words over for nonce words on the basis of palatalized surface forms. However, such an enhancing effect of wordhood did not occur for nasalized or lateralized surface forms.

Third, unlike our expectation, the positive transfer of L1 phonological rules was not robust for L2 perceptual compensation. The degree of restoration of underlying phonemes was lower for palatalized forms than for lateralized or nasalized ones (39% vs. 49% vs. 55%). This suggests that English palatalization did not facilitate phonological inferencing for Korean palatalized words more greatly than for Korean lateralized or nasalized words.

Last, we obtained interesting findings for the effect of homophony on phonological compensation. When listeners heard homophonous stimuli with no contexts (e.g., [kači] for /kathi/ 'together' and /kači/ 'value'), the underlying words were activated and restored less successfully for palatalized words than when they heard non-homophonous stimuli (31% vs. 37%). However, when the identical stimuli were given with contexts, the former was recovered more accurately than the latter (55% vs. 45%).

Our findings contribute to revealing novel and more comprehensive nature of L2 listeners' phonological compensation. First, this study contributes to revealing that contextual enhancement might vary in accordance with types of L2 phonological rules. Second, the status as real words does not necessarily facilitate perceptual compensation, implying that the mechanism of phonological backward inferencing might depend on the internalization of L2 phonological rules [6]. Third, it is suggested that the existence of L1 phonological rules does not necessarily positively assist the restoration of L2 words which have undergone similar L2 phonological rules. Finally, contexts assist L2 listeners to activate L2 homophonous words positively whereas L2 listeners greatly suffer from recovering the speakers' intended words without contexts.

(1) Listening stimuli (real words vs. nonce words)

Real words

a. palatalized words	b. nasalized word	c. lateralized words
/pat^{h}+i/ [pači] 'field+NOM'	/akma/ [aŋma] 'devil'	/cinli/ [cilli] 'truth'
/kut+i/ [kuǰi] 'willingly'	/nat^{h}mal/ [nammal] 'word'	/nɔnli/ [nɔnlli] 'logic'
/kət+i/ [kəči] 'surface+NOM'	/papmat/ [pamm$\underline{\text{a}}$t] 'rice taste'	/nanli/ [nalli] 'disaster'

Nonce words

[uči]	[innu]	[sinnat]
[əči]	[amnu]	[kanni]

(2) Context effects (% correct)

	Palatalized words	Nasalized words	Lateralized words	Total
No context	29	52	35	39
Context	49	54	35	46
	F[1,6822]=325, p<.0001*	*F*[1,896]=-1.85, p>.05	*F*[1,740]=.00, p>.05	

(3) Wordhood effects (% correct)

	Palatalized words	Nasalized words	Lateralized words	Total
Nonce words	34	55	50	46
Real words	43	55	50	49
	F[1,6822]=23.7, p<.0001*	*F*[1,1832]=.05, p>.05	*F*[1,2300]=.02, p>.05	

(4) Context and Homophony effects (% correct) for palatalized words

	Non-homophony	Homophony	
No Context	37	31	*F*[1,1754]=5.87, p=.01*
Context	42	55	*F*[1,2378]=44.45, p<.0001*

References

[1] Gow, D. W., Jr., and Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language* 51,279-296

[2] Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes* 8, 291-309.

[3] Mitterer, H., and Ernestus, M. (2006). Listeners recover /t/s that speakers reduce. Evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34, 73-103.

[4] Darcy, I., Perperkamp, S., and Dupox, E. (2007). Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9*, 411-442.

[5] Dilley, L, and Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America* 122, 2340-2353.

[6] Gaskell, M. G., and Marslen-Wilson, W. D. (1998). Mechanism of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24, 380-396.

Temporal Organization of the Prenuclear Glides in Taiwanese Southern Min*

Feng-fan Hsieh & Yueh-chin Chang

National Tsing Hua University (Taiwan)
ffhsieh@mx.nthu.edu.tw, ycchang@mx.nthu.edu.tw

This study examines the temporal organization of the prenuclear glides /j, w/ within a syllable (a.k.a. the "medial" in Chinese traditional phonology; see, e.g., [1] for a recent survey of previous studies). In addition, the experimental results are compared with the coordination patterns of the prenuclear glides /j, w/ in Standard Chinese (SC) as an effort to develop a typology of prenuclear glides in Sinitic languages. Kinematic data were collected from four Taiwanese Southern Min (TSM) speakers (one female), using Electromagnetic Articulography (NDI Wave).

Following [2, 3], among others, the gesture timing of the prenuclear glides was compared in CV, CGV, and GV contexts to investigate the phonetic expression of (sub-)syllabic organization. More specifically, mean relative timing of C/G gestures to a (heterosyllabic) consonantal anchor /l/ were calculated using the lag between achievement of constriction of the C/G and /l/. By comparing a given subject's mean lags for Cs/Gs with CG sequences, we calculated the subject's leftward shifts (C vs. CG: *C-shift*) and rightward shifts (G vs. (C)G: *G-shift*) associated with /pj-/, /pw-/, /kj-/ and /kw-/. The analytical heuristics are schematized in (1). Some discussion is in order. In (1a), the articulatory targets of the C/G gestures among the three syllable types occur simultaneously, hence "No shift." This pattern can be interpreted as secondary articulation (e.g., labialized /k/) because CV, CGV and GV are indistinguishable regarding the intervals between the targets of the onset consonants/glides and a common anchor point (C(onsonant)-lag or G(lide)-lag). The C-center effect in (1b) is a well-established coordination pattern for "genuine" consonant clusters (e.g., *pr-* or *sk-* in English). As for (1c), one possible interpretation is that C in CGV may be "extrasyllabic," since only C-lag is significantly longer in duration ([CGV]>[CV]; *p<.05*). Finally, (1d) means that the glide is pushed rightward significantly ([GV]>C[GV]; *p<.05*) as a result of the addition of C in the context of CGV.

Each syllable type (CV, CGV and GV) is embedded in a carrier phrase and is repeated ten times in a randomized order. Data are post-processed and analyzed with the help of MView ([4]).

The results in (2) seem to suggest that there is *no* consistent (sub-)syllabic organization for the prenuclear glides in TSM. In other words, it may not be the case that the prenuclear glides are invariably "secondary articulations" across (Sinitic) languages (e.g., [5], et seq.). As we can see, the "N.S/No shift" pattern in (1a), or, the articulatory timing pattern for secondary articulations, is outnumbered. Nevertheless, some general tendencies can still be observed, even though we find considerable inter-speaker variation in (2). Firstly, C-shift-only (1c) and G-shift-only (1d) *do not* co-occur in the results reported in (2). In other words, the relative temporal stability across all speakers can be, to some extent, warranted, as the C-shift-only vs. G-shift-only patterns are presumably "incompatible" with each other if these patterns are found within the same syllables among different speakers. Precisely, it is unlikely that for some speakers, [wa] = k[wa], while for some other speakers, [wa] = [kwa]. That would mean that the glide /w/ has two mutually conflicting impacts on articulatory timing: /w/ pushes /k/ more leftward if [wa] = k[wa], while /k/ pushes /w/ more rightward if [wa] = [kwa]. Put differently, /w/ does not seem to have a fixed role in this scenario, which is predicted to be impossible under the assumption that the phonetic expression of (sub-)syllabic organization is supposed to be more or less identical across speakers (see, e.g., [3], among others). As a matter of fact, it is remarkable that the results from the two younger speakers (M3 and F1) are identical in (2). Secondly, and more importantly, the G-shift pattern in (1d) is most frequently attested in (2), while the C-shift pattern in (1c) is mostly found among the triad *po-pio-io* produced by the two younger speakers (M3 and F1). Taken together, the results, again, suggest

* This research is supported by an International Scholarly Exchange Research Grant (#9500023/RG003-P-14) from the Chiang Ching-kuo Foundation, Taiwan and MOST 103-2410-H-007-036-MY3.

that "subsyllabic affiliation" of the prenuclear glides are not always the same because different shift patterns are indeed attested. It is also worth mentioning that places of articulation of the consonant onset plays a role (here, labial vs. velar) in articulatory timing.

Comparison with Standard Chinese: The results in (2) are compared with those found in Standard Chinese (SC) under a similar experimental setting in all conditions. The speakers are in their 20s and are from Northern or Northeastern China (i.e., Beijing, Hebei, or Heilongjiang). Each syllable type (CV, CGV and GV) was embedded in a carrier phrase and was repeated ten times. The results are given in (3). We can see from (3) that SC does not pattern alike with TSM as far as the temporal organization of the prenuclear glides /j, w/ is concerned. Firstly, a "new" pattern, i.e., the C-center effect, is attested for the triad *ka-kwa-wa* produced by F1 and F4 (note that C-center is *not* attested in TSM in (2)). Secondly, and more importantly, it appears that there is no C-shift-only pattern for CGV /*pje, tja, kwa*/, while the C-shift-only pattern is attested among /*kwej*/ only, suggesting that, again, different places of articulation (here, coronal vs. velar) have a bearing on the articulatory phasing of CG, as well as syllable composition (i.e., CGV vs. CGVG).

In this study, we have shown that the prenuclear glides /j, w/ are not "created equal," at least in terms of articulatory timing patterns. There is no consistent timing pattern across the board, as evidenced in (2) and (3), while inter-speaker variation is attested (but notice again that is not unexpected since different structures are not contrasting in meaning, e.g., consonant cluster vs. secondary articulation: [pj] vs. $[p^j]$ in Russian). Finally, time permitting, we will also discuss implications of these results for segmental phonology in Sinitic languages.

(1) a. No shift	b. C-center	c. C-shift only	d. G-shift only
[kwa] = [ka] [wa] = k[wa]	[kwa] > [ka] [wa] > k[wa]	[kwa] > [ka] [wa] = k[wa]	[kwa] = [ka] [wa] > k[wa]

(where [ ] = lag between C/G gestures to a common anchor point; >/< means longer/shorter in duration significantly (*p*<.05))

(2) TSM	*pwe*	*pjo*	*kwa*	*kwe*	*kwi*	*kja*	*kju*	*kjo*
M1	G-shift	N.S.	C-shift	G-shift	N.S.	G-shift	N.S.	G-shift
M2	G-shift	N.S.	N.S.	N.S.	G-shift	G-shift	G-shift	G-shift
M3$_{\text{Young}}$	G-shift	C-shift	N.S.	N.S.	G-shift	G-shift	G-shift	G-shift
F1$_{\text{Young}}$	G-shift	C-shift	N.S.	N.S.	G-shift	G-shift	G-shift	G-shift

(where N.S. = No shift in (1a); Young=speaker in their 20s, others in their 40s.)

(3) Standard Chinese	*pje*	*tja*	*kwa*	*tjow*	*twej*	*kwej*
M1 (Beijing)	N.S.	N.S.	G-shift	G-shift	G-shift	C-shift
M2 (Heilongjiang)	N.S.	G-shift	G-shift	G-shift	N.S.	N.S.
F1 (Beijing)	G-shift	N.S.	C-center	N/A	G-shift	N.S.
F2 (Heilongjiang)	G-shift	N.S.	N.S.	N.S.	G-shift	C-shift
F3 (Hebei)	N.S.	N.S.	G-shift	G-shift	C-center	C-shift
F4 (Heilongjiang	G-shift	N/A	C-center	N/A	N/A	N/A

(where N/A means "no or insufficient tokens are available.")

References

[1] Myers, J. (2015). Stuck in the middle: Mandarin medials in articulation, parsing, and association. In Y. E. Hsiao & L.-H. Wee (Eds.) *Capturing Phonological Shades within and across Languages* (pp. 101-119). Cambridge, UK: Cambridge Scholars Publishing.

[2] Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in Phonetics and Speech Science*, 239-250.

[3] Shaw, J.A., Gafos, A.I., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology* 28:3: 455-290.

[4] Tiede, Mark. (2010). MVIEW: Multi-channel visualization application for displaying dynamic sensor movements. In development.

[5] Duanmu, S. (2007). *The Phonology of Standard Chinese* (2nd ed.). Oxford: Oxford University Press.

Articulatory Strategies to Mark Prominence in Consonants

Tomas O. Lentz[1] & Hayo R. Terband[2]

[1]University of Amsterdam (Netherlands), [2]Utrecht University (Netherlands)
lentz@uva.nl, h.r.terband@uu.nl

Introduction Stressed or pragmatically prominent syllables are typically produced with longer durations and a more careful articulation. For instance, stressed vowels in English are not reduced to schwas, while unstressed vowels often are. In Dutch, unstressed vowels are not always reduced to a schwa, but still produced with a different spectral tilt, i.e. relatively less energy in the higher frequency bands [1,2,3]. Articulatory strategies to mark pragmatic prominence on vowels have been identified, but seem to vary between speakers [4]. One way is to move the tongue to a more extreme position, the other is to increase the sonority contrast in more prominent syllables, by making consonants less and vowel more sonorous [5].

However, vowel and consonants might not be subject to prominence marking in the same way. Consonants could be shortened to increase the sonority contrast within a syllable, but they could also be articulated at more extreme positions to protect their identifiability. If a consonant has to be articulated in a shorter time with the articulators still reaching the target position, movement velocity has to increase. For lingual coda consonants, there might be a need to compensate for more extreme tongue positions reached during the preceding vowel articulation. If a speaker does not speed up the articulator movements or lengthens the consonant, she should accept loss of quality, such as lenition or coarticulation. Hence, multiple predictions on the effect of prominence on consonants can be made and different speakers might apply different strategies. In addition, prominence derived from lexical stress and prominence from pragmatics have effects on articulation [6,7], but these two ‘prominences’ might be marked differently.

Given the myriad of possible effects of prominence on consonant articulation, the present study features an in-depth exploratory analysis of a subset (4 pp) of Electromagnetic Articulography (EMA) recordings of 40 speakers of Dutch. Our goal is to formulate hypotheses on possible strategies employed to mark prominence in consonant articulation that can be tested on the remaining data in a later stage.

Method The study focussed on coda consonants, because 1) they might be more susceptible to prosody-driven lengthening than onset consonants [5], 2) they are susceptible to coarticulation in Dutch and 3) they allowed a relatively elegant elicitation of lexical and pragmatic prominence. We analysed /n/ and /t/ in coda position of morphemes ‘in’ (/ɪn/) and ‘uit’ (/œyt/) before a labial. Stress was manipulated by placing the morpheme in a compound verb (/'ɪnpɑkə/, *to pack: /'ɪnbakə/, to season:*) or as preposition in a phonemically identical phrase (/ɪn 'pɑkə/, *in packages*). The /n/ is more sensitive to coarticulation with the labial to [m] than the /t/. The morpheme was made pragmatically prominent to three degrees by embedding it in a given sentence that served as an answer to a given question. Pragmatic prominence levels varied between low (after a different phrase with narrow focus), neutral (the whole sentence or phrase contains new information), and high (the morpheme alone corrects an assumption).

Tongue tip (for /n,t/) and labial articulatory movements (distance upper-lower lip, for /b,p/) were segmented using Mview, courtesy Mark Tiede. Syllable and vowel durations were extracted from the acoustic signal. A linear mixed model was fitted for all dependent variables, with lexical stress, pragmatic prominence as independent variables and participant and item as random effects.

Results Due to The tongue was less front with low focus (p = 0.03); for the /t/, the tongue was lower for stressed syllables (p < 0.05). The distance the articulator travels depends on prominence, interacting with the consonant (p < 0.05*). For /n/, the tongue moved less for high focus compared to low (B = -4.4mm, p = 0.009**). For the /t/, stress and focus interact (p = 0.03*). For stressed /t/, the tongue moves less (B = -3.9mm, p = 0.004**). Focus prominence now had a positive influence, with more distance for high focus than low focus consonants (B = 3.9mm, p = 0.019*), but only on the stressed syllables (B = 6.9mm, p = 0.004**). For onset velocity, there was a significant interaction between phoneme and stress (*p* = 0.04*); only for /t/ the unstressed consonants are articulated significantly slower (*B* = -3.5 cm/s, *p* = 0.02*). No effect for entry velocity peak or focus was found, neither were overlap, lags, exit velocities or articulatory durations significantly affected by any prominence level. Stressed vowels seem slightly longer acoustically, but not significantly.

Discussion As no evidence was found that prominence affects articulatory or acoustic duration of consonants, we remain agnostic on the possibility that sonority is increased by reducing consonant duration as well as the possibility that stress or focus are expressed by syllable lengthening. However, consonants might be (only) proportionally shorter, if the vowels are longer when more prominent; we did not find an effect on vowel length now, but it might found in the full data set. If so, phoneme identifiability can be preserved, while the sonority contrast is still enhanced.

Consonants were found to be produced at more extreme positions when pragmatic prominence was higher, but the speakers did not consistently make larger articulatory movements to achieve this. We hypothesise the following general strategy exists: when a prominent consonant is realised, the tongue might be positioned for it during the preceding vowel, suggesting that carryover coarticulation can follow from anticipatory planning. The resulting sacrifice of precision of the vowel articulation is, however, only made if needed; we hypothesise the /t/ needs less protection (probably because it is a stop, but lexical neighbourhood size happens to be a confound). If the consonant needs more protection, vowel and consonant are articulated at more extreme positions, causing an increase in movement.

We cannot yet identify one consistent strategy that is completely supported by the data. However, our exploration does allow for a focussed use of the larger dataset for hypothesis testing. In addition, we plan to not only analyse the group-wide strategy for prominence articulation in coda consonants, but also to identify participant-specific strategies.

References

[1] Rietveld, A. C. M. & Koopmans-van Beinum, F. J. (1987) Vowel reduction and stress. *Speech Communication*, 6, 217–229.

[2] Sluijter, A. M. C. & van Heuven, V. J. (1995) Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. *Phonetica,* 52,71–89.

[3] Sluijter, A. M. C., van Heuven, V. J. & Pacilly, J. J. A. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America*, 101(1)

[4] Harrington, J., Fletcher, J., & Beckman, M. (2000). Manner and place conflicts in the articulation of accent in Australian English. In Broe, M. B. & Pierrehumbert, J. B. (Eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon*, 40–51. Cambridge, UK: Cambridge University Press.

[5] Beckman, M., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In Docherty, G. J. & Ladd, D. R. (Eds)., *Papers in Laboratory Phonology II: Gesture, segment, prosody*, 68–89. Cambridge, UK: Cambridge University Press.

[6] Turk, A. (2012). The temporal implementation of prosodic structure. In Cohn, A.C, Fougeron, C., & Huffman, M. K. (Eds.), *The Oxford Handbook of Laboratory Phonology,* 242–253. Oxford, UK: Oxford University Press.

[7] Beckman, M. E. & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In Keating, P.A. (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, 7-33. Cambridge, UK: Cambridge University Press.

Context dependent voicing characteristics of the Hungarian /h/

Andrea DEME[1,2], Márton BARTÓK[1,2], Tekla Etelka GRÁCZI[3,2], Tamás Gábor CSAPÓ[4,2] & Alexandra MARKÓ[1,2]

[1]Eötvös Loránd University, Hungary; [2]MTA-ELTE Lendület Lingual Articulation Research Group, Hungary; [3]Research Institute for Linguistics HAS, Hungary; [4]Budapest University of Technology and Economics, Hungary

deme.andrea@btk.elte.hu

In Hungarian phonology, /h/ is most commonly mentioned in connection with its asymmetric behaviour in voicing assimilation: it triggers regressive devoicing, but does not undergo voicing before voiced obstruents [1]. It is suggested that its allophonic voiceless-voiced alternation is clear-cut; /h/ is claimed to surface invariantly as [ɦ] intervocalically [1] (hereafter, in V*h*V). There is only one phonologically conditioned exception from this rule: if the containing syllable bears an accent (hereafter, V#*h*V), /h/ is claimed not to be voiced [2]. In phonetics, the place of articulation, and the voicing characteristics of /h/ go back to a long debate [3]. With respect to voicing, an apparent agreement is reached, as, in line with phonology, phonetic textbooks from the last decades unanimously asserted that Hungarian /h/ is (breathy) voiced in V*h*V [3][4][5], as well as in all sonorant-/h/-vowel (Son*h*V) contexts [3][5]. Empirical research, however, has only partially addressed and corroborated the above suggestions. Through the analysis of /h/ in unsystematically varying contexts, [6] showed that /h/ is very likely to undergo voicing in V*h*V, but it is more so, if it is produced between front vowels. More recently, however, using controlled phonetic contexts we concluded that there is no clear effect of vowel backness, and openness on the ratio of the voiced part (RVP) in /h/, and that all V*h*V contexts trigger voicing to a large extent. We also found that the voicing of /h/ is not affected by the presence/absence of pitch-accent either, since the RVP was equally high both in V*h*V, and in V#*h*V (*h*V bearing a pitch accent), while in the baseline (post-pausal word onsets, #*h*V), the amount of voicing was significantly lower. Finally, we could draw the conclusion that even though the RVP measure uniformly reflected voicing in all V*h*V contexts, the fine-grained phonetic details, i.e., the quality of voicing differed remarkably as a function of vowels. The latter finding indicates that the vertical position of the larynx which differs across vowels also effects the larynx position in the enclosed /h/, and thus the (breathy) quality of the /h/ is affected accordingly (for this analysis we obtained the harmonics-to-noise ratio, HNR) [7]. The questions, however, are still open, (1) if Son*h*V contexts also facilitate the voicing of /h/ to the same extent as V*h*V contexts do, and (2) if there is a difference in voice quality as a function of contexts (vocalic vs. sonorant vs. obstruent, and their place of articulation feature).

In an attempt to answer the first question, we studied the RVP of onset /h/s in the following conditions: (i) V_1hV_1, (ii), $V_1\#hV_1$, (iii) $\#hV_1$, (iv) $\mathrm{Son}hV_1$ and (v) $\mathrm{Plos}hV_1$, where ‘Plos’ refers to /p t c/ stop consonants matched with the /m r ɲ/ sonorant ‘Son’ consonants by the place of articulation. V was systematically assigned the qualities /u ɒ i ɛ/, and $\#hV_1$ condition served as baseline (where /h/ is predicted to bear a glottal place of articulation and no voicing). Hence, in this study we retested the effect of vowel quality features, the presence of pitch-accent, and we additionally examined the effect of the Son*h*V contexts (compared to matching voiceless consonant contexts) for the first time in Hungarian phonetics. In an attempt to answer the second question, we estimated and compared several acoustic measures suggested for the study of voicing in fricatives [9]: center of gravity (COG), harmonics-to-noise ratio (HNR), the fricative-to-right-vowel-intensity ratio (FTR), and the low-frequency-intensity-to-total-intensity ratio (LFT). Stimuli were recorded from 21 female speakers (5 repetitions each, 180 tokens per speaker in total). RVP was measured by the help of Praat’s voice report (VR) [8], on the basis of a pre-test, which showed high congruence of VR with manual segmentation [7].

Preliminary analysis of 12 speakers’ RVP and HNR data revealed a 3-way interaction effect for both measures: RVP ($F(4,44) = 2.78$, $p < 0.05$) (Fig 1); HNR ($F(4,44) = 5.15$, $p < 0.05$) (Fig 2) (3-way repeated measures ANOVA; factors: *condition, vowel height, vowel backness*). Despite the found interaction, the data clearly indicate that the main factor affecting the voicing of /h/ is the condition: intervocalic and Son*h*V positions pattern together, and are clearly distinguished from the baseline condition (which, on the other hand, patterns with the plosive context, as expected) (Fig 1). This finding reflects the more categorical nature of the process which induces intervocalic voicing of /h/. However, the fine-grained acoustic measures refine this picture, as they indicate noisier realizations in the context of plosives than in the baseline, and more sonorous realizations in the context of high vowels than in low vowels (Fig 2). These findings suggest that the allophonic occurrence of [ɦ] is phonetically motivated and that the [h] – [ɦ] distinction is realized in a rather categorical fashion, but the quality of voicing in [ɦ] exhibits fine subphonemic differences as a function of contexts. Further phonetic data and their analysis will also be presented in detail.

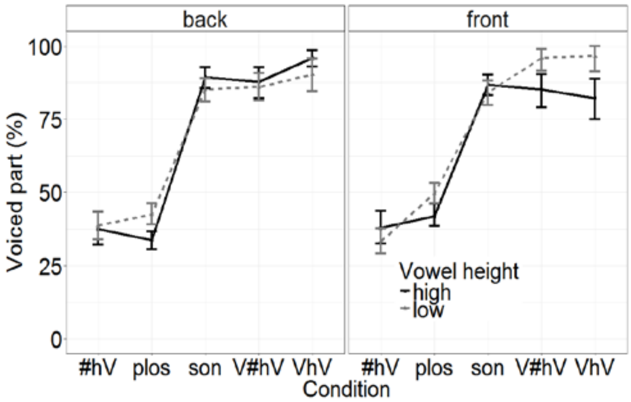


Fig.1 Ratio of voiced part in /h/ as a function of vowel backness and vowel height

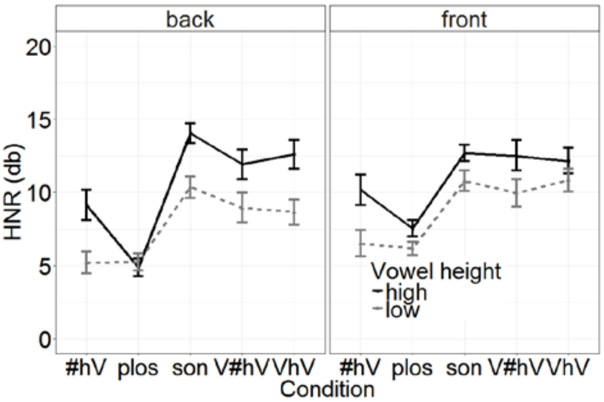


Fig.2 HNR for /h/ as a function of vowel backness and vowel height

References

[1] Siptár, P. & Törkenczy, M. 2007. *The Phonology of Hungarian.* OUP, New York.

[2] Siptár P. 1994. A mássalhangzók. In Kiefer Ferenc (ed.): *Strukturális magyar nyelvtan 2. Fonológia.* Akadémiai Kiadó, Budapest. 183–272.

[3] Laziczius Gy. 1963/1979. *Fonetika.* Nemzeti Tankönyvkiadó, Budapest.

[4] Bolla K. 1995. *Magyar fonetikai atlasz.* Nemzeti Tankönyvkiadó, Budapest.

[5] Kassai I. 1998. *Fonetika.* Nemzeti Tankönyvkiadó, Budapest.

[6] Gósy M. 2005. *A /h/ zöngésedése két magánhangzó között.* Beszédkutatás 5–20.

[7] Deme, A., Bartók, M., Gráczi, T. E. Markó, A., Varjasi, G., Csapó, T. G. 2017. Intervocalic voicing of the Hungarian /h/, presentation given at the *13th ICSH*, 29-30 June, 2017.

[8] Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4.02, retrieved 15 December 2014 from http://www.praat.org/

[9] Gradoville, M. 2011. Validity in Measurements of Fricative Voicing: Evidence from Argentine Spanish. In S. M. Alvord (Ed.), *Selected Proc. of the 5th LARP Conference* 59–74.

Aerodynamics and Laryngeal Features of Contrast

Luis M. T. Jesus and Maria C. Costa

University of Aveiro, Portugal
lmtj@ua.pt, lopescosta@ua.pt

Stop voicing is phonologically active in some languages and in others, it is passive [1]. A clear relation between phonetic cues and phonological features that support this has yet to be found. Some measures of voicing category, designed for use with the acoustic signal, have been shown to be of limited use "since periodicity in the oral airflow continues even when associated only weakly with acoustic excitation of the vocal tract" [2, p. 634]. Studies such as ours, based on new aerodynamic data that is more closely related to laryngeal behaviour, could contribute toward clarifying these issues.

Oral airflow and electroglottographic (EGG) recordings of four Portuguese speakers, producing a corpus of nine isolated words with /b, d, g/ in initial, medial and final word position, and the same nine words embedded in 39 different real sentences, were collected. Recordings were made in a quiet room using an oro-nasal circumferentially vented mask and an EGG processor, connected via an audio interface to a notebook.

Slope of the stop release (SLP), Voice Onset Time (VOT), release and stop durations, steady state oral airflow amplitude characteristics preceding and following the stop, were analysed. Differences between independent groups and correlations between variables were studied; generalised linear mixed effects models were developed. A classification of stop's voicing was automatically extracted.

All stops presented periodicity in the oral airflow waveform during closure, i.e., according to our data, there was only one voicing distribution/shape [3, p. 42]: Continuous (weak) voicing throughout the whole closure. Therefore, all stops were produced without complete closure, i.e, there was leakage in the closure.

Men's release slopes were significantly steeper than female's, and both SLP and VOT were significantly different for the three places of articulation.

The VOT was significantly correlated to the relative amplitude of oral airflow (A12 – relative preceding vowel to stop amplitude), more specifically, a longer VOT resulted in lower A12 values, meaning that those stops presented stronger voicing than those with shorter VOT values. Shorter releases (RLS) resulted in steeper slopes (SLP), which were also significantly correlated to higher relative oral airflow values.

Steeper slopes (SLP) were correlated to shorter releases (RLS) in bilabial and dental stops, and to shorter (STP) dental stops. A particularly striking (high correlation coefficient values at a 0.01 significance level) result was that for STP/VOT correlations: Longer stop durations clearly resulted in shorter VOT values, for all places of articulation. Higher A12 values were correlated to shorter dental and velar stops VOT and MOA/VOT were significantly (negatively) correlated, for velar stops. Finally, significant positive correlations were found between STP and RLS of dental and velar stops.

Results of voicing classification, based on the mean relative amplitude of the oral airflow signal [2, p. 635], revealed that 57% of stops were weakly voiced (/b/: 62%; /d/: 60%; /g/: 47%). The mean relative amplitudes of the oral flow were significantly different between genders.

Generalised linear mixed effects models were used to test for the fixed effects of VOT, SLP and the factors Place of Articulation [Bilabial; Dental; Velar], Gender [Male; Female] and Vowel Context [High; Low] (without interaction terms) on the mean oral airflow. By-speaker variation, considered as a random effect with random intercept, was found to explain a considerable part of the variability in mean oral airflow. No significant heteroscedasticity nor deviations from normality were found in the analysis of the residuals of the proposed model.

Results presented in this abstract provide new evidence toward the view that stop voicing is not phonologically active in European Portuguese (EP) as previously suggested for this language [4], German [1] and English [5]. Only around 40% of this study's stops were strongly voiced, a percentage which is even lower than what has been previously reported (around 60%) for German [1, p. 271] and English [5, p. 158]. The decrease in the amplitude of oral airflow (i.e., the amplitude of voicing) relative to the adjacent vowel (see Figure 1), observed in EP stops is very similar to what has been previously observed for German [1, p. 272], and suggests that the laryngeal feature of contrast in EP is not [voice], as recently suggested for German and English by Beckman et al. [1]. The laryngeal contrast in EP could be between stops with no laryngeal specification and those specified as [spread glottis], evidenced by the low oral airflow amplitude oscillations (shown in Figure 1) which are very likely generated by mucosal oscillations for a spread glottis [6, pp. 199–201].

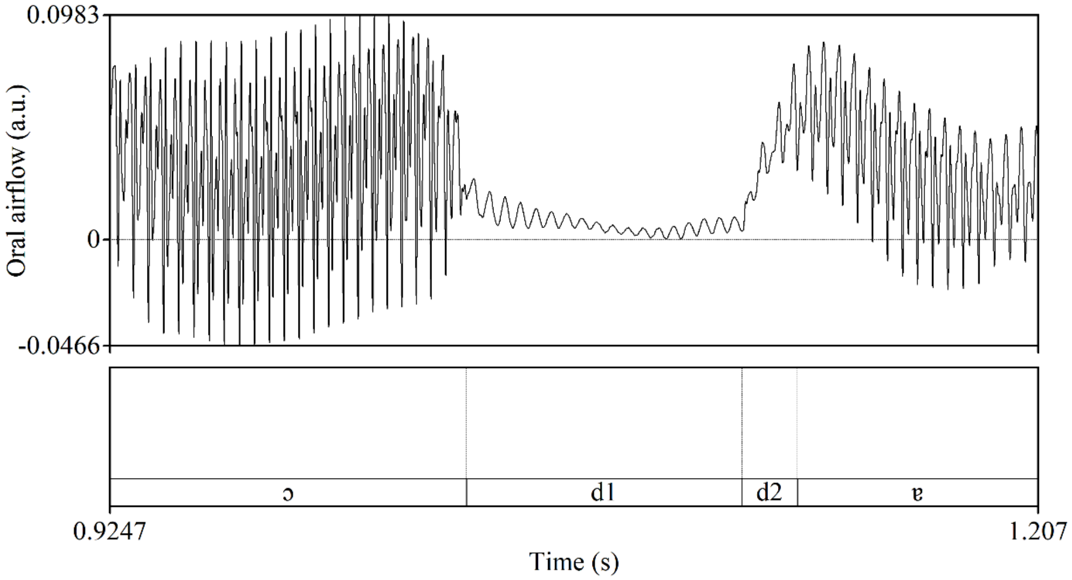


Fig. 1 Oral airflow and EGG signals for file HV065 (<Diga pode alucinar por favor> [ˈdi.gɐ ˈpɔd ɐ.lu.siˈnaɾ ˈpuɾ fɐˈvoɾ]). From top to bottom: Oral airflow waveform; annotation (phone [d] is divided into two intervals – closure (d1) and release (d2) – only the interval corresponding to the [ɔdɐ] phone sequence is shown).

References

[1] Beckman, J. Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, 49(2) 259–284.

[2] Pinho, C. M. R., Jesus, L. M. T., & Barney, A. (2012). Weak Voicing in Fricative Production. *Journal of Phonetics*, 40 (5), 625–638.

[3] Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, 54(1), 35–50.

[4] Pape D. & Jesus, L. M. T. (2015). Stop and Fricative Devoicing in European Portuguese, Italian and German. *Language and Speech*, 58(2), 224–246.

[5] Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents*. Berlin: Foris.

[6] Esposito, A. (2002). On vowel height and consonantal voicing effects: data from Italian. *Phonetica*, 59(4) 197–231.

Oral Presentations (Day 2)

Sensitivity of the N400 to the Phonetics but not the Phonology of Mandarin Tones

Stephen Politzer-Ahles[1], Jueyao Lin[1] & Lei Pan[1]

[1]*The Hong Kong Polytechnic University (Hong Kong)*
sjpolit@polyu.edu.hk, hansy.lin@polyu.edu.hk, bernice.pan@polyu.edu.hk

Introduction. The present study tested whether abstract linguistic knowledge modulates the neural response to sounds. It has often been observed that hearing a sound that doesn't match one's expectation (e.g., hearing the word "dog" after seeing a picture of a flower) elicits a negative-going event-related potential component (N400 or phonological mapping negativity), relative to hearing the same sound when it does match one's expectation (e.g., Malins and Joanisse [1], among others). We examine whether this mismatch response is attenuated when the sound heard has an abstract phonological relationship with the expected sound.

To do so, we focused on tone alternation in Mandarin Chinese. Mandarin has four lexical tones, but some of these change in contexts; particularly, Low tone changes into Rising tone in some contexts. For instance, the character 使 is normally pronounced *shi*^{L}, but in the compound word 使者 ("envoy", *shi*^{R} *zhe*^{L}) it is pronounced with a Rising tone. Thus, hearing a syllable with Rising tone (e.g., *shi*^{R}) when one expects it in Low tone (e.g., *shi*^{L}) might not engender a serious mismatch, since *shi*^{R} is a legal variant of *shi*^{L}. On the other hand, the reverse should not apply: hearing *shi*^{L} when *shi*^{R} is expected should engender a mismatch, because there is no context in standard Mandarin where a Rising tone can change to a Low tone, and thus *shi*^{L} is simply the wrong tone.

Methods. We recorded EEG from 29 native Mandarin speakers (of a planned 80) while they saw Chinese characters (each of which had only one possible pronunciation) and then heard Mandarin syllables and judged whether the syllable matched the character. The sound either matched the character (e.g., hearing *shi*^{L} after seeing a character pronounced *shi*^{L}, or hearing *shi*^{R} after seeing a character pronounced *shi*^{R}) or mismatched it by having a different tone. The mismatching tone was either phonologically unrelated to the expected tone (High or Falling tone, when Low or Rising tone was expected) or phonologically related (Low tone when Rising tone was expected, or vice versa). Each participant was exposed to 64 trials of each match type and 32 of each mismatch type, as well as fillers. We predicted that hearing a Rising tone when a Low tone is expected should attenuate the mismatch signal (compared to hearing a Rising tone when an unrelated tone is expected); on the other hand, hearing a Low tone when a Rising tone is expected should not yield as much attenuation (compared to hearing a Low tone when an unrelated tone is expected). We also recorded a total mismatch condition (e.g., hearing *hua*^{H} after seeing a character pronounced *shi*^{L}) as a manipulation check to ensure that this paradigm would elicit N400 effects.

Results and discussion. The results from analysis of 27 participants with sufficient artifact-free trials are shown in Figure 1. All mismatch conditions (phonologically-related and –unrelated tone mismatches, as well total mismatches) elicited robust N400s compared to the match condition. Contrary to our prediction, however, phonologically-related and phonologically-unrelated tone mismatches (the two black lines) did not substantially differ; phonological relationship did not appear to attenuate the N400 mismatch effect (such a trend was only observed on the posterior channels when hearing Low-tone targets, whereas the effect we predicted should have appeared when hearing Rising-tone targets). This suggests that either abstract phonological knowledge does not influence the matching process that generates the N400 (see also Nieuwland et al. [2]) or that it does not do so without a licensing context (given that the stimuli were monosyllables presented without any context that could trigger tone sandhi)

Another interesting point, separate from the initial predictions of the study, is the obvious difference between the response for total mismatch and the response for tone mismatch (whether phonologically related or phonologically unrelated). Malins and Joanisse (2012) found similar N400s for whole-syllable mismatch and tone-only mismatch, suggesting a tone mismatch is enough to make the whole syllable be considered "wrong", and that whole-syllable mismatch is not merely a sum of other mismatches. The present study found a qualitatively different pattern: whole-syllable mismatch appeared to elicit a P3a component followed by a stronger N400 component than tone-only mismatch, suggesting that these mismatches triggered different processes (possibly consistent with Sereno and Lee [3], who found that tone-only mismatch still yielded some priming relative to total-mismatch controls). As the only differences between these experiments were the task (Malins and Joanisse [1] used picture primes, whereas the present study used character primes) and the precise proportions of various conditions and fillers relative to the match condition, it is an open question why these studies yielded different results; however, it suggests that there may still be much we do not yet understand about the contribution of tone to syllable recognition.

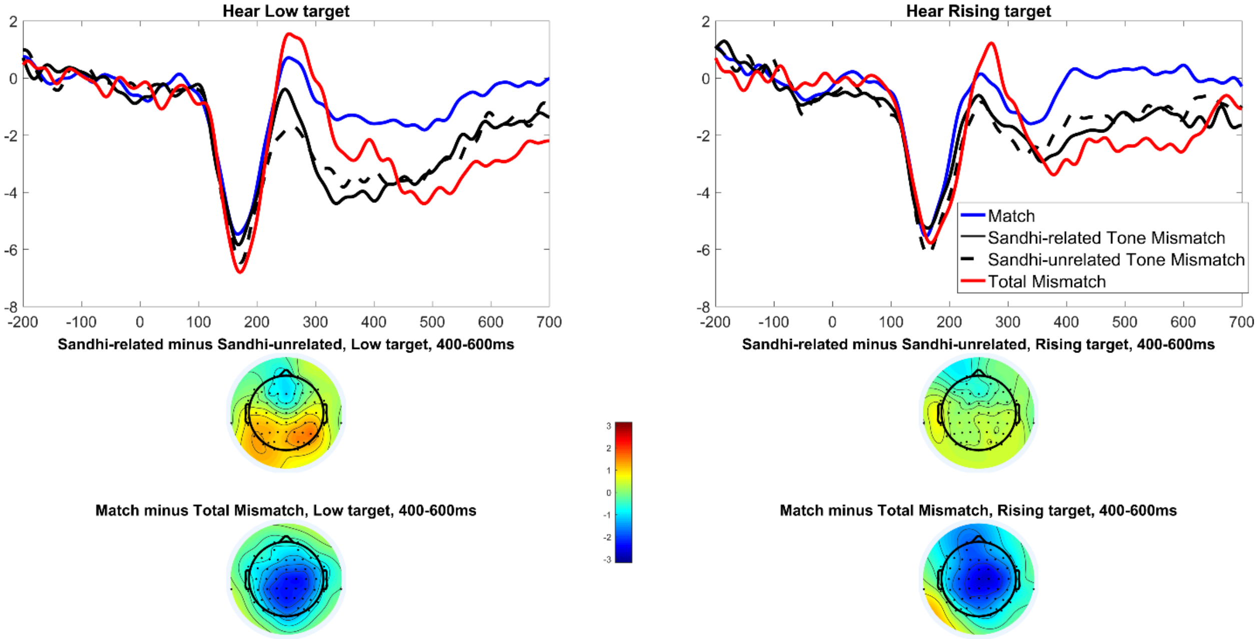


Fig.1 Waveforms and topographic plots when hearing Low tone (left) or Rising tone (right), as a function of how it relates to what was expected. Negative is plotted down.

References

[1] Malins, J., & Joanisse, M. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*, 2032-2043.

[2] Nieuwland, M., Politzer-Ahles, S., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., …, & Huettig, F. (in press). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*. *7*, e33468.

[3] Sereno, J., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, *58*, 131-151.

Contrastive Context Effects of Tone Modulated by Second Language Experience: The Case of Korean L2 Learners of Mandarin Chinese

Sang-Im Lee-Kim

National Chiao Tung University (Taiwan)
sangimleekim@nctu.edu.tw

Contrastive effects of extrinsic context are pervasive in speech perception. Tone language speakers identify a tone as high in a low-frequency context, while a physically identical tone is perceived as low in a high-frequency context (e.g. Moor & Jongman 1997, Wong & Diehl 2003, Francis, et al. 2006). Likewise, for an ambiguous stop spectrum, listeners tend to hear more [d]s (spectrally higher stop) following a spectrally lower liquid [ɹ] while [g] perception (spectrally lower stop) is more frequent following a spectrally higher liquid [l] (e.g. Mann 1980, Lotto & Kluender 1998, Kingston et al. 2014). Building upon the literature, the present study reports a novel case where a contrastive effect of neighbouring tones may be further integrated to give rise to a systematic asymmetry in stop perception: lenis stops (associated with low F0 at vowel onset) in a high-tone context vs. fortis stops (with high F0 at vowel onset) in a low-tone context by Korean listeners. By comparing Korean listeners with or without an experience with a tonal second language, Mandarin Chinese, we show that stop perception modulated by general extrinsic context effects is further shaped by an acquired sensitivity to F0 cues through learning of a tonal second language.

In a perception experiment, target stops based on Mandarin unaspirated stops (short-lag VOTs, unvoiced word-medially as well as word-initially) were preceded or followed by two tones to provide extrinsic contexts: word-medial stops preceded by low (e.g. **ma^{21}.**ta^{33}) or high (e.g. **ma^{55}**.ta^{33}) contexts and word-initial stops followed by low (e.g. ta^{33}.**ma^{21}**) or high (e.g. ta^{33}.**ma^{55}**) contexts. F0 frequency at vowel onset was digitally manipulated to range between 160 and 280 Hz on a 7-step continuum for a female voice based on two contours: Level (e.g. ma^{21}.**ta^{33}**) or Rising (e.g. ma^{21}.**ta^{35}**). The original tone contours were retained by raising or lowering the entire F0 contour so that the effects of onset F0 and tone contour can be assessed independently. In forced-choice identification tasks, 15 experienced learners and 14 naïve listeners judged unaspirated stop stimuli as either lenis ([ma**t**a]) or fortis ([ma**t'**a]) presented in Korean orthography.

The results summarized in Figure 1 demonstrate similarities and differences between the two groups. For word-medial stops with a preceding context (Fig 1, left), mixed-effects logistic regression analyses revealed a main effect of Context (p <.001) with no significant interaction with Group, indicating that the two groups behaved similarly for extrinsic context: more lenis judgments in a high-tone context than in a low-tone context. Furthermore, a significant Context*Contour interaction was found for both groups, suggesting such extrinsic context effect was greater for a level than for a rising tone regardless of L2 experience. However, an interesting group difference was captured by an OnsetF0*Group interaction (p <.001): fortis judgments were overall more frequent as onset F0 increased, but this pattern was pronounced more clearly for the learners (i.e. stiffer slops). For word-initial stops with a following context (Fig 1, right), however, similar context effects were observed only for the learners, namely more lenis responses in the high-tone context (p <.001) and larger context effects in the level tone condition (p <.05). As in the preceding context condition, learners relied on onset F0 cues more heavily in stop identification than naïve listeners did.

The results of the present study show the ways in which various aspects of tones exert complex but systematic influences on the perception of stops. Some patterns can be attributed to domain-general properties of cognitive processes. Specifically, both learners and naïve Korean listeners showed large contrastive effects of extrinsic tones such that following a high context a tone is perceived lower which, in turn, provides positive evidence for lenis stops. In addition, such effect was attested more clearly for the level tone than for the contour tone, suggesting that simple intrinsic F0 characteristics of a level tone have triggered greater reliance on extrinsic tonal properties (Moore & Jongman 1997). On top of those global effects of extrinsic context, however, it was shown that minor details of stop perception were further modulated by specific linguistic experiences. In particular, while naïve listeners incorporated onset F0 cues into stop perception to some extent as indicated by positive slopes across all conditions, the learning experience of a tonal second language seems to have enabled greater cue-weighting on F0 cues for stop identification as shown by stiffer slops for all cases under comparison. Furthermore, the learners' acquired sensitivity to F0 cues seems to be so pervasive to trigger long-distance contrastive effects of the following context for initial stop identification, a presumably unfavorable environment for context effects to arise.

Taken together, the results shed light on the role of learning in contrastive effects of context reported in past research (Mitterer 2006, Kang et al. 2016). Extrinsic context effects of tone were shown to be contrastive in the auditory domain reflecting broad cognitive biases, but generalizations arising from a particular linguistic experience, e.g. cue-weighting strategies for stops, may further refine the ways in which stops are perceived.

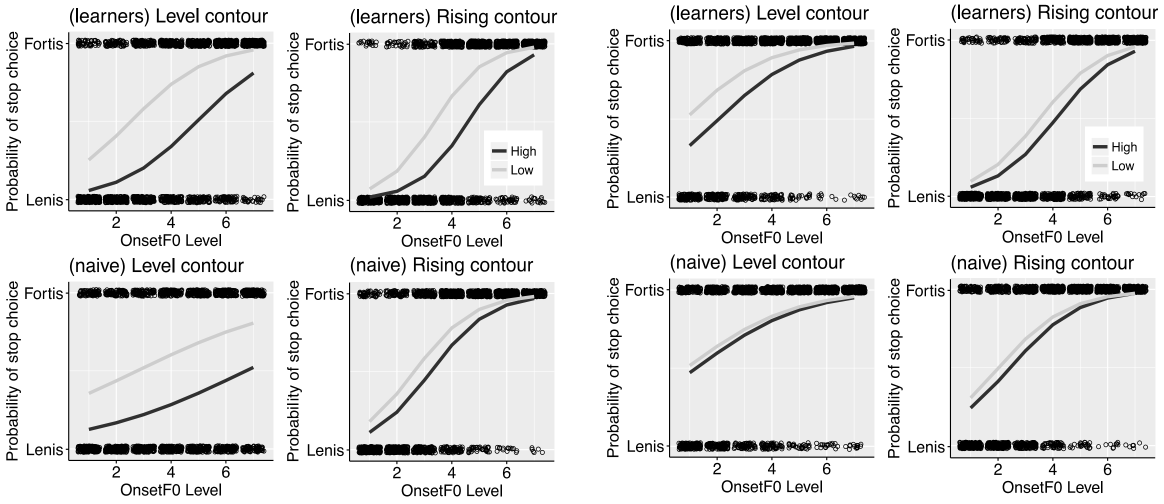


Fig. 1 Predicted logit-curves for the identification of word-medial (left) and word-initial (right) stops by experienced (top) and naïve listeners (bottom). Context tones varied between high (black) and low (gray).

References

[1] Moore, C. & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102, 1864–1877.

[2] Wong, P. C. M. & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, 46, 413–421.

[3] Francis, A., Ciocca, V., Wong, N., Leung, W. & Chu, P. (2006). Extrinsic context affects perceptual normalization of lexical tone. *Journal of the Acoustical Society of America*, 119(3), 1712–1726.

[4] Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics*, 28, 407–412.

[5] Lotto, A. J. & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysic,s* 60, 602–619.

[6] Kingston, J., Kawahara, S., Chambless, D., Key, M., Mash, D. & Watsky, S. (2014). Context effects as auditory contrast. *Attention, Perception, & Psychophysics*, 76(5), 1437–1464.

[7] Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception and Psychophysics*, 68 (7), 1227–1240.

[8] Kang, S., Johnson, K. & Finley, G. (2016). Effects of native language on compensation for coarticulation. *Speech Communication*, 77, 84–100.

Intonational Structure Mediates Speech Rate Normalization in the Perception of Speech Segments

Jeremy Steffman

University of California, Los Angeles (USA)
jsteffman@g.ucla.edu

It is well known that the phonetic properties of a given sound vary systematically as a function of its prosodic position (e.g. [1-3]). However, if and how speech perception is sensitive to this prosodically driven variability remains an open question. The present study explores this in light of recent research.

Kim & Cho [4] argued that listeners are sensitive to initial strengthening of VOT in perception, requiring longer VOT for a /p/ response (when categorizing a /p/-/b/ VOT continuum) when the target sound was IP-initial. Because VOT is longer in IP-initial position (e.g. [2, 5]), a shift in categorization might reflect sensitivity to initial strengthening and compensation, implicating listener awareness of phrasal domains (and their phonetic encoding). However, this same finding has been reanalyzed more recently by Mitterer et al. [6] as possibly originating only from speech rate normalization. Because Kim & Cho manipulated duration as a cue to a preceding IP boundary, the IP-initial target in their experiment was preceded by a lengthened segment. This durational difference would be expected to shift categorization via speech rate normalization alone (as local modulations in rate shift categorization of VOT, e.g. [7]). Mitterer et al. showed that more global slowing shifts categorization in the same direction (using the same stimuli as [4]) and also that flattening the pitch preceding the target does *not* shift categorization. Both of these experiments do not offer support for listener awareness of prosodic structure, leading the authors to conclude that speech rate normalization may be the driving force behind the effect. The current study further addresses this question in a new way: by orthogonally varying duration and F0 (intonation) as a cue to prosodic structure in a 2x2 design. By holding duration constant and manipulating intonation as a cue to boundary, its independent effects can be analyzed.

In Experiment 1, listeners categorized a VOT continuum (0-45 ms in 5ms steps) as /p/ or /b/ (in a 2AFC task). All crucial manipulations were made (with PSOLA resynthesis) to the vowel [eɪ] immediately preceding the target "pa/ba" in the carrier phrase "I'll say pa/ba again" (with H* pitch accents on "I'll" and the target). The two length conditions are LONG and SHORT, where the LONG condition reflects IP final lengthening, and the SHORT condition is an IP medial vowel. The two intonation conditions are named LH and FLAT, where LH is a low rising (L-H%) boundary tone, and FLAT is high flat F0. The choice of these contours was informed by English intonational phonology (e.g. [8]): in the LONG condition both are interpretable as IP boundary tones (L-H% and H-L% respectively). However, in the SHORT condition, FLAT intonation on un-accented "say" should be interpreted as the transition between adjacent pitch accents. In contrast, LH intonation is *only* interpretable as L-H% (cuing a boundary), because in English intonational phonology a non-prominent syllable with two pitch targets and an early-aligned L target must be L-H% (i.e. an unambiguous boundary tone). Therefore it is predicted that if listeners are sensitive to prosodic structure in their perception of segmental contrasts, there should be an effect of intonation in the SHORT condition only. Following Kim & Cho's [4] original argument: if LH intonation is perceived as a boundary tone (in the SHORT condition) it should shift categorization so that longer VOT is required for a /p/ response (decreasing /p/ responses overall).

50 monolingual speakers of American English participated in Experiment 1. Results are assessed with a linear regression with logistic linking and the lsmeans post hoc test (for interactions) in R. Firstly, length had a main effect where the LONG condition significantly decreased /p/ responses ($p < 0.001$) (replicating [4, 6]). The predicted asymmetrical effect of F0 was borne out in a significant interaction ($p = 0.015$): as predicted, intonation had a significant effect in the SHORT condition ($p = 0.011$), and no effect in the LONG condition ($p = 0.52$). However the directionality of the effect is contra Kim & Cho's original prediction: LH intonation significantly *increased*, not decreased, /p/ responses (see Fig. 1 below). This effect suggests listeners are using intonational structure to compute speech rate. That is, because L-H% naturally occurs with a phrase final lengthening, when compressed onto a SHORT vowel it is interpreted as an increase in speech rate, shifting subsequent VOT categorization.

If this interpretation is correct the effect should be replicable with other temporal contrasts. Experiment 2 successfully replicated the effect by examining the categorization of /t/~/d/ from a vowel duration continuum in the carrier sentence "I'll say coat/code now", with the same intonation and durational manipulations. Vowel length is a reliable cue to word final obstruent voicing in English and is used by listeners in perceiving "voicing" contrasts (e.g. [9]). Based on Experiment 1, it was predicted that LH intonation should increase "code" responses in the SHORT condition only. That is, if a compressed L-H% gives the impression of increased speech rate, shorter durations of the vowel in the target word would be perceived as "code" (meaning "code" responses will increase in the LH versus the FLAT condition). This would be explainable as *intonationally informed* rate normalization, not awareness of initial strengthening, because unlike VOT, the duration of a vowel (in an IP-initial CV) does not increase via initial strengthening [2]. The results of Experiment 2 confirm these predictions. As with Experiment 1, length had a main effect ($p < 0.01$), and intonation and length also showed a significant interaction ($p < 0.01$). In the SHORT condition *only*, LH intonation significantly increased "code" responses ($p < 0.01$; see Fig. 2), while it had no effect in the LONG condition ($p = 0.46$). These results are further confirmation of the hypothesis that listeners interpret the compressed boundary tone as increased speech rate, modulating perception of vowel duration in the target word, as with VOT.

These results of these experiments suggest that prosodic/intonational structure is indeed relevant in the perception of temporal contrasts, being used by listeners to compute speech rate. This predicts that listeners with different intonational systems/linguistic backgrounds will normalize for rate differently. Results will be discussed further in terms of their extension to cross linguistic study and their implications for how linguistic systems might mediate language-general normalization processes.

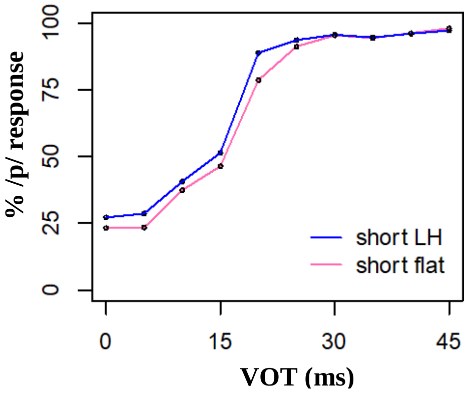


Fig. 1 Exp. 1 categorization in the SHORT condition only, split by intonation condition.

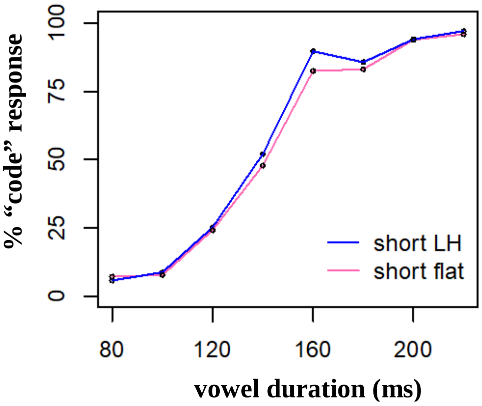


Fig. 2 Exp. 2 categorization in the SHORT condition only, split by intonation condition.

References

[1] Jun, S.-A. (1993). *The Phonetics and Phonology of Korean Prosody*. The Ohio State Univ.

[2] Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *JPhon*, *37*(4), 466–485.

[3] Keating, P., Fougeron, C., Hsu, C., & Cho, T. (2003). Domain initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Lab Phon VI*. Cambridge Univ. Press.

[4] Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *JASA*, *134*(1), EL19–EL25.

[5] Pierrehumbert, J., & Talkin, D. (1992). Lenition of / h / and glottal stop. In G. Doherty & D. R. Ladd (Eds.), *Papers in Lab Phon* (pp. 90–116). Cambridge Univ. Press.

[6] Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *JPhon*, *54*, 68–79.

[7] Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology*, *7*(5), 1074–1095.

[8] Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(01), 255–309.

[9] Raphael, L. J. (1972). Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English. *JASA*, *51*(4B), 1296–1303.

Effects of contextual prosodic structural cues on phonological inference: a case of the post-obstruent tensing rule in Korean

Sahyang Kim[1], Holger Mitterer[2], Taehong Cho[3]

Hongik University, Seoul[1]; University of Malta, Malta[2]; Hanyang University, Seoul[3]
sahyang@hongik.ac.kr, Holger.mitterer@um.edu.mt, tcho@hanyang.ac.kr

A phonological rule may or may not apply, depending on where in the prosodic structure it operates [1,2,3]. When it does, it often creates a potential ambiguity during the word recognition process (e.g., ‘right ([p]) berry’ vs. ‘ripe berry’ due to assimilation [4,5]). ‘Post-Obstruent Tensing Rule (**POT**)’ in Korean changes a lax into a tense consonant after an obstruent within, but not across, an Accentual Phrase (AP) [3]. In this case, listeners should be able to infer the underlying form of the altered segments by making reference not only to the phonological context [6,7] but also to the prosodic structure that conditions it (cf. [8,9,10,11]). Much is, however, unknown about how the phonological inferencing is indeed conjoined with prosodic structural information available in the input. The present study explores this question by investigating how Korean listeners process phonologically altered forms due to POT in reference to prosodic structure. Specifically, it utilizes an eyetracking paradigm to examine the extent to which the underlying form is activated given the altered surface form, and how the activation is modulated over the timecourse by prosodic cues that signal prosodic structure.

Three eyetracking experiments were carried out with 60 subjects in total (20 each). Target words were 24 minimal pairs that differed only with the underlying word onset (lax vs. tense). Auditory stimuli were used to guide the listener to click on a target. Crucially, there were an obstruent /k/ context that triggers POT and a nasal (control) context that does not (1a, b). Prosodic contexts varied in each experiment. Exp. 1 compared effects of Intonational Phrase (IP) (along with final lengthening and a boundary tone) across which POT does not apply vs. a phrase-internal word (Wd) boundary across which POT is expected to operate. Results showed significantly *less* looks to the lax target in the Wd than in the IP condition, reflecting the fact that the lax target was phonologically altered (i.e., tensified) in the Wd condition. Crucially, however, the phonological inferencing effect was not observable: Upon hearing a phonologically tensified lax stop even in the licensing (/k/+lax) Wd condition, listeners were faithful to the phonetic input, looking more to the competitor (the underlying tense) rather than to the intended target (the lax).

In an effort to further test the phonological inferencing effect, Exp. 2 continued to test to what extent listeners would accept the phonologically tensified phonetic form as intended (i.e., as the underlying lax) in the licensing Wd condition, but this time listeners were allowed to choose either the intended lax target or a “no answer” option if they think there is no word on the screen that matches with the auditory input. As can be seen in Fig.1, results now showed a phonological inferencing effect: Listeners accepted tensified lax forms more as the underlying lax targets in the obstruent /k/ context that triggers POT than in the control context especially between 200-400ms *and* 800-1000ms time windows. Exp.3 then tested how such a phonological inferencing would be modulated by prosodic structure when cued only by F0 (i.e., without any lengthening) that was consistent with a tonal pattern for an AP boundary. Results showed that the phonological inference effect observed in Exp. 2 disappeared: There was no difference between the POT-triggering /k/ context and the control (/n/) context (Fig.2). Furthermore, a statistical comparison of the results between Exps. 2 and 3 indicated that phonological inferencing was observed in a later time window (i.e., 800-1000ms), suggesting that the post-lexical F0 effect kicks in later in speech processing. These results taken together suggest that the acoustically same input for the target is processed with differential degrees of lexical activation over the timecourse depending on the computed prosodic structure; and a low-level F0 cue alone can be exploited in this computation process, fine-tuning lexical processing. This also implies the importance of understanding speech processing in conjunction, or in parallel, with prosodic analysis [8,11].

(1a) an obstruent context (which licenses POT)
Hwamyeon-eseo *bora-se**k*** # ***puri**-wi semo-lul* *nureuseyo*
'Screen-on' 'purple-colour' # '**beak**-above' 'trianlge-ACC' 'press'
'On the screen, click on the triangle above the purple beak.'

(# indicates the critical boundary between the target word and the preceding obstruent /k/;
Note that POT is applied only in AP-internal Wd boundary; The target word is underlined.)

(1b) a nasal context (which does not license POT)
Hwamyeon-eseo *yeonnora**ng*** # ***puri**-wi* *Semo-lul* *Nureuseyo*
'Screen-on' 'yellow-colour' # '**beak**-above' 'trianlge-ACC' 'press'
'On the screen, click on the triangle above the yellow beak.'

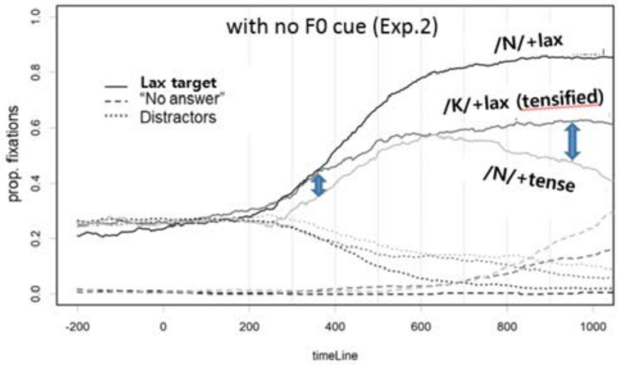


Fig.1 Looks to the lax target, "No answer", and the distractors in **Exp.2 (with no F0 cue)**. The 1st solid line shows the looks to the lax target in the /N/+lax context, the 2nd in the /k/+lax (tensified) context, and the 3rd in the /N/+tense context.

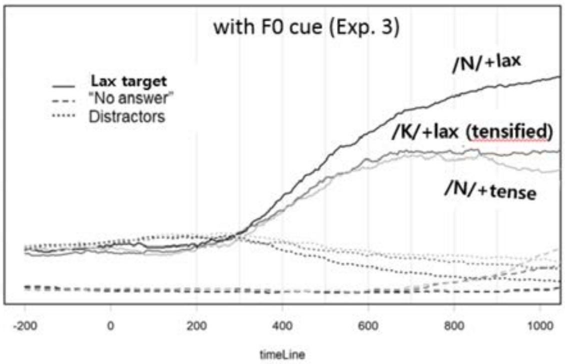


Fig.2 Looks to the lax target, "No answer", and the distractors in **Exp.3 (with F0 cue)**. The 1st solid line shows the looks to the lax target in the /N/+lax context, the 2nd in the /k/+lax (tensified) context, and the 3rd in the /N/+tense context.

References

[1] Nespor, M., & Vogel, I. 1986. Prosodic Phonology (Studies in generative grammar 28). *Dordrecht: Foris.*

[2] Kuzla, C., Cho, T., & Ernestus, M. 2007. Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics, 35*(3), 301-320.

[3] Jun, S. A. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology, 15*(2), 189-226.

[4] Gow, D. W. 2002. Does English coronal place assimilation create lexical ambiguity?. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 163.

[5] Gow, D. W., & Im, A. M. 2004. A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*(2), 279-296.

[6] Gaskell, M. G. 2003. Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics, 31*(3), 447-463.

[7] Mitterer, H., Kim, S., & Cho, T. 2013. Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language, 69*(1), 59-83.

[8] Cho, T., McQueen, J. M., & Cox, E. A. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics, 35*(2), 210-243.

[9] Mitterer, H., Cho, T., & Kim, S. 2016. How does prosody influence speech categorization?. *Journal of Phonetics, 54*, 68-79.

[10] Kuzla,, C., Ernestus, M. & Mitterer, H. 2010. Compensation for Assimilatory Devoicing and Prosodic Structure in German Fricative Perception. In C. Fougeron, B. Kühnert, M. D'Imperio, & N Vallée (Eds.), *Laboratory Phonology 10* (pp. 731-758). Berlin: Mouton de Gruyter.

[11] Cutler, A. 2012. Native Listening. Cambridge, MA: MIT Press.

Do working memory and autistic traits predict L2 prosody perception?

Grace Kuo

National Taiwan University (Taiwan)
graciakuo@ntu.edu.tw

Prosody is crucial in language production, comprehension and acquisition ([1][2][3]; and among others). In order to interpret an utterance correctly, listeners must understand its prosodic structure which include the examination of *prosodic prominence* and *prosodic phrasing* ([4]; and among others).

The present study aims to investigate the role of the *working memory* and *autistic traits* in predicting the prosodic prominence and the prosodic boundaries in a second language. The acquisition of a new sound system for an L2 learner is often influenced by the cognitive, social, and psychological factors. Among these factors, the ones emphasizing individual differences have gained growing interest in second language development. *Working memory* and *autistic traits* are of interest here in that (a) working memory was found closely related to second language acquisition, and (b) disordered prosody was found to be a feature of impaired communication, and general population with higher level of autistic traits tended to perceive prosody differently from those with lower level of autistic traits.

Very little work has been done on the relation between individual differences and prosody perception. [5] investigated the effect of *working memory* on offline decision concerning attachment preferences. They found that readers who had low memory spans, regardless of the attachment preference of their native language, had a greater tendency to break up large segments (i.e., chose high attachment). [6] reported how the presence of a prosodic boundary influenced the parsing of ambiguous relative clauses and they also found a positive correlation between listeners' performance and their autistic traits. [7] further investigated the influence of autistic traits on prominence and boundary perception among native English speakers. Results showed that there was an interaction between the types of the pitch accent and autistic trait. More specifically, individuals with higher levels of autistic traits identified fewer pitch accented words, especially those with lower pitch levels (such as L*). However, the interaction between autistic trait and prosodic boundary perception is not apparent. Individuals did not vary significantly in their identification of prosodic boundaries, i.e., intermediate phrase (ip) and intonational phrase (IP) in particular.

The present study examined the role of *working memory* and *autistic traits* in predicting the perception of prosody in neurotypical L2 learners of English. A group of university students (N=80) in Taiwan were asked to complete two questionnaires for cognitive and personality assessment – autism-spectrum quotient (AQ) questionnaire [8] and working memory questionnaire [9]. Then, each subject will have to participate in an auditory task – Rapid Prosody Transcription Task [10], in which they were required to make speeded identifications of prominent words and locations of prosodic juncture. The auditory material was a 10-minute political speech ("Weekly Addresses" recorded by Barack Obama) transcribed previously for prosodic events using the ToBI (Tones and Break Indices) conventions [11] by trained phoneticians. Subjects were divided into three groups based on their performance in the AQ test and the working memory test.

In [12] and [13], L2 learners and. Non-native speakers performed as well as or even better than the native speakers when it came to prosody perception task. In other words, prosodic cues, rather segmental information, was being used by listeners as a primary cue. Thus, it is predicted that L2 learners would perform similarly as the native speakers in [7].

Preliminary results (only 20 subjects' data has been analysed so far) support the hypothesis. The results are shown in Figure 1 and 2 – the three groups (Low, Mid, High) are defined according to subjects' AQ score and working memory score – higher score means the subjects are more autistic-like and have better working memory. The findings with L2 learners were similar to the findings with native speakers in [7]'s study except that the percentages of correctly identified prominence and boundary are lower in general (i.e., L2 learners' accuracy in identifying prominence and boundaries are approximately 12% and 10% respectively, but native speakers' accuracy were 32% and 25% respectively). In addition, there was a modest interaction between accent type and the individual difference (autistic traits + working memory) but the interaction between boundary perception and individual difference is rather weak.

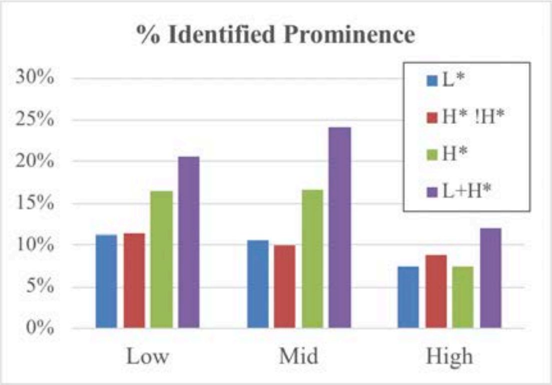


Fig.1 Percentage of words identified as prominent by L2 learners, as function of ToBI accent type. The learners were grouped by their perforance in AQ and working memory test.

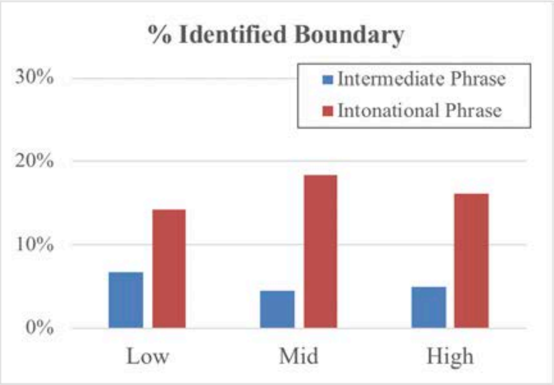


Fig.2 Percentage of prosodic boundaries – intermediate phrase (ip) and intonational phrase (IP) – identified by L2 learners. The learners were grouped by their performance in AQ and working memory test.

References

[1] Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech 40*, 2, 141-201.

[2] Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *SSLA 28*, 1-30.

[3] Wagner, M., & Watson, D. (2010). Experimental and theoretical advances in prosody: a review. *Language and Cognitive Process 25*, 905-945.

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

[5] Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: a psychometric approach. *Journal of Experimental Psychology 136*, 1, 64-81.

[6] Jun, S., & Bishop, J. (2015). Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier and E. Gibson (eds.): *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor*. Studies in Theoretical Psycholinguistics 46. Springer.

[7] Bishop, J., & Kuo, G. (2016). Do "autistic-like" personality traits predict prosody perception? *LabPhon 15 Satellite Workshop.*

[8] Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders 31*, 5-17.

[9] Vallat-Azouvi, C., Pradat-Diehl, P., & Azouvi, P. (2012). The working memory questionnaire: A scale to assess everyday life problems related to deficits of working memory in brain injured patients. *Neuropsychological Rehabilitation*, 1-16.

[10] Cole, J., Mo, Y.-S., & Baek, S.-D. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, *25*, 1141-1177. [3] Bishop, J., & Kuo, G. (2016). Do "autistic-like" personality traits predict prosody perception? *LabPhon 15 Satellite Workshop.*

[11] Beckamn, M. E., & Elam, G. A. (1993). Guideliens for ToBI Labelling. The Ohio State University Research Foundation.

[12] Carlson, R., Hirschberg, J., & Swerts, M. (2005). Cues to upcoming Swedish prosodic boundaries: subjective judgements studies and acoustic correlates. *Speech Communication 46*,, 326-333. [5] Reinisch, E., Wozny, D., Mitterer, H. & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, *45*, 91-105.

[13] Kuo, G. (2013). Perception and acoustic correlates of the Taiwanese tone sandhi group. Ph.D. dissertation. UCLA.

Exemplar Encoding of Intonation in Imitated Speech

Suyeon Im[1] & Jennifer Cole[2]
[1]University of Illinois at Urbana-Champaign (USA), [2]Northwestern University (USA)
sim16@illinois.edu, jennifer.cole1@northwestern.edu

Introduction: Speakers are able to imitate gradient aspects of speech related to segmental [1,2] and suprasegmental features [3,4]. These findings offer support for the claim of exemplar theory [5] that speakers encode non-contrastive phonetic detail of heard speech and are able to reproduce it in later production of the similar speech. Recent evidence, however, suggests that speakers are able to reproduce the phonetic detail only if it is phonologically relevant [6], and that speakers are more accurate at producing phonological features than phonetic detail [7]. The question of whether speech encoding captures all aspects of phonetic detail is interesting for intonation because phonological specification of intonational feature is particularly sparse. What is the extent of the phonological domain in which f0 contours are encoded? This study addresses these question through the analysis of imitated speech, comparing the similarity between imitated and stimulus f0. **We hypothesize that the domain in which the imitated and stimulus f0 is the most similar corresponds to the domain of f0 encoding.**

Method: 33 American English speakers (23 females) were asked to reproduce 12 sentence stimuli of 7-13 words each, produced by a female American English speaker. Participants were instructed to repeat the utterances they heard, in the manner the model speaker said them. The complex subject noun phrases (e.g. *The adversarial prosecutor*…) were produced by the model speaker with one or two pitch accents and a following prosodic break. We analyzed f0 contours of the subject noun phrases in four prosodic domains: intermediate phrase (D1), pitch accented syllable (D2), stressed syllable (D3), and syllable (including stressed and unstressed--D4). The f0 contours were downsampled to 30 equidistant points in each domain, as shown in Fig.1. Four Generalized Additive Mixed Models [8], one for each domain, were run to model the f0 difference (ERB) between imitated and stimulus speech as a function of f0 time-points, pitch accent pattern, subject's gender as fixed effects, and subject and item as random effects.

Results: Model comparison yields deviance explained values of 79.2% for D1, 77.1% for D2, 77.7% for D3, and 77.5% for D4. All the domains show good model fit, including the analysis that includes every syllable--the smallest domain--which captures the temporal dynamics of f0 in fine detail. D1 shows the best model fit. It models the f0 contour holistically, as a continuous, time-varying pattern over the whole phrase. These findings lend support for a theory of the cognitive encoding of f0 that represents temporal dynamics in fine detail.

This study investigates the phonological interval that defines the domain of cognitive encoding of intonational phonetic detail using imitated speech of American English. This study examines the similarity of f0 contour between imitated sentences and their stimuli over domains of varying size and prosodic status, from the syllable to the prosodic phrase. Results show evidence for the cognitive encoding of the phonetically detailed f0 contour over an entire prosodic phrase (ip). The findings do not support a model of encoding that excludes phonologically unspecified regions. The findings contribute to previous studies showing speakers' adaptation to fine phonetic detail and call for an extension of exemplar models to include phonetically detailed representation of f0 patterns.

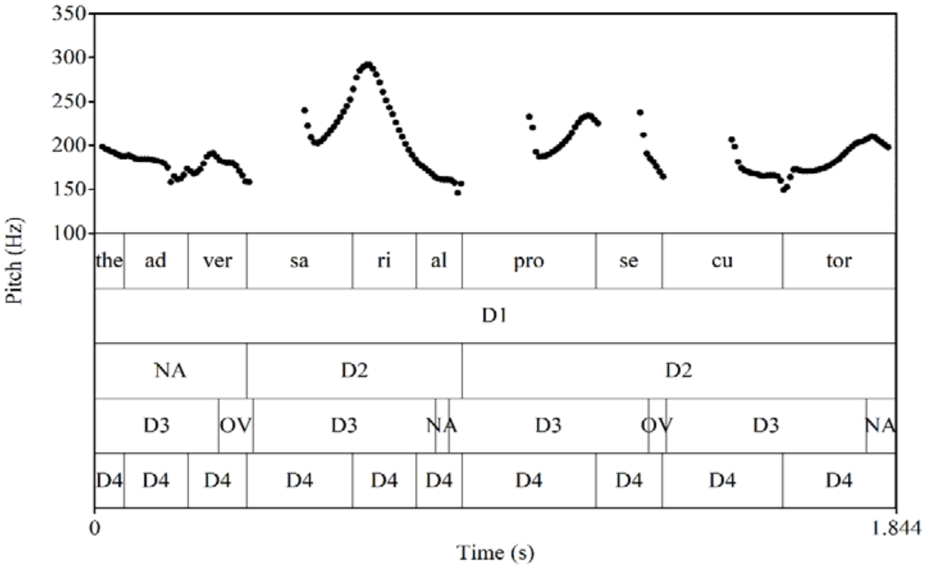


Fig.1 Four domains (D1-D4) in decreasing order for the subject noun phrase “the adversarial prosecutor”: D1 for ip, D2 for pitch accented syllable, D3 for stressed syllable, and D4 for syllable. NA indicates the regions outside the domains of analyses. OV stands for the overlapping regions between intervals. Each domain covers different regions over the subject noun phrase. D1 and D4 cover the entire subject noun phrase. D2 and D3 cover the phonologically specified regions over the subject noun phrase.

References

[1] Nye, P. W., & Fowler, C. A. (2003). Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, 31(1), 63-79.

[2] Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382-2393.

[3] D’Imperio, M., Cavone, R., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*, 5, 1-10.

[4] German, J. S. (2012). Dialect adaptation and two dimensions of tune. In Q. Ma, H. Ding and D. Hirst (Eds), *Proc. of the 6th International Conference on Speech Prosody* (pp. 430-433). Shanghai: Tongji University Press.

[5] Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251-279.

[6] Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1), 168-173.

[7] Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: what do listeners imitate?. In *Twelfth Annual Conference of the International Speech Communication Association*.

[8] Wood, S. (2006). *Generalized additive models: An Introduction with R*. Texts in Statistical Science.

Phonetic Convergence in VOT in a Cue-Distractor Paradigm

Stephen Tobin[1,2], Marc Hullebus[1] & Adamantios Gafos[1,2]

[1]Universität Potsdam (Germany), [2]Haskins Laboratories (USA)
tobin@uni-potsdam.de, hullebus@uni-potsdam.de, gafos@uni-potsdam.de

We aimed to uncover immediate effects of phonetic convergence within a single perception-production cycle using a cue-distractor paradigm [1, 2, 3]. It is already known that phonetic convergence can take place over long time scales [4]. Heretofore, however, it has not been possible to demonstrate convergence at the shortest time scale of individual perception-production cycles (which must be the basis for convergence at the longer time scales).

In a cue-distractor paradigm, participants are instructed to utter syllables in response to visual cues (“if you see ** say ka, if you see ## say ta”). Once the visual cue appears, but before the participant has produced their response, participants hear a distractor over headphones and are told ignore it. Previous applications of this paradigm have revealed fine-grained variation in response times (RTs) according to the degree of featural (in)congruency (voicing, place of articulation) between cue and distractor. Here we extend the domain of application of this paradigm to test whether effects of the distractor on the response appear not only in reaction times but also in phonetic parameters (VOT in the present experiment) of the response.

In our extension of this paradigm, we controlled, for the first time, the phonetic distance between the mean VOT of a particular speaker-hearer’s speech and the VOTs with which that speaker-hearer is presented via the distractors. This was made possible by registering participant-specific VOTs in a baseline block of 100 trials (50 *ta*, 50 *ka*) without any distractor. This baseline block preceded the experimental block. In the experimental block (*N*=720) distractors were presented on every trial (360 *ta*, 360 *ka*; 50% matching cue+distractor, 50% mismatching cue+distractor). Both *ta* and *ka* distractor stimuli were drawn from 9-step VOT continua ranging from 45-85 ms. The continua were created by resynthesis in Praat [5]. In order to ensure sufficient variation in phonetic distance, participants were grouped by VOT (high vs. low with a cut-off of 65 ms) and were assigned to hear the short VOT distractors (5 steps: 45-65 ms) of the long VOT distractors (5 steps: 65-85 ms) by counterbalancing.

Measurements were made with reference to the waveform and consisted of (i) oral release (the first substantial deviation from zero amplitude), (ii) phonation onset (the earliest indication of a periodic signal) and (iii) phonation offset (return of amplitude to zero amplitude or background-level noise). All measurements were made semi-automatically in a software package developed in explicitly for this purpose [6].

In order to detect subtle effects of convergence (cf. [4]) and to adequately represent the predictor variable of phonetic distance, we derived two variables before beginning our analysis. First, to represent phonetic distance between participant baseline and distractor tokens, we subtracted the mean of participants’ baseline VOT (henceforth VOT^{mb}) from each trial’s distractor VOT (henceforth, δ^{db}). On this scale, zero means that the VOT of the distractor was identical to that of the participant VOT^{mb}. Positive values indicate that distractors were above the participant VOT^{mb} and negative values indicate that distractors were below the participant VOT^{mb}. Next, to detect subtle effects of convergence among response VOTs in spite of considerable variation in syllable duration, we normalized response VOTs by syllable duration ($t_{\text{release burst}} - t_{\text{phonation offset}}$). Syllable duration covaries with VOT [7] and could confound subtle effects of convergence. Thus, we divided response VOTs by syllable duration (VOT/σ). As with δ^{db}, we converted this quotient into a difference score ($[\mathrm{VOT}/\sigma]_{\text{distractor task}} - [\mathrm{VOT}/\sigma]_{\text{baseline task}}$) on a participant-wise basis (henceforth, δ^{rb}). On this scale, zero means that the normalized response VOT quotient was identical to that of the participant’s normalized baseline VOT quotient. Positive values indicate that normalized response VOT quotients were above the baseline quotients and negative values indicate that they were below the baseline quotients. Given the two variables defined above, convergence is indicated when increases (/decreases) in δ^{db} yield increases (/decreases) in δ^{rb}.

Pending further data measurement and analysis, we present the results of 22 participants in the *ta* visual cue condition. Using linear mixed effects regression, we observe significant effects of (within-trial) RT, and significant interactions of δ^{db} x Match, RT x Match, and δ^{db} x RT x Match. Figure 1 depicts four regression lines from the Match (top row) and the Mismatch (bottom row) conditions, taken from representative points along our continuous predictor of RT (left to right). We highlight our two main results. First, there is an effect of δ^{db} (distance between distractor and baseline VOT [x-axis]) on our dependent variable δ^{rb} (distance between response and baseline VOT [y-axis])—positive slopes in the relation between the two variables indicate convergence. Second, just as notable are the modulating effects of RT and Match. That is, later responses yield less convergence (flatter slopes at longer RTs) and mismatching cue+distractor pairs yield less convergence than matching ones (flatter slopes in Mismatch condition). Overall, our results constitute the first demonstration of trial-to-trial phonetic convergence on VOT values of response syllables to the VOT values of distractor syllables.

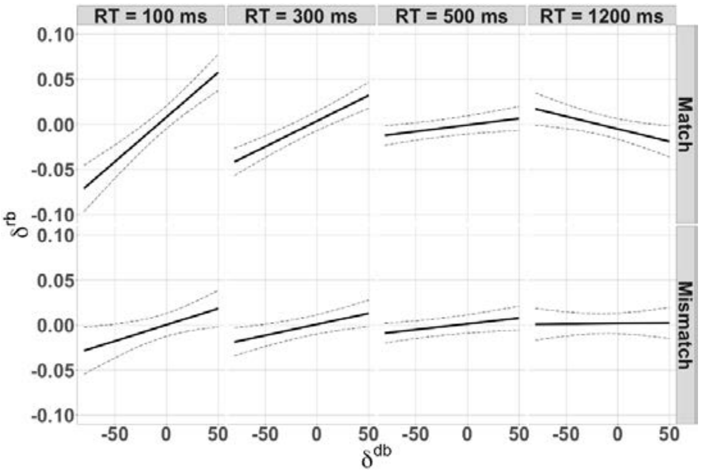


Fig. 2 δdb x δrb regressions with 90% CIs. Four representative points from continuous RT are shown from left to right. Match vs. mismatch cue-distractor conditions are shown in the top vs. bottom row, respectively. Positive slopes indicate convergence.

References

[1] Kerzel D. & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility, *JEP:Human Perception & Performance 26(2)*, 634-647.

[2] Gallantucci, B., Fowler, C. & Goldstein, L. (2009). Perceptuomotor compatibility effects in speech, *Attention, Perception & Psychophysics 71(5)*, 1138-1149.

[3] Roon, K. & Gafos, A. (2015). Perceptuo-motor effects of cue-distractor compatibility in speech: beyond phonemic identity, *Psychological Bulletin & Review 22(1)*, 242-250.

[4] Tobin, S. Nam, H, Fowler, C. (2017). Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics 65*, 45-59.

[5] Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 3 February 2018 from http://www.praat.org/

[6] S. Kuberski, S., Tobin, & A. Gafos. (2016). A landmark-based approach to automatic voice onset time estimation in stop-vowel sequences, *Proceedings of IEEE GlobalSIP*, 60-65.

[7] Allen, S., Miller, J. & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America 113(1)*, 544-552.

Poster Presentations (Day 1)

Analysis of the Influence of Word Frequency in Auditory Perception

Shen Lue[1], Stephen Politzer-Ahles[1]

[1]*Hong Kong Polytechnic University (Hong Kong)*
14110709d@connect.polyu.hk

Lexicality has been shown to have a top-down influence on people’s perception of sounds [1]. When people hear an ambiguous stimulus, their perception will be influenced by the word context: for example, if they hear a sound that is somewhere in between being a good token of /t/ and a good token of /d/, they tend to hear it as /t/ more often in contexts where /t/ would make it a real word (*task*) than in contexts where /t/ would make it a nonword (**tash*), and likewise for /d/.

Based on Ganong's [1] finding, one would also expect that people are more likely to perceive /t/ in a context where it yields a relatively high-frequency word compared to one where it yields a relatively low-frequency word. In the experiment, we tested Mandarin speakers' perception of ambiguous sounds (with voice onset times [VOTs] that put them in between /d/ and /t/) in different contexts. In one context, if the ambiguous sound is perceived as /d/ it would yield the high-frequency word *duìhuà* (对话, "conversation") and if it is perceived as /t/ it would yield the low-frequency word *tuìhuà* (退化, "degeneration"). In the other context, if the ambiguous sound is perceived as /d/ it would yield the low-frequency word *duìyì* (对弈, "play chess") and if it is perceived as /t/ it would yield the high-frequency word *tuìyì* (退役, "retire"). If frequency has a similar top-down influence on perception as lexicality does, we predict that people will hear /t/ more in the *{d/t}uìyì* context, where /t/ would give them a high-frequency word, than in the *{d/t}uìhuà* context, where /t/ would give them a low-frequency word. We also tested the same ambiguous sounds in two contexts that yield words or non-words, in order to attempt to replicate the Ganong effect and to serve as a manipulation check. In the *{d/t}uìchū* context, where /t/ would give them a real word: *tuìchū* (退出, “quit”), and in the *{d/t}uìfāng* context, where /t/ would give them a nonword: **tuìfāng*, but /d/ would give them a word *duìfāng* (对方, “counterpart”). Finally, as a control, we tested perception of the same ambiguous sounds in a context where perceiving the sound as either /d/ or /t/ would yield a nonword: neither **duìchī* nor **tuìchī* is an existing word in Mandarin. We expect that at the two endpoints of the continuum (where the sound is presented with very low or very high VOT) participants would perceive the sound they had heard consistently as /d/ or /t/ regardless of lexicality and frequency, but during the middle part of the continuum (where the sound is ambiguous) they will perceive /t/ more often in the context where it yields real words or high-frequency words, compared to the context where it yields nonwords or low-frequency words. The stimuli were chosen from the SUBTLEX-CH corpus [2].

To create the stimuli, we recorded a token of *tuì* with a long VOT, and successively cut more and more of the aspiration period out to yield tokens with VOTs of 10 msec, 15 msec, 20 msec, 25 msec, 30 msec, 35 msec, 40 msec, 45 msec, and 50 msec. The tokens were identical in all ways except the aspiration period. Each VOT token was spliced onto each of the five second syllables (*huà*, *yì*, *chū*, *fāng*, and *chī*) to create the five contexts described above; this yielded 45 stimuli (9 VOTs × 5 contexts), each of which was repeated ten times in the experiment. The experiment consisted of ten blocks; in each block, the participant heard each stimulus once, in a random order. Participants were instructed to indicate whether the initial consonant they heard was /t/ or /d/.

Figure 1 shows the proportion of /t/ perceptions in each context and VOT, from 35 native Mandarin speakers. The blue lines represent the conditions for which we expect more /t/ perception (where /t/ perception yields real words and high-frequency words), the red lines the conditions for which we expect less. The dotted lines indicate the baseline Ganong effect comparison (where the stimuli can be perceived as real words or nonwords), and the solid lines the critical comparison (where the stimuli can be perceived as high-frequency or low-frequency words). For the Ganong effect, The blue dotted line standing for perceptions of the real word *tuìchū* (退出) is higher than the control black line where both perceptions yield nonwords, but the red dotted line standing for perceptions the nonword **tuìfāng* is lower than the control line. This shows that our experiment was able to capture the expected Ganong effect. More importantly, the solid blue line standing for perception of the high frequency word *tuìyì* (退役) is higher than the solid black control line, and the solid red line standing for perception of the low frequency word *tuìhuà* (退化) is lower than the control line, which represents that more perception of /t/ in the *tuìyì* (退役) context than in *tuìhuà* (退化) context. 32 of the 35 participants showed a difference in this direction. The result shows a lexical effect, in which the high frequency words are more commonly being perceived than words with low frequency. Thus, word frequency will also influence individual's speech perception.

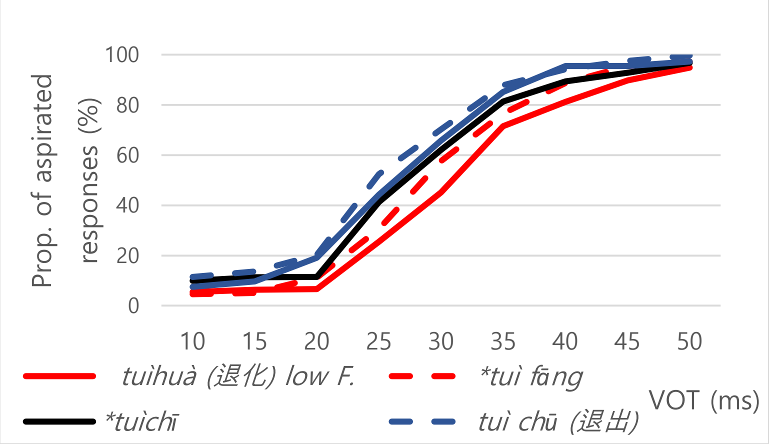


Fig. 1 Experiment Result

Reference

[1] Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110.

[2] Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos ONE, 5(6), e10729.*

Native speakers' perception of Mandarin lexical tones in contemporary pop music

Janice Wing-Sze Wong[1], Jung-Yueh Tu[2]

[1]Hong Kong Baptist University (Hong Kong SAR), [2]Shanghai Jiao Tong University (China)

janicewong@hkbu.edu.hk, jytu@sjtu.edu.cn

Previous research in tone-melody mapping has mainly surveyed across a number of songs in the target tone language and calculated the degree of parallelism between the direction of two adjacent musical notes and the direction of lexical tones of the two corresponding sung words. This is what Wong and Diehl [1] named *ordinal mapping*, through which Vondenhoff [2] reported that only less than 40% of tone and melody showed matches in contemporary Mandarin songs. Interestingly, even though these Mandarin songs display such a high degree of mismatches between tone and melody, native speakers of Mandarin seldom report that the tone-melody conflicts make the songs sound unnatural or odd. However, only very few studies have directly examined the strategies or acoustic cues that listeners utilized when identifying the sung syllables through looking at native speakers' perception of the lyrics. The present study, as a preliminary report of a larger project that analyses a greater corpus of contemporary Mandarin and Cantonese pop songs, attempts to add a new dimension to the understanding of the interaction between lexical tones and melody, by investigating how the perception of Mandarin lexical tones of individual syllables in a song, i.e., decontextualized sung syllables, is affected by the height of the melodic tune.

We used a perceptual experiment with stimuli extracted from a contemporary Mandarin pop song called "A Bit of Fortune" available on the market. An analysis of the lexical tones of the lyrics showed that almost each lexical tone (Tone 1: *high-level*; Tone 2: *rising*; Tone 3: *fall-rising*; Tone 4: *falling*), except for Tone 3 at note B5, was found sung at all seven notes within an octave ranging from C5 (523.25Hz) to B5 (987.77Hz). This constitutes a total of 27 stimuli (4 tones × 7 musical notes − 1 Tone 3 at note B5 = 27). The presence of almost all possible tone-note combinations, i.e., each lexical tone is not associated with a consistent musical note, raises the question about whether or not and to what extent the observed freedom in tone-melody mapping in Mandarin affects actual comprehension of the lyrics.

A total of 24 native Taiwan Mandarin speakers (17M7F), with age ranged from 18 to 20, participated in this perceptual identification experiment. They all had less than two years of music experience and all reported no hearing or speaking deficits. The experiment was conducted in two different sound-proof language laboratories located in Shanghai and Taipei. Two repetitions of all 27 stimuli were presented to the participants in a random order and they were asked to identify the tone, i.e., T1, T2, T3, or T4, of the stimulus they heard.

The mean of accurate identification of the four Mandarin tones among all participants was 33.52% (SD = 21.75%; chance level = 25%). T1 was usually correctly identified, followed by T3 and then T2. T4 was the least accurately perceived. A chi-square test of independence confirmed that the relationship between the Mandarin tones produced by the singer and tones identified by the listeners was significant, $\chi^2(9, N = 1263) = 95.28$, $p < .0001$. A standardized Pearson residual cell-wise post-hoc analysis showed that only T1 and T3 were more accurately identified than the other two tones. T2 also contributed the highest number of confusion with T4. Table 1 on the next page shows a confusion matrix for the targets and responses.

Four different Pearson chi-square tests of independence, separated based on their target tones, were also performed to examine the relationship between the musical notes adhered to the target tones produced and the lexical tones perceived. Musical notes and the tones perceived was dependent of each other (all at $p < .0001$). When T1 was sung at notes D and B, they were more accurately perceived than those which were sung at other notes. When T2 was sung at note D, it was usually perceived as T3; while T2 was sung at notes E and B, it was identified as T4 more significantly often. T3 was more correctly recognized when sung at note C whereas it was usually perceived as T1 when sung at note D. There was more confusion when perceiving T4 at different notes: it was wrongly perceived as T3 when sung at note C, as T2 when at note D, and as T1 when at note G. Only when it was sung at note E did it show significant and more accurate identification.

The results in the present study has shown that, besides the impact of the direction of tone-melody transitions, there are other strategies or acoustic cues that listeners used to understand lyrics. The pitch characteristics of individual syllables and the pitch range they were sung at should also be taken into consideration in tone-melody mapping. Listeners tended to be consistent in choosing T1 as their most favored choice, regardless of the tone produced. The default and most favored choice, T1, appears to show that the effects of melody have somehow overridden the prosodic cues in lexical tones of individual syllables. That is, the musical pitch may have been regarded as a part of the tone information. This is more evident when the tones were sung at higher notes, as they were more often identified as T1.

Although tone-melody conflicts are usually found in Mandarin songs in which listeners seldom report them awkward, the current results reflect that only T1 and T3 could be accurately identified more often. Interestingly, T1 and T3 are not the most frequently occurring tones in Mandarin, but T4 is [3]. While T1 was consistently a more favored choice, that T3 received higher accuracy rates may be due to its creaky feature which still stands out in singing. However, accurate identification of T2 was only near chance level (25%). Meanwhile, the tone which mostly favored T4 as the target was T2. This seems counterintuitive as T2 and T4 have opposite contours (rising vs. falling). This might suggest that when tones are sung with a melody, the contour cue might be neutralized or diminished due to the musical note, hence playing a less significant role as a cue. In addition, the notes did not show any pattern in influencing the tones perceived. More works are needed to investigate the underlying interaction, but the present preliminary findings have shown that tone-melody disagreement neither burdens nor eases lexical tone perception in lyrics sung in Mandarin, even in a decontextualized condition.

Table 1. Confusion matrix showing the number of actual identification tokens of 4 Mandarin lexical tones

		Actual Performance			
		T1	T2	T3	T4
Targets	T1	172 (52.91%)	80 (24.84%)	54 (16.77%)	19 (5.90%)
	T2	89 (27.64%)	94 (29.19%)	74 (22.98%)	72 (22.36%)
	T3	78 (28.26%)	66 (23.91%)	98 (35.51%)	39 (14.13%)
	T4	109 (33.85%)	96 (29.81%)	70 (21.74%)	53 (16.46%)

References

[1] Wong, P. C. M., & Diehl, R. L. (2002). How can the lyrics of a song in a tone language be understood? *Psychology of Music, 30*(2), 202-209.

[2] Vondenhoff, M. (2009). *An Optimality Theoretical Model of the Influence of a Sung Melody on the Interpretation of Mandarin Lexical Tones* (MA Thesis). University of Amsterdam.

[3] Wan, I. -P., & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology, 15*(3), 417-461.

Production of Trisyllabic Third Tone Sandhi in Mandarin by L1 and L2 Speakers

Jung-Yueh Tu [1], Janice Wing-Sze Wong [2] & Jih-Ho Cha[3]

[1] *Shanghai Jiao Tong Univ. (China),* [2] *Hong Kong Baptist Univ. (Hong Kong),*
[3] *National Tsing Hua Univ. (Taiwan)*
jytu@sjtu.edu.cn, janicewong@hkbu.edu.hk, toddcha@gmail.com

In Mandarin tonal inventory, the Tone 3 sandhi rule is straightforward in disyllabic words, where a Tone 3 followed by another Tone 3 is changed to a rising tone, similar to Tone 2 (indicated as T2, hereafter) [1, 2]. Nevertheless, when the Tone 3 sandhi rule is applied to trisyllabic words or polysyllabic phrases where all are Tone 3 syllables, it becomes more complicated and involves both the prosodic and morpho-syntactic domains [3, 4]. The application of the third tone sandhi rule to trisyllabic words then becomes quite challenging for second language (L2) learners of Mandarin. In particular, it is common that L2 learners may oversimplify a rule in a target language.

This study investigates how L2 learners with another tone language experience could master the Mandarin Tone 3 sandhi rule. Specifically, the study intends to focus on the production of Tone 3 sandhi in trisyllabic Mandarin words by Hong Kong Cantonese speakers. Mandarin uses four distinctive tones (high-level, mid-rising, low-dipping, and high-falling), to convey lexical meanings [2] whereas Cantonese has six lexical tones (high-level, mid-rising, mid-level, lo w falling, low rising, and low level) [5, 6]. In the current study, 30 Cantonese speakers and 13 Mandarin speakers were requested to produce 15 trisyllabic words and 5 hexasyllabic sentences with Tone 3 in sequences. The trisyllabic sandhi patterns can be viewed as [1+(2+3)] (e.g., mǐlǎoshǔ 'Mickey Mouse') and [(1+2)+3] (e.g., yǎnjiǎnggǎo 'speech script') patterns. The former sandhi pattern is realized as [T3+(T2+T3)] while the latter is realized as [(T2+T2)+T3]. Based on their phrasal (morpho-syntactic) structures and sandhi patterns, the stimuli can be put into six categories, as given in *Table 1*. The recordings were judged by two phonetically trained native speakers of Mandarin, who further identified the tonal errors made by the participants. The two native speakers evaluated the recordings and labelled the tone of each syllable in the trisyllabic words and hexasyllabic sentences with a choice among the four lexical tones.

The results for the L1 group showed that the Mandarin speakers robustly applied the Tone 3 sandhi rule in the same way as described (accuracy rate: 98.95%), except for only 3 tokens out of 286. Those for the L2 group showed that the overall accuracy rate of trisyllabic words was higher than that of hexasyllabic words, which indicated that Tone 3 sandhi was more accurately applied to the domains at the lexical level than the sentential level. Among the trisyllabic words, the accuracy rate for individual categories showed the effects of phrasal structure on the production of Tone 3 sandhi by the Cantonese speakers. The accuracy rate for Category 2 (VPs) was significantly higher than that for Category 1 and Category 3 (NPs with [1+(2+3)] and [(1+2)+3] sandhi patterns, respectively). The results demonstrated that the Tone 3 sandhi rule was better applied to VPs than NPs by the Cantonese speakers. Similarly, the accuracy rate in the hexasyllabic sentences also showed the effects of morph-syntactic structures on the application of Tone 3 sandhi. The accuracy rate for Category 5 (NP+VP) was significantly higher than that for Category 6 (NP+Adv-VP), which may imply that the sentences consisting of phrases other than NPs+VPs are more difficult for L2 learners to identify the sandhi domains in trisyllabic tones. The misproductions by L2 group were analysed in terms of four major types of error patterns: overgeneralization, under application, combination, and others. The results showed that the overall error rate in the "overgeneralization" pattern was the highest among the four error patterns. The average error rate of each error pa ttern are presented in the interaction plot, as shown in *Figure 1 & 2*. It is demonstrated that Cantonese speakers tend to apply Tone 3 sandhi in trisyllables in an overgeneralized way, changing Tone 3 into a rising tone in trisyllables with Tone 3 in a row. The findings suggested the effects of phono-syntactic interactions of Tone 3 sandhi and general error patterns in the acquisition of Mandarin tonal system by Cantonese speakers.

Table 1 The categories of stimuli.

Category	Syllables	Phrasal structure	Sandhi pattern	Tonal realization
1	trisyllabic	NP	[1+ (2+3)]	[T3+(T2+T3)]
2		VP	[1+ (2+3)]	[T3+(T2+T3)]
3		NP	[(1+2) +3]	[(T2+T2+T3)]
4		VP/AP	[1+ (2+3)]	[T3+(T2+T3)]
5	hexa-syllabic	NP+ VP	[1+ (2+3)]+ [1+ (2+3)]	[T3+(T2+T3)]+ [T3+(T2+T3)]
6		NP+ (Adv.-VP)	[(1+2)+3]+ [1+ (2+3)]	[(T2+T3)+T3]+ [T3+(T2+T3)]

(NP=noun phrase, VP=verb phrase, AP=adverbial phrase)

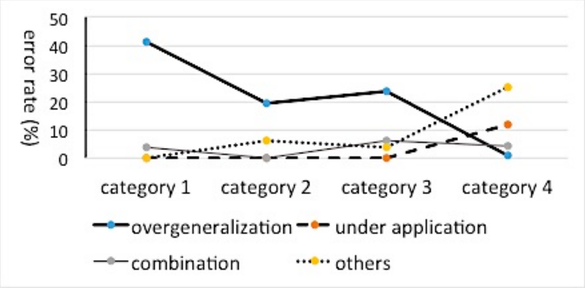


Fig. 1 Error rate of each pattern for category 1-4 in the trisyllabic words

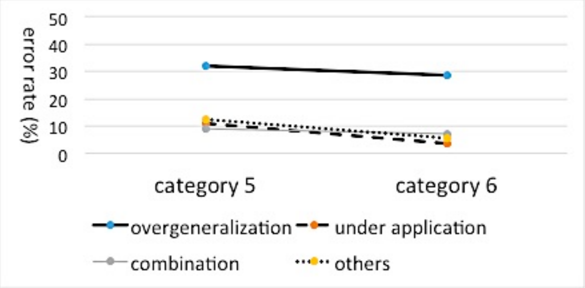


Fig. 2 Error rate of each pattern for category 5-6 in the hexasyllabic sentences

References

[1] Chao, Y. R. (1930). *A system of tone letters*. La Maître phonétique 45, 24-27.
[2] Lin, Y. H. (2007). *The sound of Chinese*. Cambridge: Cambridge University Press.
[3] Chen, M. Y. (2000). *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge: Cambridge University Press.
[4] Duanmu, S. (2000). *The Phonology of Standard Chinese*. Oxford University Press.
[5] Chao, Y. R. (1947). *Cantonese Primer.* Westport: Greenwood Press
[6] Bauer, R. S. & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Berlin and New York: Mouton de Gruyter.

Korean-speaking Toddlers' Perceptual Mapping Based on the VOT and F0 Dimensions

Gayeon Son[1, 2]

[1]*Kwangwoon University (Korea),* [2] *Univ. of Pennsylvania (USA),*
son.gayeon@gmail.com

Korean stop contrasts (lenis, fortis, and aspirated) have historically been phonetically differentiated by Voice Onset Time (VOT); however, as Seoul Korean is undergoing tonogenetic sound change in which there has been a loss of VOT differentiation in young adults' production, the role of fundamental frequency (F0) in differentiating Korean stop contrasts has increased as reported in [1], [2], and [3]. Recent production studies with Korean children indicated that the acquisition of fortis stops is strongly related to VOT development, while phonetic accuracy in lenis and aspirated stops increases with F0 development (e.g., [4] and [5]). However, despite the fact that the critical role of F0 in articulatory distinction among Korean stop contrasts has been consistently reported in speech of young adults and children, it is not yet understood how Korean young children's perceptual distinction develops in two-dimensional acoustic space.

Therefore, the current study investigates how toddlers' perceptual categorization of Korean stops develops along both VOT and F0 dimensions through a perception experiment. This study conducts a perceptual identification test by using a point-to-a picture task with 48 Korean monolingual children aged 24 to 47 months (Table 1). The children identified nine minimal pairs, which included every possible pair of lenis-fortis-aspirated homorganic stops. Multi-level quantitative analysis revealed that perceptual accuracy is higher for fortis or aspirated stops than for lenis stops, and that between 2 and 4 years of age, there is a significant interaction between children's age (in months) and successful perception of lenis stops ($p < 0.001$), suggesting that significant phonemic development has occurred in the F0 dimension. In addition, the results also showed that the same phoneme can be perceived differently depending on which phonetic dimension is dominantly involved in the perception. Lenis stops were more successfully identified with fortis counterparts compared to the cases when lenis stops were paired with aspirated counterparts by all age groups (Figure 1). This indicates that when VOT differences were amplified more than F0 differences, the toddlers more correctly perceived the stimuli. Therefore, it is suggested that children's perception system for Korean stop contrasts has started to develop mainly in the VOT dimension and to operate to distinguish fortis or aspirated stops after 2 years of age. F0 has not yet developed enough to allow perfect phonemic categorization of lenis stops before 4 years of age.

These findings indicate that during the process of acquisition of Korean stop contrasts, toddlers' perceptual mapping functions are distinctively based on the VOT dimension rather than the F0 dimension. In addition, it is suggested that this perceptual developmental pattern is related to articulatory development in the VOT dimension. This study provides essential evidence for further understanding how native phonological contrasts develop in young children.

Table 1. Child participant information for a perception test

Age group (years;months)	*M*	*SD*	Male	Female	Total
2;0–2;6	2;3.21	0;2	5	6	11
2;7–3;0	2;9.18	0;1.18	4	4	8
3;1–3;6	3;3.3	0.1;12	8	7	15
3;7–3;11	3;8.27	0;1.3	3	11	14

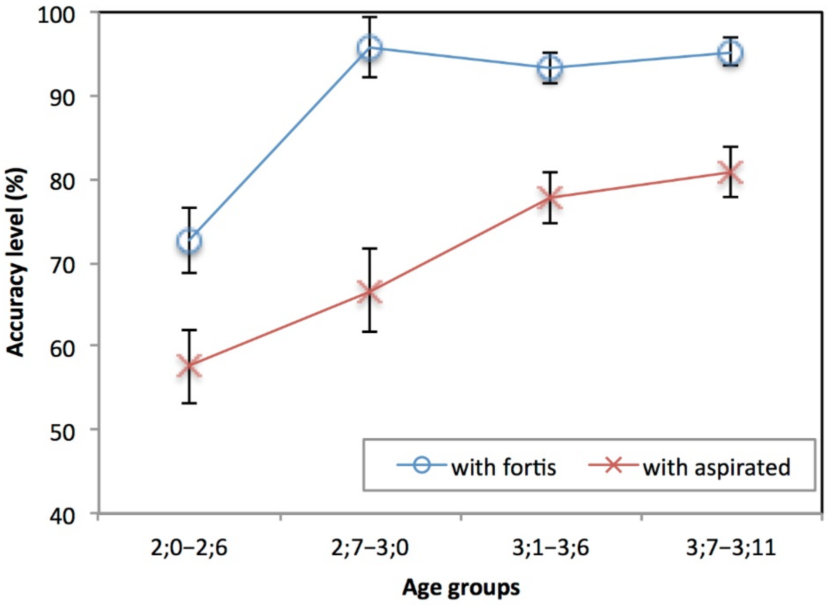


Fig. 1 Comparison of accuracy levels in the perceptual identification of lenis stops by age group, when provided with fortis and aspirated minimal pairs. Error bars represent standard errors.

References

[1] Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: a corpus study. Journal of Phonetics, 45, 76–90.

[2] Kang, K. H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America*, 124, 3909–3917.

[3] Wright, J. D. (2007). *Laryngeal Contrast in Seoul Korean*. Ph.D. dissertation, University of Pennsylvania, Philadelphia, PA.

[4] Kong, E. J., Beckman, M. E., & Edwards, J. (2011). Why are Korean tense stops are acquired so early?: The role of acoustic properties. *Journal of Phonetics*, 39, 196–211.

[5] Son, G. (2017). Interactive *Development of F0 as an Acoustic Cue for Korean Stop Contrast*. Ph.D. dissertation, University of Pennsylvania.

The Development of English Tense Agreement Morphology in Welsh-English Bilingual Children with and without Specific Language Impairment (SLI)

Hyowon Kwon[1, 2*] & Vicky Chondrogianni[1]

[1]*University of Edinburgh(UK),* [2]*Ghent University Global Campus(Korea)*
*Corresponding author: hyowon.kwon@ghent.ac.kr, v.chondrogianni@ed.ac.uk

To date, much research concerning the acquisition of English tense-marking morphology as a clinical marker has focused on comparing monolingual children with SLI to their TD peers (Leonard, 1998; Rice & Wexler, 2001). Studies contrasting children with and without SLI acquiring English as a second language (L2) have shown that L2 children have more difficulties with tense morphology in production than monolinguals, and English L2 children with SLI show exceptional deficits in tense morphology compared to their L2-TD peers (Chondrogianni & Marinis, 2011; Paradis, 2005; Paradis et al., 2008). Therefore, looking at accuracy and error types of tense morphology by L2 children with and without SLI could function as the potential clinical marker that distinguishes TD from SLI in English L2 children (Blom & Paradis, 2013). To our knowledge, research tapping into bilingual children with and without SLI has mainly been limited to investigating groups of either younger or older children, or specifically their early and late years of L2 exposure. However, comparing accuracy and error types in tense-marking morphology produced by both younger and older groups of L2-TD children is essential for revealing age effects in their developmental patterns of tense morphology. Hence, our aim is to provide a cross-sectional study, which investigates how tense morphology develops among three different profiles of bilingual children acquiring English as L2 across a wide range of ages, and whether the developmental trajectories of their 3SG –*s* and past tense acquisition profiles are in line with previous studies that support the prediction of Bybee's (1995, 2001) usage-based network model (Blom & Paradis, 2013; Blom et al, 2012). Specifically, our analyses include 1) whether tense can be equally problematic for young Welsh-English bilingual children with language impairment attending Welsh-medium schools as it has reported for other English L2-SLI populations (Blom & Paradis, 2013), 2) whether there are developmental changes in tense and agreement production in older L2-TD children with a homogeneous L1 (Welsh) and 3) how child-internal and language specific factors modulate performance in L2-TD and L2-SLI children. A group of Welsh-English TD bilingual children from 7-9-years of age (Mage: 93.72 months), a younger group of L2-TD children from 4-6-years (Mage: 67 months) and an age-matched group of L2-SLI peers (Mage: 63 months) were administered the tense probe from the Test of Early Grammatical Impairment (Rice & Wexler, 2001). Responses that had been transcribed and scored on the TEIG were selected and coded in order to analyze individual tense morphology elicitation probes. The results indicated that the three groups differed in their production of 3SG –*s* and regular past tense but not in terms of accuracy on irregular past tense verbs. The L2-SLI children produced similar error types to the younger L2-TD children, who differed from their older L2-TD peers in this respect. Vocabulary size, frequency, and morphophonology differentially contributed to L2 children's performance. We discuss these results within current accounts of language development and impairment.

References

[1] Blom, E., & Paradis, J. (2013). Past tense production by English second language learners with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *56*(1), 281-294.

[2] Blom, E., Paradis, J., & Duncan, T. S. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular–s in child L2 English. *Language Learning*, *62*(3), 965-994.

[3] Bybee, J. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, *10*(5), 425-455.

[4] Bybee, J. (2001). *Phonology and language use* (Vol. 94). Cambridge University Press.

[5] Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, *1*(3), 318-345. [6] Tremblay, A. (2008). Is L2 lexical access prosodically constrained? On the processing of word stress by French Canadian L2 learners of English. *Applied Psycholinguistics, 29,* 553-584.

[6] Leonard, L. (1998). *Children with specific language impairment*. Cambridge, MA:MIT Press

[7] Paradis, J. (2005). Grammatical Morphology in Children Learning English as a Second Language Implications of Similarities With Specific Language Impairment. *Language, Speech, and Hearing Services in Schools*, *36*(3), 172-187.

[8] Paradis, J., Gavruseva, E., & Haznedar, B. (2008). Tense as a clinical marker in English L2 acquisition with language delay/impairment. *Current trends in child second language acquisition: A generative perspective*, 337-356.

[9] Rice, M. L., & Wexler, K. (2001). *Rice/Wexler test of early grammatical impairment*. Psychological Corporation.

The Three-way Contrast of Conversational Korean Stops

Jae-Hyun Sung

Yonsei University (Korea)
jsung@yonsei.ac.kr

Conversational speech has received considerable attention from linguists and speech scientists over the past few decades [1,2]. Speech sounds in casual conversation differ dramatically from those found in citation forms, owing to various phonetic reduction processes such as shorter duration or deletion of segments or syllables, shorter distance between sound categories, and less available cues in speech signals [3,4,5,6]. Instrumental acoustic studies of conversational speech have, however, focused on widely spoken languages such as English or Dutch. In order to diversify and expand on conversational speech research, this study examines the unique system of stop contrast in Korean [7,8,9] and their acoustic cues at play in spontaneous phone conversation. More specifically, the study investigates some key acoustic properties of intervocalic Korean stops – phonemically lenis, fortis, and aspirated – in casual speech, adding the unique sound system to the growing body of languages represented in the conversational speech literature.

Comparisons of Korean stops from 10 Korean speakers' careful reading and casual conversation (see Fig 1 for an example) show greater acoustic variability in conversational speech than in careful reading, as noted in the previous literature. Despite the substantial variability in acoustic signals, conversational Korean stops contrast with one another in three different acoustic dimensions: relative intensity difference (Fig 2), VOT (Fig 3), and closure duration (Fig 3), allowing some degree of idiosyncratic patterning among speakers (Fig 4). The findings from this study point to manner-specific adjustments in conversational Korean stops relative to their more commonly studied careful speech counterparts, and add weight to individual variation in phonetic cue uses in speech production.

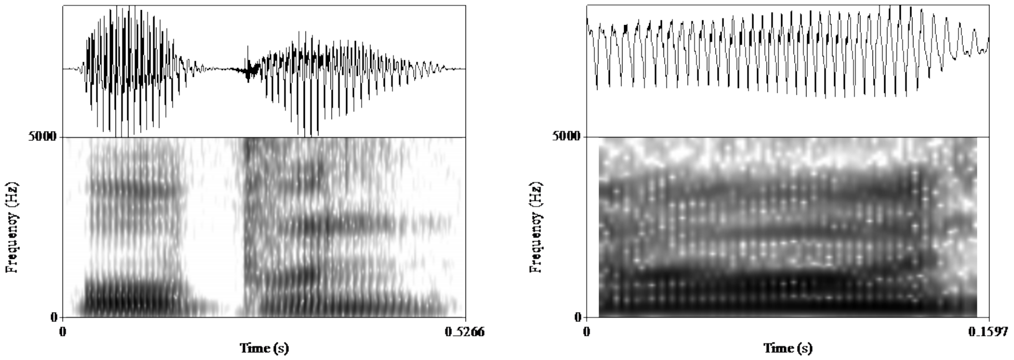


Fig.1 Read (left) and conversational (right) tokens of /opun/ '5 minutes', Speaker 8

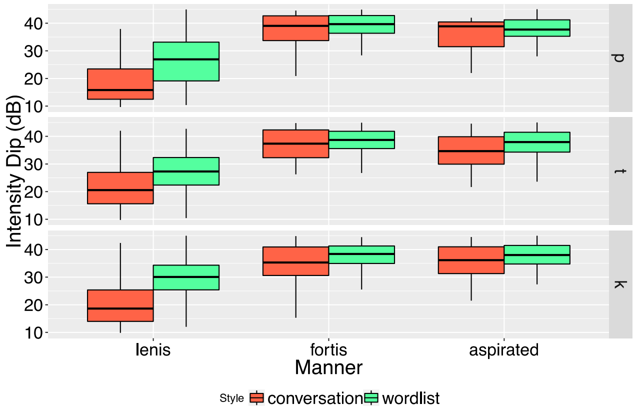


Fig.2 Intensity differences for conversational (red) and read (green) tokens of Korean stops

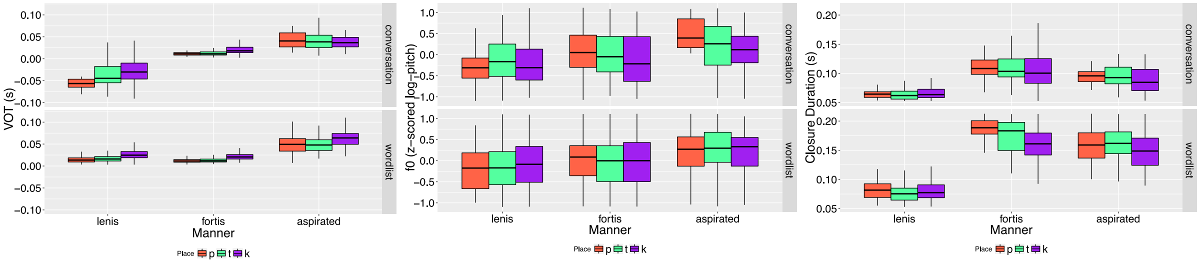


Fig.3 VOT, f0 (z-scored), and closure duration for conversational and read tokens of Korean stops

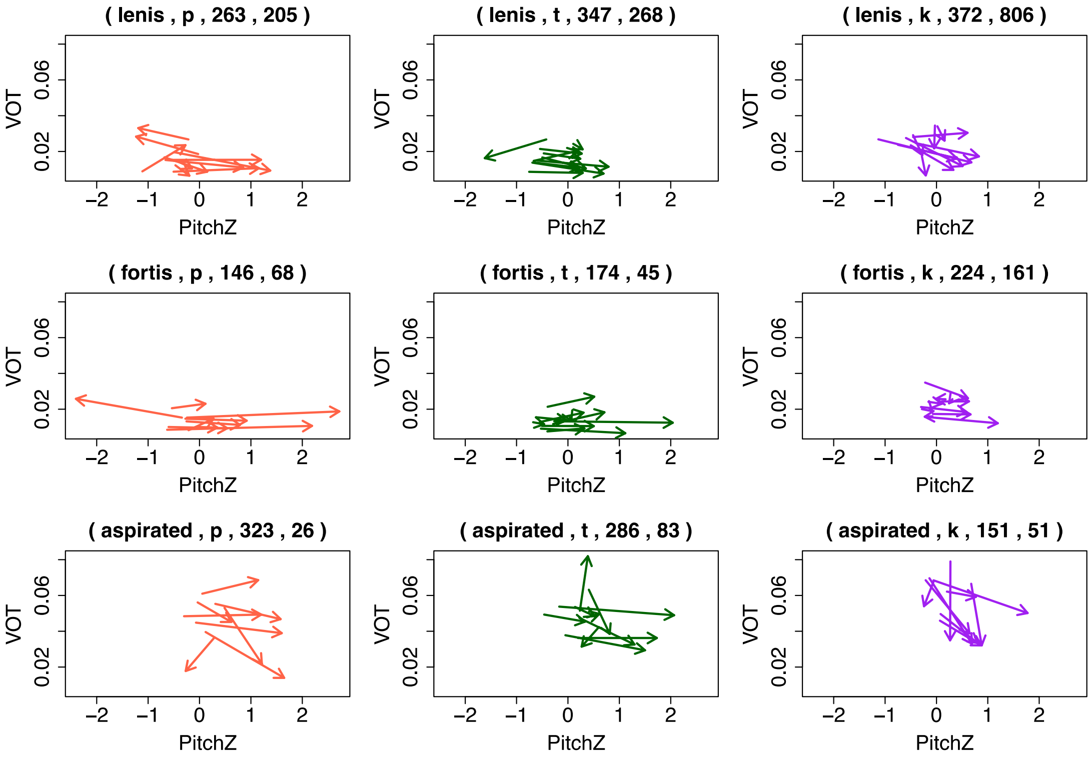


Fig.4 VOT*f0 cue variation in the careful to casual shift ; numbers represent the numbers of tokens in the wordlist (left) and the conversation (right) contexts ; each arrow in the plot represents individual speaker, in which the direction refers to the direction of shift (careful to casual)

References

[1] Barry, W. & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, 31, 51-66. Chen, A. (2011).

[2] Ernestus, M. & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39, 253-260.

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

[4] Warner, N. & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America*, 130, 1606-1617.

[5] Cheng, C. (2012). *Mechanism of Extreme Phonetic Reduction : Evidence from Taiwan Mandarin*. PhD dissertation, University College London.

[6] Brenner, D. (2015). *The Phonetics of Mandarin Tones in Conversation*. PhD dissertation, The University of Arizona.

[7] Cho, T. Jun, S-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30, 193-228.

[8] Kang, K-H., & Guion, S. G. (2008). Clear speech production of Korean stops : Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America*, 124, 3909-3917.

[9] Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics*, 41, 117-132.

The Role of Within-Category Duration Differences in Speech Perception

Seongjin Park & Natasha Warner

University of Arizona (USA)
seongjinpark@email.arizona.edu, nwarner@email.arizona.edu

The aim of the present study is to examine the role of the duration of a segment in facilitating perception of the segment. Specifically, this study investigates whether simply being longer leads to more accurate recognition of sounds, even if the sound is relatively steady-state and longer duration does not provide additional acoustic cues. For example, if one token of /f/ is 160 ms long while another is 190 ms long, is the latter perceived more accurately? If acoustic cues of the segment are uniformly distributed within the segment, additional duration might aid perception, but if cues are concentrated in a certain area within the segment (e.g. areas near Stevens' [1] landmarks), additional steady-state duration would not improve perception.

To investigate whether the duration of each stimulus affected the identification of each segment, we analyzed data from the English Diphones project [2]. In that project, participants identified gated fragments of all 2288 possible combinations of two segments in English (e.g., diphones /hə/, /dt/, /ab/, /iɛ/). Each of two segments in diphone stimuli (e.g., /h/ and /ə/ in /hə/) were divided into three intervals with same duration, so there were six gates for each diphone (see Fig. 1). For example, in diphone /hə/, Gate 1 was from the beginning of the stimulus to one third through /h/; Gate 2 was from the beginning to the two thirds of /h/; Gate 3 was to the end of the first segment; Gate 4 included from the beginning of the first segment to one third through /ə/; Gate 5 to two thirds of the way through /ə/; Gate 6 to the end of the diphone. For stops and affricates (e.g., /p/, /ʧ/), the first gate point in the segment was positioned halfway through the closure and the second just before onset of the burst (rather than at thirds of the segment), so that the entire burst and aspiration/affrication noise consistently fell within the final gate of the segment. Listeners heard each gated stimulus, and had to respond with what two sounds they heard or might have heard the beginning of. For example, a listener might hear /sa/ at Gate 6 and correctly identify both segments, or a listener might hear /ða/ at Gate 1 (ending 1/3 through the first segment) and guess incorrectly at both segments. This dataset provides probabilistic information about how listeners use acoustic cues as they become available over the course of the speech signal.

We measured the duration added by each gate (e.g., the duration difference between Gate 2 and Gate 1) and the listeners' average improvement in accuracy at each gate (e.g., how much more accurately listeners identified /s/ in /sa/ when hearing Gate 2 than Gate 1). The segments of each manner of articulation were analyzed separately. Consonants were categorized as stops, fricatives, affricates, liquids, nasals, or glides; vowels were categorized as tense, lax, or diphthongs. The added duration and the change of accuracy were normalized with min-max normalization. If longer duration facilitates perception, one would predict a positive correlation between duration added by a gate and improvement in perception at that gate.

For the consonants, the most notable finding was that affricates had a significant positive correlation between the added duration and the accuracy improvement whether the consonant was the first (e.g. /tʃa/) or second (e.g. /atʃ/) segment of the diphone (segment 1: $r = 0.78$, $p < 0.001$; segment 2: $r = 0.60$, $p < 0.001$). This significant correlation was caused by the longer duration of Gate 3 and Gate 6 for affricates, due to the end of Gate 2 or 5 being positioned just before the burst. The burst plus affrication noise in affricates is both long, and provides exceptionally rich perceptual cues, leading to very large improvement in perception at that gate. For the vowels, the results showed that the diphthongs had a significant correlation between duration added and accuracy improvement (segment 1: $r = 0.19$, $p < 0.001$; segment 2: $r = 0.15$, $p < 0.001$), although this was limited to Gates 2 and 4. This probably reflects the fact that at these gates, a longer duration of sound is more likely to contain enough formant change to provide cues to the second vowel quality of the diphthong. Thus, the correlations between duration added and accuracy improvement were significant for affricates and diphthongs only at certain gates because specific acoustic cues were embedded in those gates and were more likely to be present the longer the gate portion was, not because the duration of each gate was simply longer. That is, the apparent correlation actually reflects a categorical effect of presence/absence of specific perceptual cues, not a continuous and gradual effect of longer duration being inherently more recognizable. There were few other significant positive correlations apart from those in the affricates and diphthongs, and none were consistent across gates.

The present study suggests that longer duration does not inherently help listeners to identify the segments they have heard. The correlation between added duration and improvement in accuracy was only significant when particular acoustic cues occurred within the interval. This suggests that acoustic information tends to be concentrated in particular portions of the signal, and that listeners perceive speech primarily through those information-rich acoustic areas.

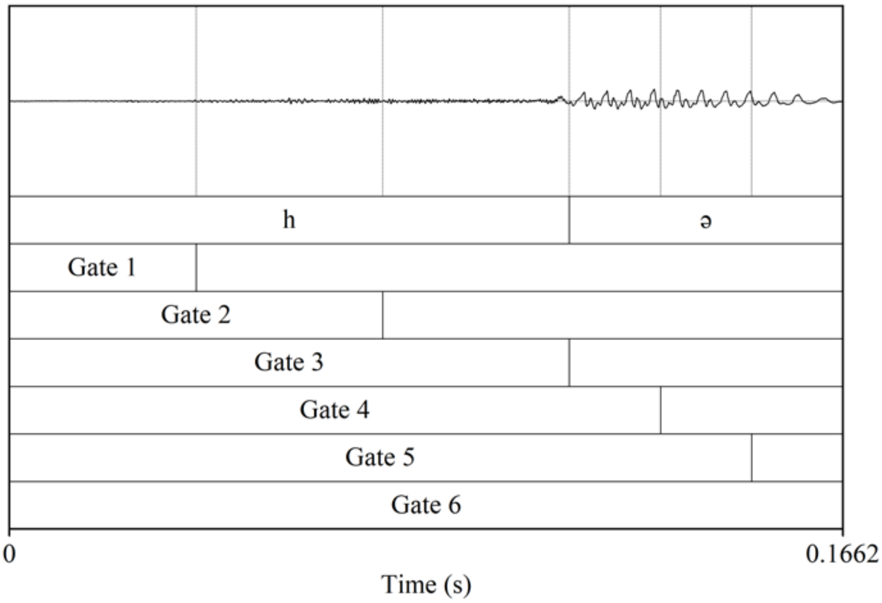


Fig.1 Six gates for diphone /hə/

References

[1] Stevens, K. N. (2000). Diverse acoustic cues at consonantal landmarks. *Phonetica*, 57, 139–151.

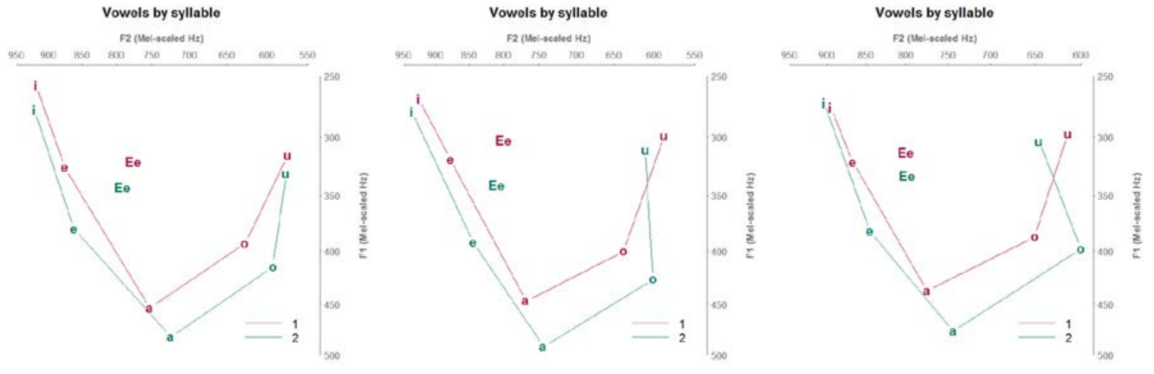
[2] Warner, N., McQueen, J., & Cutler, A. (2014). Tracking perception of the sounds of English. *Journal of the Acoustical Society of America*, 135, 2995-3006.

Role of syllable weight on vowel acoustic space: A case in Assam Sora

Luke Horo & Priyankoo Sarmah

Department of HSS, IIT Guwahati (India)
luke@iitg.ernet.in, priyankoo@iitg.ernet.in

This study argues that syllable weight affects vowel acoustic space and evidence is drawn from Assam Sora, a transplanted variety of the Sora language. Sora is a South Munda language of the Austroasiatic language family spoken in parts on Orissa and Andhra Pradesh by 252,519 individuals [1]. Assam Sora emerged due to the migration of Sora people from Orissa to Assam as indentured tea laborers in the 19th century [2]. The study by [3] reveals that Assam Sora has six vowels and disyllables always have weak-strong word prosody. This study examines the role of syllable weight on vowel acoustic space in Assam Sora vowel system. Vowel acoustic space is generally analysed to examine the difference in vowel system of two dialects [4] or to examine gender variations in the vowel system of the same language [5]. Thus, analysis of vowel acoustic space provides significant information regarding similarity or dissimilarity in the overall shape and structure of vowel systems. Therefore, by examining the vowel acoustic space of Assam Sora vowel system separately in first and second syllable of disyllables, this work aims to determine the role of syllable weight on the overall shape and structure of Assam Sora vowel system across syllables. For this purpose, 48 disyllables are recorded from 30 native Assam Sora speakers (15 Males and 15 Females) living in Sonitpur, Lakhimpur and Udalguri district of Assam. Subsequently, vowel acoustic space is calculated in first and second syllable with the help of formant frequency measurements (F1 and F2) of all six vowels in Assam Sora. Figure 1 illustrates Assam Sora vowel plots, based on average F1 and F2, of the three districts showing difference in vowel area space across first and second syllable. This shows that Assam Sora vowel area space differs as a function of syllable position. Subsequently, by estimating the actual vowel area space, it is revealed that vowel area space in Assam Sora vowel system is always larger in the second syllable then in the first syllable. Hence, this work provides evidence that syllable weight significantly affects the vowel area space in Assam Sora whereby, vowel area space is larger in the heavy syllable than in the weak syllable. Moreover, considering the argument that weak-strong word prosody is typically found in Southeast Asian Austroasiatic languages (e.g. Mon-Khmer), but not in South Asian Austroasiatic languages (e.g. Munda) [6] this work suggests that weak-strong word prosody is an invariant feature in Assam Sora.



Sonitpur Lakhimpur Udalguri

Fig.1 Assam Sora vowel plots across first and second syllable

References

[1] Registrar General of India. (2001). Distribution of the 100 Non-Scheduled Languages – India, States & Union Territories – 2001 Census. New Delhi: Office of the Registrar General and Census Commissioner, India.

[2] Kar, R. K. (1981). Savaras of Mancotta. New Delhi: Cosmo Publications.

[3] Horo, L., & Sarmah, P. (2015). Acoustic analysis of vowels in Assam Sora. North East Indian Linguistics, 7, 69–88.

[4] Jacewicz, E., Fox, R. A., & Salmons, J. (2007). Vowel space areas across dialects and gender. In International congress of phonetic sciences (Vol. 16, pp. 1465–1468).

[5] Lengeris, A. (2016). Comparison of perception-production vowel spaces for speakers of standard modern Greek and two regional dialects. The Journal of the Acoustical Society of America, 140 (4), 314–319.

[6] Donegan, P. J., & Stampe, D. (2002). South-east Asian features in the Munda languages: Evidence for the analytic-to-synthetic drift of Munda. In Annual meeting of the Berkeley linguistics society (Vol. 28, pp. 111–120).

Australian English listeners' cue weighting of spectral change and duration in the categorization of front vowels

Daniel Williams[123], Paola Escudero[23] & Adamantios Gafos[1]

[1]University of Potsdam (Germany), [2]Western Sydney University (Australia), [3]ARC Centre of Excellence for the Dynamics of Language (Australia)

daniel.williams@uni-potsdam.de, paola.escudero@westernsydney.edu.au, gafos@uni-potsdam.de

Research on vowel segments has been guided by a belief that relevant spectral information lies within a relativity "steady-state" portion of their formants. An alternative view is based on observations that vowels formants show regular patterns of frequency change, referred to as vowel inherent spectral change (VISC) [1]. Recent research is beginning to find that VISC is perceptually relevant, e.g., by playing a role in distinguishing vowel contrasts. For instance, the mean F2 frequencies of Standard Southern British English (SSBE) /i:/ and /u:/ are quite similar, but the F2 of /i:/ increases over time, whereas the F2 of /u:/ decreases [2]. Mirroring these acoustic patterns, vowels containing rising F2 frequencies are more likely to categorized as /i:/, whereas vowels containing falling F2 frequencies are more likely to be categorized as /u:/ by SSBE listeners [3].

The Australian English (AusE) front vowels /iː/, /ɪ/ and /ɪə/ (as in "bead", "bid" and "beard", respectively) are strikingly close to one another in a conventional F1 × F2 vowel space. The three vowels exhibit distinct patterns of VISC, as shown in Figure 1 for male speakers [4]. Specifically, /iː/ and /ɪə/ have large F1 × F2 trajectory lengths whereas /ɪ/'s trajectory length is much shorter. With respect to trajectory direction, /iː/ resembles a "closing" vowel, while /ɪ/ and /ɪə/ are both "centering" vowels. The three vowels also differ in duration and the male speakers in [4] produced /iː/, /ɪ/ and /ɪə/ with durations of 168 ms, 101 ms and 206 ms, respectively.

Using the acoustic investigation of AusE vowels in [4] as a starting point, the present study tested the perceptual relevance of VISC and duration in AusE /iː/, /ɪ/ and /ɪə/. Specifically, we were interested in the properties of VISC which are perceptually most relevant and therefore investigated the role of F1 × F2 trajectory length as well as F1 × F2 trajectory direction, along with vowel duration.

20 monolingual native AusE listeners took part in a multiple-alternative forced-choice identification task in which they heard one vowel stimulus per trial and responded by selecting /iː/, /ɪ/ or /ɪə/ on a computer screen. The vowel stimuli were based on average values from male speakers [4] and varied in vowel duration (four logarithmically equal steps), F1 × F2 trajectory length (four equally spaced ERB magnitude steps) and F1 × F2 trajectory direction (two steps), but were identical in all other respects. Two further F1 × F2 trajectory steps corresponding to zero spectral change and an exaggerated F1 × F2 trajectory length acted as continuum endpoints.

The categorization results are shown in Figure 2. To determine the perceptual relevance of F1 × F2 trajectory length, F1 × F2 trajectory direction and duration, mixed-effects logistic regression models were fit to the data with the three acoustic cues serving as predictors and random intercepts included for subject and by-subject slopes added for the acoustic cues. The models' results confirmed a closing F1 × F2 trajectory is well associated with the vowel /iː/, whereas a centering F1 × F2 trajectory is more strongly associated with both /ɪ/ and /ɪə/. Vowel duration was also important for categorizing all three vowels, especially for /ɪ/. Finally, F1 × F2 trajectory length was negatively associated with categorizing the vowel stimuli as /ɪ/.

In order to measure the relative weightings of individual acoustic cues, relative importance analyses were conducted, which partition the variance explained by the predictors even if they are correlated [5]. The analyses revealed F1 × F2 trajectory direction is weighted very strongly in the perception of /iː/, while duration plays a rather limited role. For /ɪ/, duration is by far the strongest cue, followed by F1 × F2 trajectory length. For /ɪə/, both F1 × F2 trajectory direction and duration are weighted equally. Lastly, F1 × F2 trajectory length is a very weak cue for identifying both /iː/ and /ɪə/, despite their relatively large F1 × F2 trajectory lengths in speech production.

With respect to the role of VISC in speech perception, the results show that the direction of spectral change serves as a critical cue for phonemic identity in vowels which typically exhibit more substantial spectral change. Interestingly, the magnitude of the change itself is far less relevant, suggesting it may simply enhance the cue of direction or, since there is some variability in how /iː/ and /ɪə/ are realized by different sub-populations of AusE speakers [6], it may contribute more strongly to other aspects of speech perception, such as speaker identity.

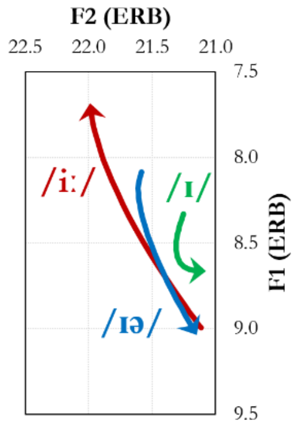


Fig. 1 Average F1 × F2 trajectories for AusE male speakers reported in [4]

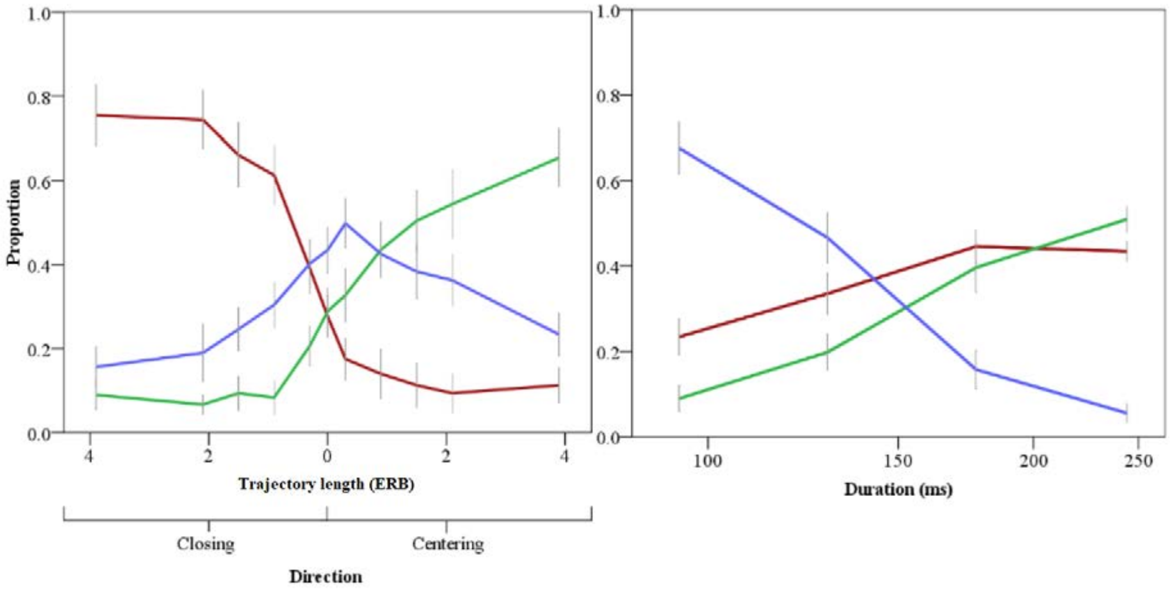


Fig. 2 Categorization of the stimuli as /iː/ (red), /ɪ/ (blue) and /ɪə/ (green) according to F1 × F2 trajectory length and F1 × F2 trajectory direction (left) and duration (right)

References

[1] Nearey, T. M. and Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification, *Journal of the Acoustical Society of America*, 80, 1297-1308.

[2] Williams, D. & Escudero, P. (2014). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English, *Journal of the Acoustical Society of America*, 136, 2751-2761.

[3] Chládková, K., Hamann, S., Williams, D. & Hellmuth, S. (2017). F2 slope as a perceptual cue for the front-back contrast in Standard Southern British English, *Language and Speech*, 60, 377-398.

[4] Elvin, J., Williams, D. & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English, *Journal of the Acoustical Society of America*, 140, 576–581.

[5] Tonidandel, S. & LeBreton, J. M. (2011). "Relative Importance Analysis: A useful supplement for regression analysis, *Journal of Business Psychology* 26, 1-9.

[6] Cox, F. (2006). /hVd/ vowels in the speech of some Australian teenagers, *Australian Journal of Linguistics*, 26, 147-179.

Prosodic Structure Constrains the Processing of Denasalized Nasals in Korean Lexical Access

Seulgi Shin & Annie Tremblay

University of Kansas, USA
seulgi.shin@ku.edu, atrembla@ku.edu

This study investigates whether prosodic structure modulates listeners' interpretation of phonetic variants in lexical access. It does so by examining Korean listeners' processing of denasalized nasal consonants in word-initial position.

Prosodic context is one of the phonological factors that has been shown to influence spoken word recognition. For example, several studies have demonstrated that acoustic cues to prosodic boundaries help listeners resolve segmentation ambiguities in the speech signal [1,2]. What is less clear from previous research, however, is whether the prosodic context also influences listeners' processing of words that are potentially ambiguous at the segmental level.

The present study addresses this question by focusing on the potential segmental ambiguities caused by domain-initial denasalization in Korean. Korean nasal consonants undergo nasal weakening at the beginning of the Accentual Phrase (AP); as a result, denasalized nasals share acoustic properties of AP-medial (voiced) lenis plosives [3,4]. These shared acoustic properties have been shown to be interpreted differently depending on the prosodic context in which they are heard: Using an offline phoneme identification task, Kim [5] found that denasalized nasals were perceived as nasals when spliced into the initial position of CVCV sequences but as plosives when spliced into the intervocalic position of VCV sequences. These results suggest that Korean listeners use the prosodic context to interpret the phonetic variants of nasal consonants. Yet, it remains to be seen whether the prosodic context also influences Korean listeners' interpretation of these phonetic variations in lexical access.

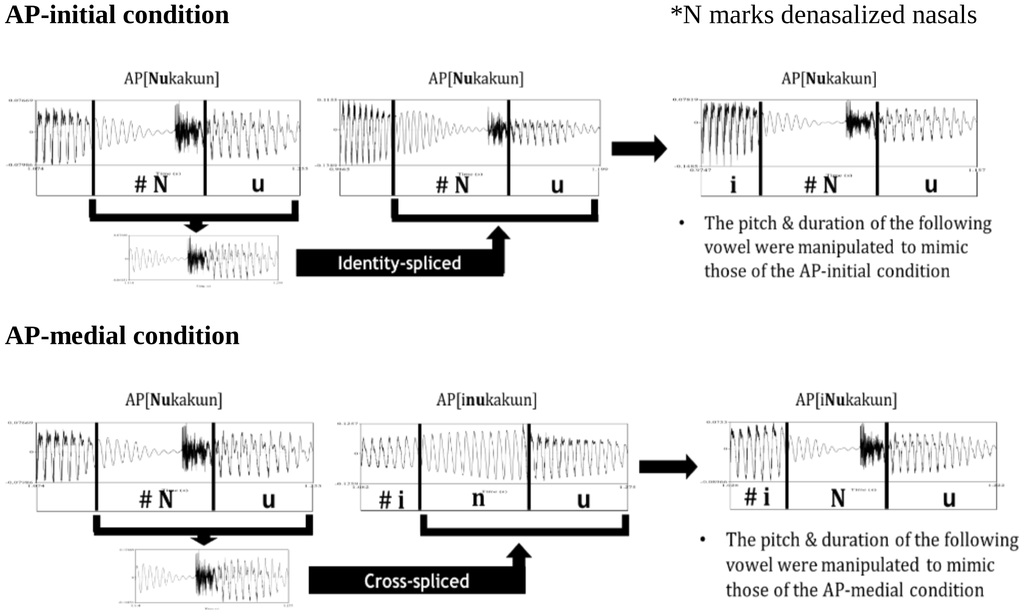
If denasalized nasals can be processed as nasals or as plosives depending on the prosodic context in which they are heard (as suggested by Kim's findings [5]), Korean listeners should activate nasal-initial words more when hearing denasalized nasals in AP-initial position than when hearing them in AP-medial position, and they should activate plosive-initial words more when hearing denasalized nasals in AP-medial position than when hearing them in AP-initial position.

Thirty-six Seoul Korean listeners completed two cross-modal priming experiments with lexical decision with at least a two-day gap between the experiments. The stimuli contained 32 /n/-/t/ disyllabic minimal pairs controlled for token frequency. In Exp. 1, the critical targets were visually presented Korean words that began with a nasal (e.g., *noru* 'roe deer'). The experimental primes were auditory words that rhymed with the target but began with a denasalized nasal (e.g., *noru* 'roe deer'), and the control primes were auditory words that were phonologically and semantically unrelated to the target (e.g., /tʃotɛ/ 'invitation'). In Exp. 2, the experimental and control primes were same as in Exp. 1, but the visual targets were Korean words that began with a plosive (e.g., *toru* 'stealing a base').

The auditory primes were recorded in two prosodic positions: in AP-initial position (e.g., [onɯl pɛun p^{h}joɦjəni #$_{AP}$ *noru*nɯn ɑnipnitɑ] 'The expression that is learned today is not *roe deer*') and in AP-medial position (e.g., [onɯl pɛun p^{h}joɦjən #$_{AP}$ i*noru*nɯn ɑnipnitɑ] 'The expression that is learned today is not this *roe deer*'). The denasalized nasal and following vowel (recorded in AP-initial position) were identity-spliced into recordings of the same word in AP-initial position or cross-spliced into recordings of the same word in AP-medial position (Fig. 1.); the pitch and duration of the following vowel were manipulated so that they would match the expected prosody of the AP-initial or AP-medial condition. Participants listened to sentences that contained the primes in AP-initial or AP-medial position and decided as quickly as possible whether the visual target (presented 50 ms after the offset of the prime) existed or did not exist in Korean. Although denasalized nasals share acoustic properties with voiced plosives and were found to be perceived as plosives in Kim's [8] VCV context, an analysis of the naturally produced primes and corresponding plosives revealed that denasalized nasals differed acoustically from AP-medial plosives.

Linear mixed-effects models were conducted on the log-transformed reaction times (Fig. 2). In Exp. 1, listeners showed sensitivity to prosodic context such that in the AP-initial condition but not in the AP-medial condition, the denasalized primes facilitated the recognition of nasal-initial words more than did the control primes. In Exp. 2, no effect of prosodic context or priming condition was found, suggesting that listeners interpreted denasalized nasals differently from plosives regardless of prosodic context (unlike the findings of Kim [5]). Given our acoustic analysis of denasalized nasals and plosives, the results from Exp. 2 suggest that acoustic differences between denasalized nasals and voiced plosives prevented denasalized nasals from being mapped into voiced plosives in lexical access.

Overall, the present results suggest that listeners are sensitive to prosodically driven fine-grained phonetic details in lexical access.


AP-initial condition

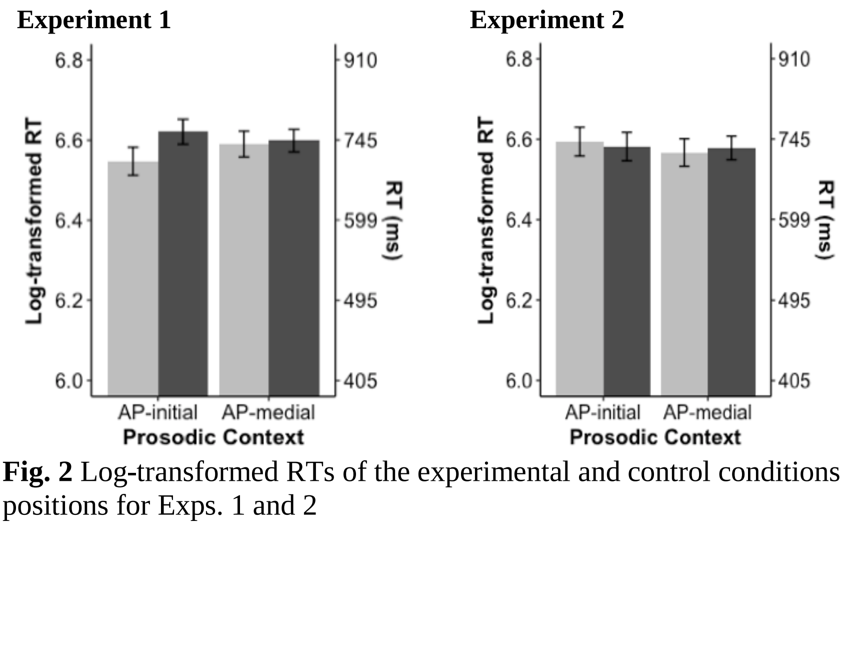
*N marks denasalized nasals

- The pitch & duration of the following vowel were manipulated to mimic those of the AP-initial condition

AP-medial condition

- The pitch & duration of the following vowel were manipulated to mimic those of the AP-medial condition

Fig. 1 Resynthesis of the stimuli for the AP-initial & AP-medial conditions


Experiment 1

Experiment 2

Fig. 2 Log-transformed RTs of the experimental and control conditions in AP-initial and AP-medial positions for Exps. 1 and 2

References

[1] Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language, 51*, 523–547.

[2] Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of native language on the use of fundamental frequency in non-native speech segmentation. *Frontiers in Psychology, 7: Phonology in the Bilingual and Bidialectal Lexicon*.

[3] Chen, M., & Clumeck, H. (1975). Denasalization in Korean: A search for universals. In C. A. Ferguson, L. M. Hyman, & J. J. Ohala (Eds.), *Nasalfest: Papers from a symposium on nasals and nasalization* (pp. 125-131). Stanford, CA: Stanford University Linguistics Dept.

[4] Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*, 155-190.

[5] Kim, Y. (2011). *An acoustic, aerodynamic and perceptual investigation of word-initial denasalization in Korean*. PhD Dissertation. University of College London.

Bilingual speakers' perception on non-native segment contrasts: a case study on Maghrebi Arabic speakers' discriminability of Korean vowels

Seung-ah Hong

Hankuk University of Foreign Studies (Korea)
essayhong@gmail.com

This study examines the perception of bilingual speakers of two different Arabic varieties and French in order to see how bilingual speakers use the vowel inventories for accommodating novel speech contrasts. Earlier studies found that a speaker has a larger vowel inventory than a target language, it is easier for him/her to facilitate novel contrasts. [1,2,3]. Based on this fact, a question is raised that how a bilingual speaker exploits his/her linguistic system under such a circumstance when they can access to both of a smaller and a bigger vowel inventories than a target language. The hypothesis is that as like a monolingual speaker, a bilingual speaker would choose a single language mode that has a larger vowel inventory.

To get an answer, a phonetic experiment was conducted on bilingual speakers of French and Maghrebi Arabic. These bilingual speakers are classified into two groups based on the Arabic dialect they use: Moroccan or Tunisian Arabic. Although both colloquial Arabic languages are derived from the same root, Classic Arabic, the vowel systems are distinct from each other; that MA has 5 monophthongs, namely /i, a, ə, u, ŭ/ and lost the duration feature [4] while TA has 7, /i, e, ä, a, ə, o, u/ and keep the vowel length as a distinctive feature [5]. Comparatively, French has more affluent monophthongs; there are 10 vowels, /i, e, ɛ, a, ɔ, o, u, y, ø, œ/ [6]. Considering these facts and the previous finding together, hypothetically bilinguals shall rely on French inventories when they discriminate Korean vowel contrasts. If so that both groups may exploit only French for the process, then there should be no difference on discriminability between the groups.

An ABX discrimination task was given to the participants. Eight contrast pairs, /ɨ-i/,/i-e/,/e-a/,/a-ʌ/,/ʌ-o/,/o-u/,/u-ɨ/,/ɨ-ʌ/, were created based on seven Korean monophthongs, /i, E, a, ʌ, o, u, ɨ/. Among them, the discriminability against two contrasts /a/ versus /ʌ/ and /o/ versus /u/ were different from the rest. Besides, there were significant discriminability differences between two groups for these contrasts. While other contrasts were clearly perceived by bilinguals across the groups, /a-ʌ/ and /o-u/ contrasts were comparatively more challenging, especially the subjects showed great difficulty to see the difference between /o/ and /u/. Comparing the discriminability of MA and TA speakers, while MA speakers had more trouble with /a-ʌ/ contrast than TA speakers (X^2(1)= 29.352, p <.05), the opposite situation was found for /o-u/ contrast (X^2(1)= 11.315, p <.05).

The result shows that the MA and TA speakers' perception is different from each other. As we supposed if the bilinguals relied solely on French categories where it is equal to the monolingual French speakers' vowel categories, then there should be no difference between two groups. However, the result was shown the opposite. This might be an evidence that French vowel inventory might be reshaped by the Arabic variety a speaker uses so even if one relied only on the French vowel categories they perceived the same thing differently since the acoustic characteristics of French vowels they hold were different. In general, we could see that Arabic can still play a role in perceiving Korean vowel contrasts. Thus, we could suggest that the effect of L1(or the dominant language in a bilingual system) is stronger than that of L2 for the perception of non-native speech contrasts even though L2 may be more beneficial than L1.

Table 1 Logit-loglinear models

#	model	L2	*df*	*p*
1	{GSC}	0.000	0	.
2	{GS}{GC}{SC}	2.140	7	.952
3	{GS}{GC}	3.571	8	.894
4	{GS}{SC}	45.536	14	.000
5	{GC}{SC}	4.419	14	.992
6	{GS}{C}	48.026	21	.001
7	{GC}{S}	6.027	15	.979
8	{SC}{G}	46.120	15	.000
9	{GC}	7.367	16	.966
10	{G}{S}{C}	48.597	22	.001

Table 2 Chi-square test result on {GC}

{GC}	Morrocan		Tunisian		Chi-square test
	correct	incorrect	correct	incorrect	
/a-ʌ/	136	80	147	21	X^2(1)= 29.352, p <.05
/a-e/	216	0	168	0	-
/ɨ-ʌ/	216	0	168	0	-
/o-ʌ/	216	0	168	0	-
/u-o/	52	164	18	150	X^2(1)= 11.315, p <.05
/u-ɨ/	216	0	167	1	$X^2$1)= 1.289, p >.05
/i-e/	215	1	168	0	X^2(1)= 0.780, p >.05
/i- ɨ/	216	0	168	0	-
total	1,483	245	1,172	172	

References

[1] Iverson, P. & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems: auditory training for native Spanish and German speakers. *Journal of Acoustical Society of America,* 126, 866-877.

[2] Mokari, P. G. & Werner, S. (2017). Comparison of cross-linguistic vowel inventory size to predict L2 vowel discrimination. *Proceedings of Seoul International Conference on Speech Science (SICSS)* (pp 126-127). Seoul, Korea: Seoul National University.

[3] Souza, Hanna Kivistö-de, Carlet, Angélica, Jułkowska, Izabela Anna, & Rato, Anabela. (2017). Vowel inventory size matters: Assessing cue-weighting in L2 vowel perception. *Ilha do Desterro, 70*(3), 33-46.

[4] Aguadé, J.(2010). On vocalism in Moroccan Arabic dialects. *The Arabic Language across the Ages*. (pp. 95-105). Wiesbaden.

[5] Maume, J.-L. (1973). L'Apprentissage du Francais chez les Arabophones Maghrebins: digglossie et plurilinguisme en Tunisie. *Langue Francaise,* 1, 90-107.

[6] Park, E.-M. (2011). Le système vocalique du français standard -structure et variation en voyelles moyennes-. *Societe Coreenne d'Enseignement de Langue et Litterature Francaises,* 36, 185-204.

Third-tone Sandhi is Incompletely Neutralizing in Perception as well as Production: Evidence from Visual World Eye-tracking

Stephen Politzer-Ahles[1], Katrina Connell[1] & Yu-Yin Hsu[1]

[1]*The Hong Kong Polytechnic University (Hong Kong)*
stephen.politzerahles@polyu.edu.hk, katrina.connell@polyu.edu.hk, yu-yin.hsu@polyu.edu.hk

Third Tone Sandhi is a phonological alternation in Mandarin Chinese whereby a syllable that is Low tone (Tone 3) underlyingly, ends up being pronounced with Rising tone (Tone 2) in certain contexts. For instance, the character 使 is normally pronounced *shi*^{L}, but in the compound word 使者 ("envoy", *shi*$^{[R]}$ *zhe*^{L}) it is pronounced with a Rising tone. A substantial amount of previous research has shown that this alternation is incompletely neutralizing, i.e., a Rising tone derived from an underlying Low tone is acoustically different than a "real" Rising tone [1,2] However, previous studies have suggested, listeners cannot perceive this difference (Peng, 1996). Such studies have been based on explicit metalinguistic judgments, in which listeners cannot accurately report a difference between the two sounds. On the other hand, some psycholinguistic studies have observed different results for sandhi-derived and non-sandhi-derived Rising tone stimuli [3,4] although such studies were not designed to test whether or not participants can accurately discriminate. Because of these conflicting sets of results, we hypothesized that listeners may be able to hear the difference between sandhi-derived and non-sandhi-derived Rising tones at the unconscious, automatic level, but not able to consciously access that for a metalinguistic judgment. We tested this with a visual world eye-tracking experiment (essentially a simplification of the design used by 3), which can track which syllables a listener is unconsciously considering even before the participant makes a behavioural response, and without requiring metalinguistic judgments.

Participants looked at arrays which included a Low-tone picture (e.g., a picture of rain, pronounced *yu*^{L} in Mandarin), a segmentally identical Rising-tone picture (e.g., a picture of a fish, pronounced *yu*^{R} in Mandarin), and two unrelated distractors. They heard instructions to click on the "*yu*^{R}"; crucially, in Underlying Rise trials this was from a natural production of a true underlyingly Rising tone, and in Derived Rise trials this was from a natural production of a sandhi-derived Rising tone. Acoustic analysis confirmed that these stimuli were physically different, as is typical in Mandarin. In the critical condition, the target words were embedded in a context that licenses tone sandhi (请将____点出来, *qing*^{L} *jiang*^{H} ____ *dian*^{L} *chu*^{H}*lai*^{R}, "Please click on the ____").

We focused on how much participants looked at the Low-tone picture when hearing the Rising-tone critical word. If participants can subconsciously perceive the difference between sandhi-derived and non-sandhi-derived Rising tones, we expect that they will look at the Low-tone picture more in the Derived condition than in the Underlying condition.

We report data from a preliminary sample of 6 native speakers (of a planned 50). We replicated the common finding that participants do not distinguish these tones behaviourally: as shown in Figure 1, half of the participants clicked on the Low-tone target more when hearing an underlying Rising tone, and half clicked on it more when hearing a derived Rising tone, suggesting that participants were not reliably conscious of this difference. However, their eye movements were reliably different between conditions. As shown in Figure 2, from about 600 ms after hearing the critical word (adjusted for the 200 ms required to program saccades), participants looked more at the Low-tone target when they heard a Rising tone that was actually derived from a Low tone, compared to when they heard a true underlyingly Rising tone. As shown in Figure 3, this pattern was observed in all participants. This suggests that participants do hear a difference between sandhi-derived and underlying Rising tones, even if they cannot access it behaviourally.

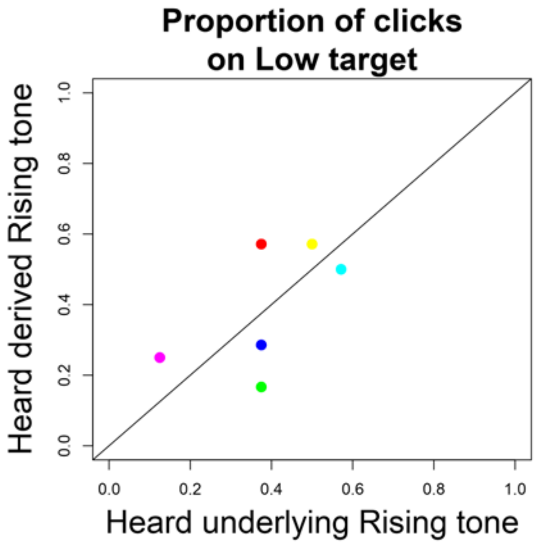


Fig.1 Behavioural results. The diagonal line indicates zero difference between conditions

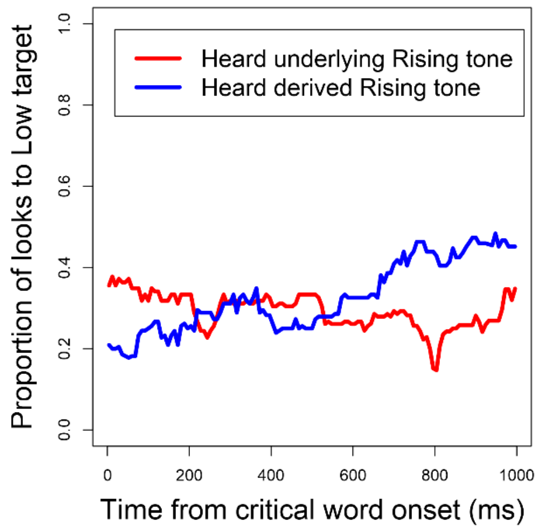


Fig.2 Eye-tracking grand averages

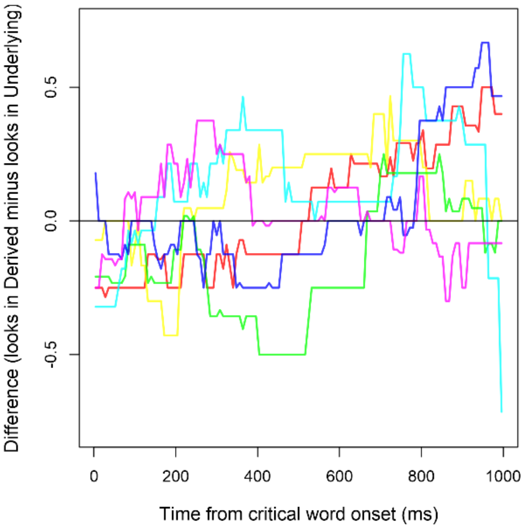


Fig.3 Eye-tracking individual results. Lines above zero indicate more looks to the Low target when hearing a derived Rising tone compared to when hearing an underlying Rising tone.

References

[1] Peng, S. (1996). *Phonetic implementation and perception of place coarticulation and tone sandhi*. Ph.D. dissertation, The Ohio State University.

[2] Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology, 27*, 153-201.

[3] Speer, S., & Xu, L. (2008). Processing lexical tone in third-tone sandhi. *Talk presented at Laboratory Phonology 11*

[4] Zhou, X., & Marslen-Wilson, W. (1997). The abstractness of phonological representation in the Chinese mental lexicon. Ed. Chen, H., *Cognitive Processing of Chinese and Related Asian Languages*, 3-27. The Chinese University Press.

Syllabification of consonant clusters by L1 Japanese L2 English speakers

James Whang[1,2]

[1]*MARCS Institute for Brain, Behaviour & Development (Australia)*
[2]*ARC Centre of Excellence for the Dynamics of Language (Australia)*
research@jameswhang.net

Introduction: The current study investigates the influences of second language (L2) proficiency on the phonological processing of loanwords in the speakers' native language (L1). In standard modern Japanese, unaccented high vowels /i, u/ optionally become deleted between voiceless obstruents [1], especially when the vowel is phonotactically predictable [2], resulting in surface consonant clusters (e.g. /masutaa/ → [mastaa] 'master'). There is debate over whether the stranded onset consonant forms a consonantal syllable by itself ([ma.s.taa]; [3]) or resyllabifies into the following syllable ([ma.staa]; [4]). The current study presents acoustic evidence that proficient L2 English speakers do seem to allow tautosyllabic clusters even when speaking L1 Japanese.

Methods: A pilot experiment was conducted with eight native Japanese L2 English speakers (four women and four men) who were born and raised in the Tokyo area. All participants were international students in New York City, who had resided in the United States for a minimum of 12 months and a maximum of five years. The stimuli were limited to lexical tokens containing /suC$_2$/ sequences because [sC$_2$] is the only legal tautosyllabic sequence in English that can also result in Japanese from high vowel deletion. The stimuli were grouped into non-devoicing native Japanese words (e.g. /sugoi/ 'amazing'), devoicing native Japanese words (e.g. /sukuu/ 'to rescue'), and devoicing loanwords (e.g. /sukii/ 'ski'). The stimuli were embedded in meaningful and unique carrier sentences, which the participants read aloud in a sound-attenuated booth. The recorded data were segmented using Praat, from which segment duration measurements were taken.

Predictions: Consonants in tautosyllabic onset clusters tend to be shorter than when in singleton onsets due to c-center effects [5], where the center of the onset, whether singleton or complex, is coordinated with the nucleic vowel (see Figure 1 below). The prediction going into the study, therefore, was that (i) /s/ duration should be longest in non-devoicing tokens, since the consonant unambiguously forms a singleton onset that syllabifies with the following vowel and that (ii) /s/ duration in devoiced tokens would be shorter than in non-devoiced tokens should there be resyllabification. Even a short exposure to L2 is known to result in phonetic drift, where L1 segments are produced more L2-like [6], and since the participants resided in the United States and actively used English for at least a year, the L2 English speakers could produce more English-like clusters with shorter /s/ durations.

Results: A linear mixed effects regression model was fit to the duration data with token type (i.e. native non-devoicing, native devoicing, and loanword devoicing), gender, and their interaction as predictors. Random intercepts by participant and word were also included in the model. Gender and the interaction terms did not have a significant effect. Token type, however, did have a significant effect. /s/ duration was longest in non-devoicing tokens as predicted, and /s/ duration was also significantly shorter in devoiced tokens, but only in loanwords. The comparison is shown below in Figure 2.

Discussion: In an electromagnetic articulography (EMA) study with Japanese monolinguals, Shaw and Kawahara [7] found no c-center effects, suggesting that a stranded onset consonant forms a consonantal syllable by itself. The acoustic results of the current study suggest the opposite, however, perhaps because the participants in the current study have had prolonged exposure to English, which allows numerous /s/-initial tautosyllabic clusters. The difference between native and loanwords may be related to how the two are represented, where loanwords could be represented with an underlying cluster (i.e. without an intervening, epenthetic vowel) due to the participants' familiarity with the source language. The current pilot study results suggest that proficient L2 English speakers allow tautosyllabic clusters even when speaking L1 Japanese, and perhaps specifically when producing English loanwords.

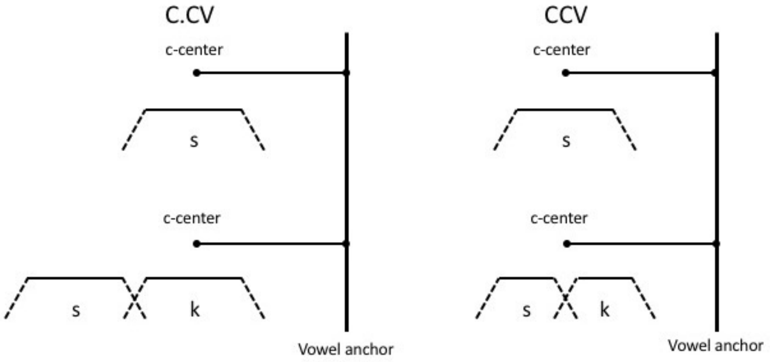


Fig.1 C-center effect in heterosyllabic (left) vs. tautosyllabic (right) clusters

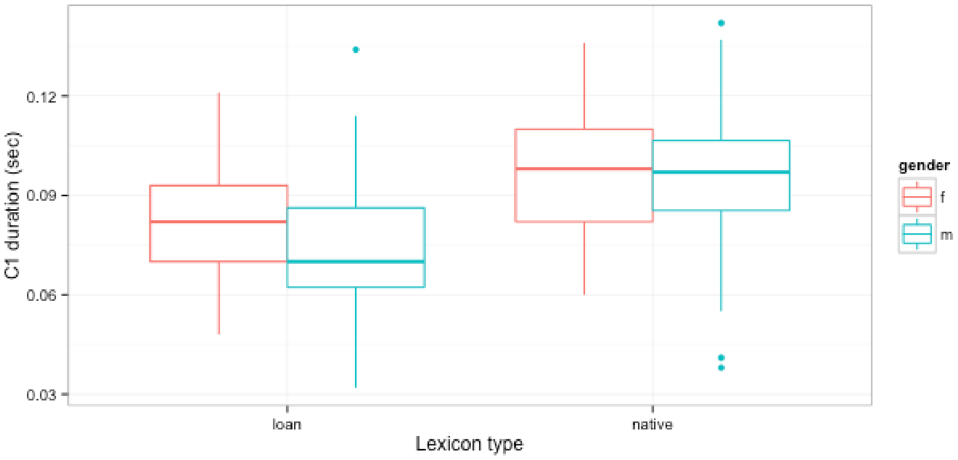


Fig.2 Duration of /s/ by lexical type and gender

References

[1] Shaw, J. & Kawahara, S. (2018a). The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics*. 66:100–119.

[2] Whang, J. (2018). Recoverability-driven coarticulation: Acoustic evidence from Japanese high vowel devoicing. *Journal of the Acoustical Society of America* 143(2): 1159-1172.

[3] Matsui, M. F. (2017). On the input information of the C/D model for vowel devoicing in Japanese. *Journal of the Phonetic Society of Japan* 21(1), 127-140.

[4] Kondo, M. (1997). *Mechanisms of vowel devoicing in Japanese*. (Ph.D.), Edinburgh, Edinburgh, UK.

[5] Browman, C. P., & Goldstein, L. (1988). "Some Notes on Syllable Structure in Articulatory Phonology". *Phonetica* 45:140-155.

[6] Chang, C. B. (2010). *First language phonetic drift during second language acquisition*. (Ph.D.), UC Berkeley.

[7] Shaw, J. & Kawahara, S. (2018b). Consequences of High Vowel Deletion for Syllabification in Japanese. In G. Gallagher, M. Gouskova & S. H. Yin (eds.), *Proceedings of the 2017 Annual Meeting on Phonology (AMP)*. New York, NY, USA.

Sound symbolism in speech directed to children by Korean mothers

Jinyoung Jo[1], Eon-Suk Ko[2]

[1]Seoul National University (Korea), [2]Chosun Univ. (Korea)
jinyoungjo710@gmail.com, eonsukko@chosun.ac.kr

Sound symbolism refers to the non-arbitrary relationship between sound and meaning of a word (e.g. *bang*). The inherent echoic relation to a referent in sound symbolism might facilitate infants' word learning [1], in part because sound symbolic words are acoustically more salient than conventional words [2]. This study investigates Korean mothers' use of sound symbolism concerning the frequency and acoustic saliency of these iconic words as a function of child age based on naturally occurring data.

Our analysis focuses on expressive lengthening, ideophones and word play. In expressive lengthening (e.g. k^{h}*ɨ::n* 'huge'), an extra elongation of the vowel indicated by the double colons augments the scalar properties of the meaning. Ideophonic words depict a wide range of sensory experiences, either auditory (e.g. *məŋməŋ* 'woof woof') or non-auditory (e.g. *tuŋkɨltuŋkɨl* 'round'). The ratio of ideophones in Korean child-directed speech is reported to be particularly high compared to English or even to Japanese [3], which makes Korean an advantageous test bed for investigating the use of sound symbolism. Word play includes cases where mothers playfully produce nonsense sounds mainly to grab the child's attention.

A total of 36 infant-mother dyads (child age: 0, 1 and 2 years old) in a 40-minute free-play session were recorded. Their speech was transcribed in which lengthened syllables, ideophones and word play were tagged using CHILDES convention [4]. We first calculated the ratio of these target words in each mother's word tokens. Three linear regression analyses were performed with age and gender of child as fixed factors and the ratio of words with expressive lengthening, ideophones and word play as dependent variables in each model. The results showed a significant effect of child age on the ratio of sound symbolism in all three categories (all *p*'s<0.01; Figure 1), indicating a decrease in the use of sound symbolism with child age.

We further investigated prosodic saliency of words containing expressive lengthening and ideophonic expressions. Duration, mean pitch and pitch range of 641 words with expressive lengthening and 1460 ideophonic words were compared with a randomly generated sample of 1885 ordinary words whose boundaries were marked by a forced-alignment toolkit. Using the lmerTest package [5] of R, mixed effects linear regression models were constructed with Word Type (i.e., expressive lengthening, ideophone, and ordinary words), Age (i.e., 0, 1, and 2) and their interaction as fixed factors, and the three acoustic measures as dependent variables. It was found that both types of iconic words were significantly longer in duration and higher in pitch than ordinary words (*p*'s<0.05). The effect of Age was non-significant in all acoustic measures (*p*'s>0.05). The interaction between Word Type and Age was only significant for mean pitch; a significant Word Type (ideophone) × Age (2) interaction term indicated that the acoustic saliency (i.e., higher pitch) of ideophones became weaker with child age (*p*<0.05; Figure 2). Thus, the findings suggest that sound symbolic words were perceptually more salient than non-iconic words and that such saliency persisted throughout the age range of children examined in this study, except for the less enhanced mean pitch of ideophones in age 2.

A decrease in the frequency of sound symbolism with child age is consistent with the claim that sound symbolism might facilitate early word learning [1], which predicts that the use of sound symbolism in child-directed speech is modulated by the maturity of child's ability to associate linguistic form with meaning. This is in line with previous studies that reported attention-grabbing function of those words is crucial only when infants are younger [6] and that sound symbolic words are less useful for learning to make fine-grained distinction among similar concepts [7]. The present study, however, also suggests that older children might still benefit from the prosodic saliency of these iconic words, evidenced by the finding that their acoustic saliency remained quite robust in the mothers' speech directed to older children.

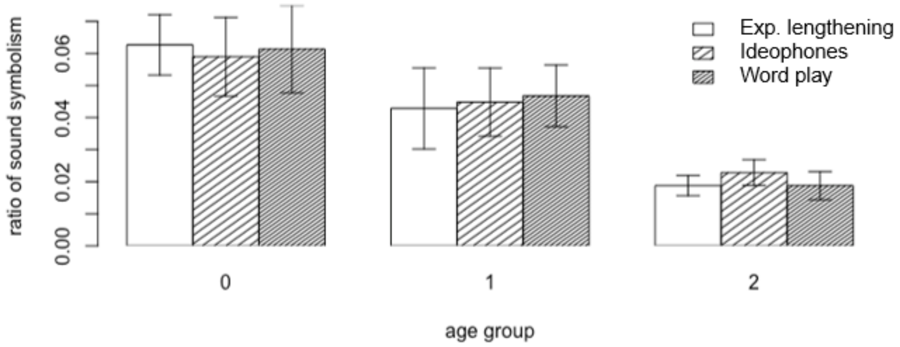


Fig.1 Ratio of sound symbolic words

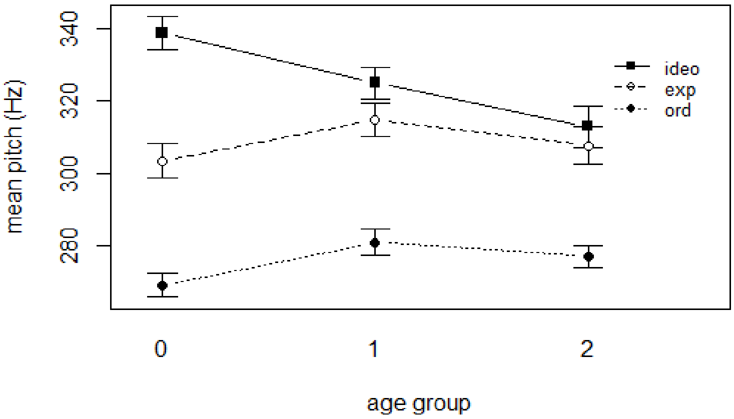


Fig.2 An interaction of Word Type and Age on mean pitch

References

[1] Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society*, *369*, 20130298.

[2] Laing, C. E., Vihman, M., & Keren-Portnoy, T. (2016). How salient are onomatopoeia in the early input？A prosodic analysis of infant-directed speech. *Journal of Child Language*, 231–268.

[3] Bae, S. B., & Park, H. W. (2012). Korean children's use of onomatopoeic and mimetic words. *The Korean Journal of Developmental Psychology*, *25*(1), 101–115.

[4] MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd editio). Mahwah, NJ: Lawrence Erlbaum Associates

[5] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

[6] Kauschke, C., & Klann-delius, G. (2007). Characteristics of maternal input in relation to vocabulary development in children learning German. In I. Gülzow & N. Gagarina (Eds.), *Frequency effects in language acquisition: defining the limits of frequency as an explanatory concept* (pp. 181–204). Berlin: Walter de Gruyter.

[7] Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *38*(5), 1152–1164.

The interaction between prosodic focus and phrasal tone in South Kyungsang Korean

Yong-Cheol Lee[1], Dongyoung Kim[2] & Sunghye Cho[3]

[1]Cheongju University (Korea), [2]Yonsei University (Korea), [3]University of Pennsylvania (USA)
soongdora@gmail.com, deewaikim@gmail.com, paae0928@gmail.com

We examined the production and perception of corrective focus in South Kyungsang Korean, using phone number strings. We found that prosodic marking of focus varies depending on the tonal patterns within a phrase. When a phrase-initial digit was in focus, a phrase beginning with a High tone (HLL and HHL) was clearly marked in production and accurately recognized in perception, compared to a phrase beginning with a Low tone (LHH and LHL). Furthermore, a HLL tonal pattern was perceived better than HHL. Unlike the HHL tonal pattern, where the focused High tone contradicted with the second High tone within a phrase, the focused High tone in HLL was clearly distinct from the other tones. This study demonstrates that prosodic marking of focus is not universal but depends on the prosodic structure, even within a single language, thus calling for a broader study involving typologically similar languages.

South Kyungsang Korean allows the following four tonal patterns in a trisyllabic word: HHL, HLL, LHL, and LHH. Given that focus has a multiplicative effect on pitch (Liberman & Pierrehumbert, 1984), it is likely that higher pitches will demonstrate a greater impact than lower pitches for focus prosody. Thus, we hypothesize that when a phrase-initial tone is focused, a phrase beginning with a High tone has the advantage of marking prosodic focus and results in a better identification rate in perception than a phrase beginning with a Low tone. In addition, since prosodic marking of focus spreads across the entire phrase in this language (Lee, 2015, 2017), a HHL tonal pattern would result in less accurate perception than HLL. Taken together, the tonal patterns under focus would be ranked in the order of perceptual accuracy as follows: HLL > HHL > LHT, where the symbol '>' means a higher rate of identification and the 'T' in LHT is either Low or High.

We used 10-digit number strings, read as connected individual digits grouped as (NNN)-(NNN)-(NNNN). We created sets of 100 10-digit sequences, designed so that each digit occurs ten times in each sequential position, and each pair of digits occurs once in each adjacent pair of positions. Five native speakers of South Kyungsang Korean (2 males; 3 females) read these digit strings in isolation as a background "broad-focus" condition, and in a Q&A dialogue for corrective focus, where someone asks for a phone number in which one of the digits is incorrect, and the speaker answers, correcting the wrong digit ("No, the number is 215-417-5623"). In a perception experiment, which was conducted with an online software Qualtrics, 40 listeners heard only the phrase with the correction, and were asked to identify which digit was corrected.

Among the 100 digit strings, we only included phrase-initial positions in three-digit groups (that is, NNN) for further analysis. The phone-number strings in the broad-focus condition were directly compared with the same sequences in the corrective-focus condition by the aggregate measures of mean pitch. Figure 1 illustrates prosodic modulation by focus in terms of mean pitch, separated by tonal patterns within a phrase, when focus was in the phrase-initial position. The HHL tonal pattern showed that, although the focused High tone showed a clear increase in mean pitch when compared to the same tone under broad focus, its prosodic marking of focus was confusing within a phrase because the subsequent tone also showed a similar degree of pitch increase. In contrast, the focused High tone of the HLL tonal pattern was clearly distinct from the broad-focus counterpart as well as the subsequent Low tones within a phrase. Finally, the focused Low tone of the LHT tonal pattern demonstrated a marginal pitch increase and its pitch value was much lower than the next High tone. The perception results support our hypothesis that HHL showed the highest identification rate (about 82%), followed by HHL (about 45%) and LHT (about 29%).

This study demonstrated that, in marking prosodic focus, a certain tonal pattern is more favorable than others within a phrase. The different modes of focus prosody seem to reflect basic typological features of languages similar to South Kyungsang Korean, where phrasal tones are determined by lexical pitch accent. The precise nature, however, will be clarified by examination of additional closely related languages.

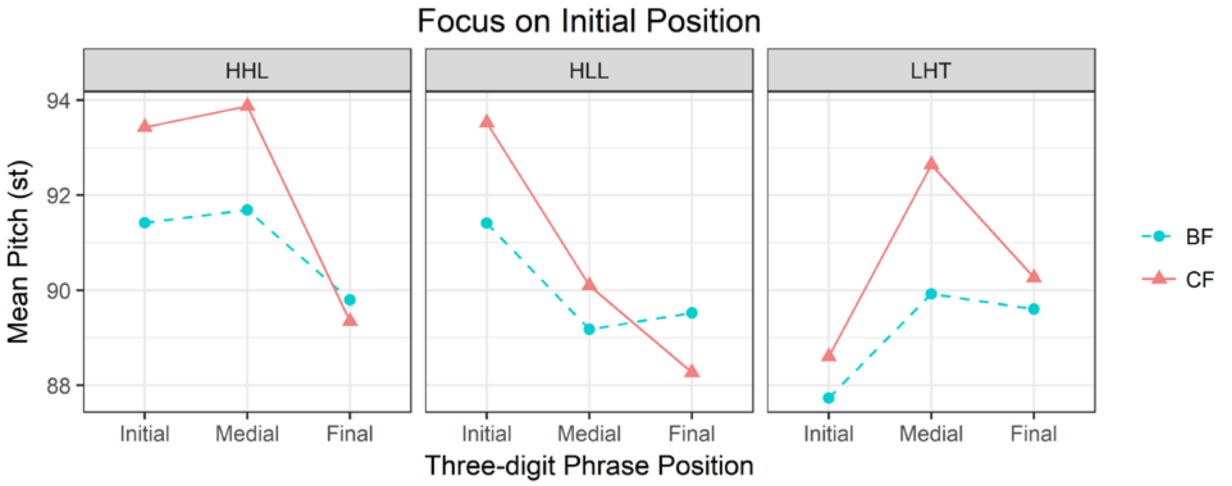


Fig.3 Mean pitch in the three positions within a phrase in the two focus conditions (BF: broad focus; CF: corrective focus)

References

[1] Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (eds.), *Language sound structure* (pp 157–233). Cambridge: MIT Press.

[2] Lee, Y. (2015). *Prosodic focus within and across languages*. PhD thesis, University of Pennsylvania.

[3] Lee, Y. (2017). Prosodic focus in Seoul Korean and South Kyungsang Korean. *Linguistic Research*, 34(1), 133–161.

Poster Presentations (Day 2)

Proposing a Method for Quantifying Speech Prosody: Some Insights from a Singing Workshop*

Hang Chan

Hong Kong Baptist University
joshuachan@hkbu.edu.hk

Every sound is made up of pitch, intensity, and length (P, I, and L). These universal parameters work together to give a sound its aural sensation. The three dimensions correspond to the acoustic parameters of a wave form: fundamental frequency (frequency of vibrations), amplitude (height of a vibration cycle), and duration of sound displacement (Roederer, 2008). This presentation will introduce a case of using P, I, and L fluctuations to appraise the effects of a prosodic training activity. The main question this project explores is whether or not singing to a song, among other benefits, can cause P, I, and L to vary in certain ways, therefore bringing about suprasegmental level of changes. The song used was ‘I Could Have Danced All Night’ from the musical *My Fair Lady*, on the theme of learning English. A musician with 15 years of coaching experience advised on the use of the song and a vocal singer from Hong Kong Youth Windophilics conducted the singing session.

This study considers a singing session as a natural context for practising P, I, and L. As a popular language arts activity, singing has wide appeal in public domains (e.g., websites), news reports, and teacher-oriented literature; it is often believed to be a good way to introduce melodic sound patterns to learners (e.g., Celce-Murcia, Brinton, & Goodwin, 2010; Kenworthy, 1987; Brown, 2012). To-date, most empirical studies involving the use of songs and experimental setups have focused on how songs aid lyric recall (Good, Russo, & Sullivan, 2015; Ludke, Ferreira, & Overy, 2014; Lehmann & Seufert, 2018; Racette & Peretz, 2007). Arguably, a focus on lyric recall (i.e., memory effects) produces quite tangible learning outcomes (i.e., whether a learner remembers more words after singing). What the present study intends to measure are the more implicit prosodic changes occurring in one’s voice after singing, that is, the P, I, and L variations.

Research Questions

Assessing Production	**RQ1:** Before singing, how will the learners use P, I, and L in a reading-aloud test?
	RQ2: After singing, will a second trial of the production test continue to reveal the same P, I, and L patterns?
Assessing Perception	**RQ3:** Which cues, P, I, or L, do the learners judge as representing stress?

Thirty-two Cantonese students (30F, 2M, with a mean age of 15.09 [SD .53]) from seven local schools participated in the singing session. The comparison group consisted of 41 English native speakers (25F; 16M, with a mean age of 20.73 [SD 4.71]) studying in different programmes at the University of Queensland, Australia. These native speakers did not undergo the singing session but they completed the tests. The research instruments included a perception test and a production test. The key question to explore via the production test (a “lyric reading” test) is whether or not singing along to a song might cause P, I, and L to be used in certain ways, while the perception test (an “auditory discriminatory test”) was administered to assess the learners’ subjective preferences of P, I, and L. The perception-production relationship aims to reveal whether or not what learners perceive as important could become what they use in producing speech.

[*] The full paper of this presentation is available online: Chan, H. (2018). *A method of prosodic assessment: Insights from a singing workshop*. Cogent Education (by Taylor and Francis). doi: 10.1080/2331186X.2018.146
1047.

Findings

Two major findings were yielded: First, while the learners judged pitch variation (P) to be important in the perception test, they relied on length variation (L) when encoding prosody in the production test. Second, singing to a song did not change the fact that length variation (L) was a dominant encoder, and pitch (P) only came second to length. The native English speakers (i.e., the control group) also used L to a much greater extent than P or I. These findings can lead to several possibilities: They may indicate that singing could affect prosody in other ways, but not how P, I, and L are varied in the voice; or, indeed, they may point to a "normal" way of encoding prosody, with L variation being the most important feature for distinguishing sounds, followed by P and then I variations.

Interpretations

Both the learners' and the native speakers' reliance on L may be explained by the nature of the production task itself. The current participants took the lyric reading test in a quiet room, where there was no need to use one's voice in extreme ways. Ladefoged (2003, p. 93) and Levis (1999, p. 43) caution against treating pitch as the only correlate of stress, as length is substitutable for pitch on many occasions. As for the perception-production mismatch (i.e., the participants' high sensitivity to pitch but more length variations in production), one explanation could be that there exist fundamental differences between "perceptual sensors" and "articulatory motors". Common experience suggests that someone being alerted by a shrill sound can hardly imply that the person emits such sounds at all times. The present results just show the importance of researching "productive prosody" in its own right in order to understand how speech is encoded.

Educational Implications

1. By measuring the vicissitudes of P, I, and L in one's voice, the current study departs from a consideration of a song's mnemonic effects, but focuses on other benefits of singing, that is, the subtle changes in one's voice.
2. Singing is just one of many pedagogical options for prosodic training. The proposed method may be useful for assessing other forms of prosodic training.

References

[1] Roederer, J. G. (2008). *The physics and psychophysics of music: An introduction* (Vol. 4). New York: Springer.

[2] Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (Vol. 2). New York, NY: Cambridge University Press.

[3] Kenworthy, J. (1987). *Teaching English pronunciation*. London: Longman.

[4] Brown, J. D. (2012). *New ways in teaching connected speech*. Alexandria, VA: TESOL International Association.

[5] Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music, 43*, 5, 627-640. doi: 10.1177/0305735614528833.

[6] Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory Cognition, 42*, 1, 41. doi: 10.3758/s13421-013-0342-5.

[7] Lehmann, J. A. M., & Seufert, T. (2018). Can music foster learning - Effects of different text modalities on learning and information retrieval. *Frontiers in Psychology, 8*. doi: 10.3389/fpsyg.2017.02305.

[8] Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition, 35*, 2, 242-253.

[9] Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Blackwell Publishing.

[10] Levis, J. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly, 33*, 1, 37-63.

A cross-linguistic comparison of word teaching strategies between Korean- and English-speaking mothers

Jihyo Kim[1] & Eon-Suk Ko[2]

[1]Chosun University (Korea), [2]Chosun University (Korea)
jeeeeehyo@gmail.com, eonsuk@gmail.com

Mothers across cultures adopt similar acoustic characteristics in addressing their children but differ in their strategies in teaching words. For example, differences in English as "noun-dominant" relative to more "verb-dominant" languages like Korean and Japanese could yield variations in maternal speech style [1]. This paper investigates the effects of structural differences in Korean as a head-final and English as a head-initial language on the mothers' word teaching strategies. We show that Korean mothers often present target words in the utterance-initial position, but they also repeat target words at the end of an utterance significantly more frequently than American mothers.

When introducing new words, American mothers place them on exaggerated pitch peaks in the utterance-final position [2]. Cross-linguistically, the right-edge of an utterance is a privileged position, with prominence rendered by utterance-final lengthening and exaggerated pitch movement. This position, however, is usually occupied by a verb in Korean. This would mean that Korean-learning infants do not benefit from the right-edge prominence in learning nouns. Korean mothers, however, might still take advantage of the prominence of the utterance-final position by, for example, repeating target nouns at the end of an utterance similar to the English tag-question. We thus investigated how Korean mothers present focused words to their children in comparison to American mothers.

We extracted the target words from corpora Korean and English and coded them with regard to their position in the utterance and the presence of repetition. The Korean corpus was constructed by recording 35 mothers spontaneously interacting with their children of three age group: 0;9 (12 dyads), 1;1 (11 dyads), and 2;3 (12 dyads). Matching data of American mother-child dyads were selected from the Brent, Brown, Post, and Providence corpus in the CHILDES database [3]. We selected 10 target nouns for analysis based on frequency in each mother-child data. We then coded the position (isolated, utterance-initial, utterance-final) and the presence of repetition (tagged, within-utterance, nearby-utterance). A total of 2812 Korean and 3774 English utterances were coded.

We conducted multiple regression analyses to test if age and language significantly predicts mothers' word teaching strategies. The results indicates that age and language significantly predicts the use of repetition (R^2=.35, F(3, 66)=11.9, p<0.001). Korean had a higher rate of repetition than English (β=21.9, p<0.01), and it decreased with age (β=-8.2, p<0.05). Both the ratio of tag-repetition and within-utterance repetition was explained by language, with Korean having a significantly higher rate of them than English (p<0.001). An interaction between age and language (β=-3.7, p<0.001) indicates that the ratio of tag-repetition is higher in speech to Korean younger children (Figure 1). The placement of focused words in the utterance-initial position is explained by language (R^2=.76, F(3, 66)=70.59, p<0.001). The placement of focus at the right-edge, however, was explained both by language and age (R^2=.83, F(3, 66)=106.2, p<0.001). An interaction between language and age (β=4.9, p<0.5) indicates that the ratio of focused words occurring at the right-edge is lower in speech to Korean older children (Figure 2).

It thus suggests that Korean mothers do make efforts to take advantage of the edge-prominence of the final position by repeating target words at the end of the utterance because making use of the right-edge of an utterance to teach focused nouns is hardly possible in Korean. American mothers, however, rarely repeat focused words within the utterance or at the end of the utterance at all because they can rather place target words at the right-edge of an utterance, making full use of the edge-prominence of the utterance-final position.

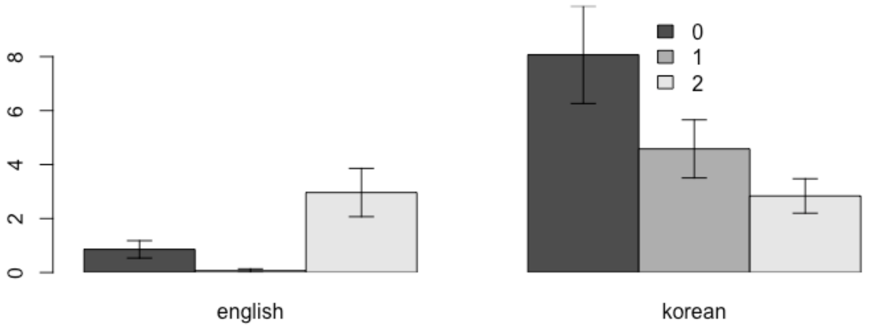


Fig.1 Ratio of nouns with a tag-repetition in English and Korean child-directed speech. Ages 0, 1, and 2 refer to groups of children aged about 9, 13, and 28 months, respectively.

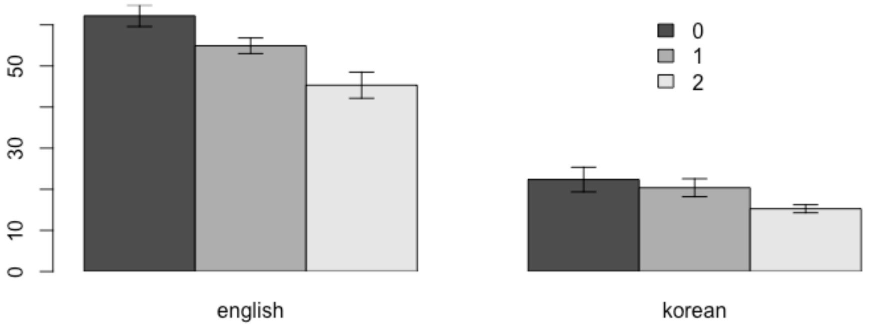


Fig.2 Ratio of nouns at the right-edge of an utterance in English and Korean child-directed speech. Ages 0, 1, and 2 refer to groups of children aged about 9, 13, and 28 months, respectively.

References

[1] Gopnik, A., & Choi, S. 1990. Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language* 10, 199-215.

[2] Fernald, A., & Mazzie, C. 1991. Prosody and Focus in Speech to Infants and Adults. *Developmental Psychology* 27, 2, 209-221.

[3] MacWhinney, B. 2000. The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mandarin monosyllables trigger long-lag identity priming but not long-lag morphological priming

Stephen Politzer-Ahles[1], Lei Pan[1] & Jueyao Lin[1]

[1]*The Hong Kong Polytechnic University (Hong Kong)*
sjpolit@polyu.edu.hk, bernice.pan@polyu.edu.hk, hansy.lin@polyu.edu.hk

One powerful technique for investigating phonological and other relationships between words in the mental lexicon is priming. Evidence from many previous experiments has suggested that phonological relationships between words are represented differently than morphological relationships; for example, in long-lag priming experiments (when many trials intervene between primes and targets), identical or morphologically-related prime-target pairs generally elicit facilitative priming, whereas phonologically-related pairs do not [1, among others]. This method has mainly been used with Indo-European languages, however, and it is unknown to what extent these patterns of results may extend to other languages with very different typological properties, such as Mandarin Chinese, which uses lexical tone to distinguish words and which has substantial homophony.

We conducted a pre-registered [https://osf.io/d3xu5/], high-powered long-lag priming experiment (N=153 native Mandarin speakers, 96 item sets) to compare reaction times for auditorily presented monosyllable targets preceded by unrelated auditory primes (e.g., *hua*^{4}…*shi*^{3} [numbers represent tone categories]) against those preceded by segmentally-related but morphologically-unrelated primes (e.g., *shi*^{1}…*shi*^{3}; superscript numbers indicate lexical tone categories) and those preceded by segmentally- and morphologically-related primes (e.g., *shi*^{2}…*shi*^{3}); because of a systematic phonological alternation (third tone sandhi [3]), *shi*^{2} is a possible allomorph of *shi*^{3} in certain phonological contexts. (See Table 1 for example.) The segmentally-related pairs shared only a phonological relationship, without a morphological relationship. Critical targets were all in either Mandarin tone 2 or tone 3, and fillers included 48 prime-target pairs with tone 1 or tone 4 targets, and 288 nonwords (phonologically legal Mandarin accidental gaps). Each target was separated from its prime by 18-52 intervening trials, and participants made a speeded lexical decision to every stimulus. In immediate auditory priming experiments (e.g., [3]), segmentally-related pairs like *shi*^{1}…*shi*^{3} typically elicit priming, but we expected that these sorts of phonologically-related pairs would not elicit long-lag priming unless they also shared a morphological relationship (*shi*^{2}…*shi*^{3}). Thus, we predicted a larger facilitative priming effect for morphologically-related pairs than for merely segmentally-related pairs. But contrary to our expectation, both kinds of related primes were associated with slower, rather than faster, reaction times on their targets (unrelated: 1066ms; segmentally-related: 1071ms; segmentally-and-morphologically-related: 1084ms); see Figure 1. Morphologically related primes in particular triggered a large inhibition effect (18ms). This finding is surprising because long-lag priming typically engenders facilitative effects (for identical or morphologically related prime-target pairs) or no reliable differences for other types of prime-target relationships.

We then ran a pre-registered high-powered replication study [https://osf.io/f8tb9/] to test whether this inhibition pattern could be replicated, and—as a manipulation check—whether long-lag identity priming works at all in Mandarin monosyllables. Preliminary results from 95 participants (right-hand side of Figure 1) replicated the lack of priming for morphologically related targets (1032ms, compared to 1031 ms for unrelated targets), while also revealing a robust priming effect for identity targets (962ms, 65ms faster than the unrelated condition).

These results suggest interesting possibilities about the nature of lexical representations and priming across languages. We hypothesize that long-lag priming may depend on being able to uniquely identify and activate a particular morpheme, which was generally not possible in this experiment: our materials were Mandarin monosyllables, the vast majority of which have a large number of homophones (for example, a participant hearing *shi*^{3} may activate [the morphemes written as] 始 ["start"], 史 ["history"], 驶 ["drive"], or 矢 ["arrow"], among others).

Accordingly, it is possible that the identity priming in the present study was episodic rather than linguistic priming [1], and ambiguous Mandarin monosyllables do not elicit morphological priming; further research will be needed to test these two hypotheses.

Table1 Sample stimulus sets

Target	Unrelated prime	Morphologically- and segmentally- related prime	Segmentally-related prime
shi^3	*hua*^{4}	*shi*^{2}	*shi*^{1}
zao^3	*hun*^{1}	*zao*^{2}	*zao*^{4}
pin^2	*bu*^{4}	*pin*^{3}	*pin*^{1}
lian^2	*yue*^{1}	*lian*^{3}	*lian*^{4}

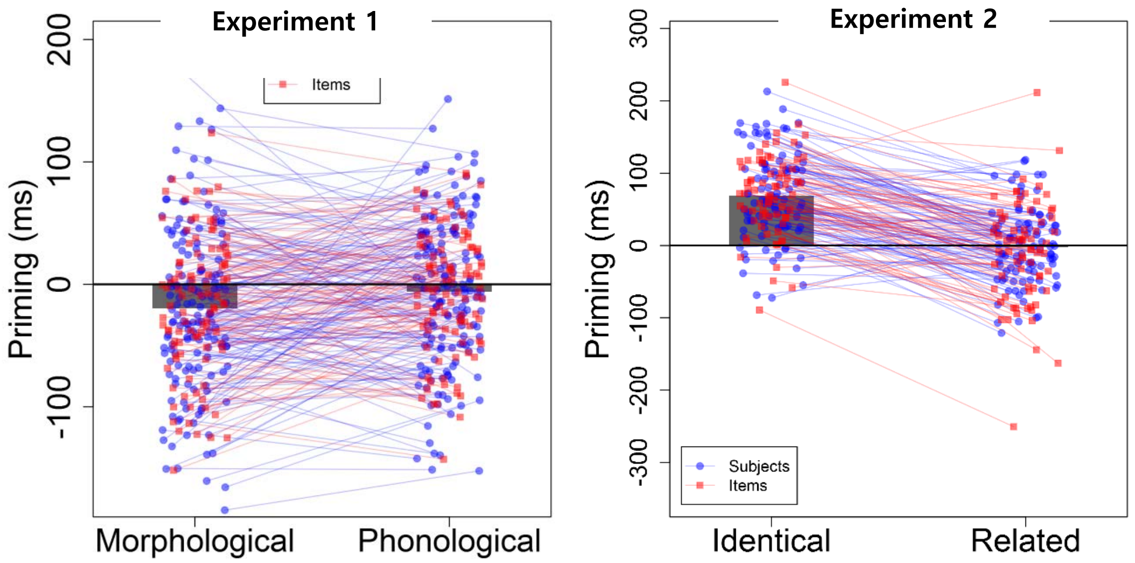


Fig.1 Data by participants and by items. Gray bars represent the mean priming effects (unrelated condition minus related condition)

References

[1] Kouider, S., & Dupoux, E. (2009). Episodic accessibility and morphological processing: evidence from long-term auditory priming. *Acta Psychologica*, *130*, 38-47.

[2] Chen, M. (2000). *Tone sandhi: patterns across Chinese dialects*. Cambridge: Cambridge University Press.

[3] Sereno, J., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, *58*, 131-151.

Linguistic and Social Dimensions of Denasalization in Seoul Korean

Hankyul Kim

Cornell University (USA)
hk783@cornell.edu

Recent studies raise suspicion about a property of nasal consonants in Korean. An interesting difference is developing between word initial and medial nasals. While both are described as being sonorous nasals, some evidence suggests that in word initial position there is denasalization. Recently the author has observed that non-native speakers' judgments, especially from English speakers who have a voicing contrast, suggest that word initial nasals and medial nasals do not show the same pattern. When English speakers hear a Korean word with a word initial nasal, they often perceive the nasal as a corresponding voiced stop, /m/>[b] and /n/>[d]. As this happened repeatedly in colloquial use of the language, this is clearly a phenomenon that requires fuller investigation.

Mention of denasalization based on the authors' impressionistic judgments goes back to [1] and a handful of studies have followed on this issue [2, 3] Considering the fact that this issue had been observed since 1924 but only small number of instrumental studies have been published, it is reasonable to question whether denasalization has been around us but ignored, or it is a recently developing change. [4] investigates this question for Busan Korean finding that older speakers produced 100% or near 100% sonorant nasals while younger speakers produce denasalized realizations with inter-speaker variation. What about Seoul Korean speakers? Do they also show a generational difference?

This also raises the issue of phonological conditioning. [5] show that Korean nasals are longer in duration and weaker in overall energy in higher domain-initial positions but do not directly address the issue of degree of nasality. What is the prosodic conditioning and how does it interact with this ongoing change? The goal in this study is to advance our understanding of denasalization by investigating speech data from native speakers of Seoul Korean regarding phonological (prosodic domain) and social-indexical (age) factors.

Participants for a phonetic production experiment include native speakers of Seoul Korean divided by age with older speakers over 60 years old and younger speakers under 30 years old. Participants read reading materials including short passages from stories, a dialogue, and short sentences. Nasal stimuli are distributed according to prosodic domains. Results indeed show some cases of denasalization. Figure 1 below shows a clear comparison of a denasalized segment and a typical nasal segment. A target word /manula-lɯl/ forms an Accentual Phrase (AP) with four syllables within a sentence. Both nasals are preceded and followed by vowels. However, while the word-medial [n] is fully sonorous nasal with a clear voicing bar and formant structures accompanied with antiresonances around 1,000Hz, denasalized [m͊] shows burst like realization after 50ms closure that even lacks voicing. The results of the current study to date show different rates of denasalization between old speakers and young speakers; young speakers are more likely to produce denasalized nasals than old speakers.

We report on the interaction of duration and denasalization at all levels of the prosodic hierarchy. Results to date suggest a more complicated pattern for denasalization than found by [5] for duration and overall energy. One striking result is a very short denasalized segment in utterance initial position as shown in Figure 2. Interestingly, this allophone might account for Cho and Keating's finding of shorter duration in utterance initial position.

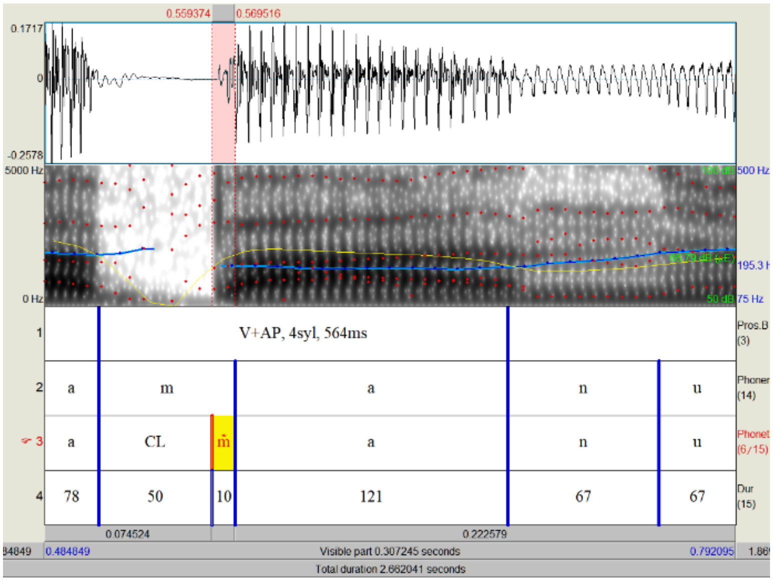


Fig.1 Sample waveform and spectrogram of denasalized [m͊] and sonorous [n]

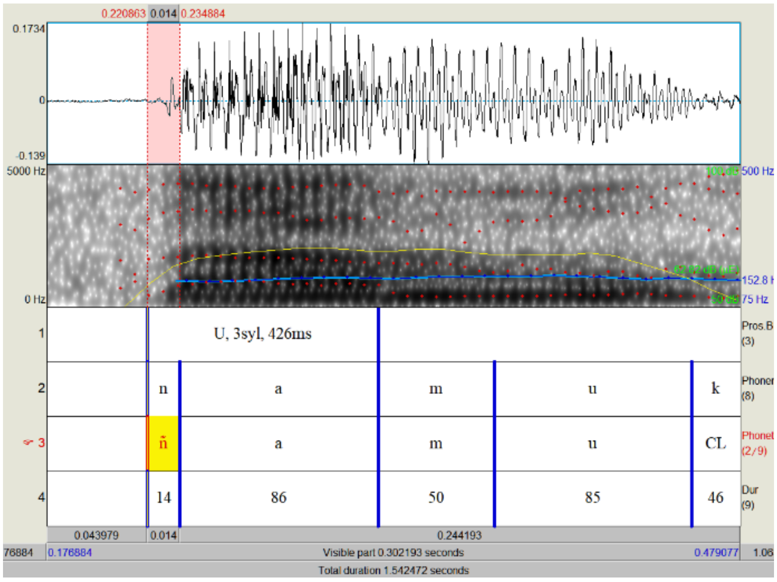


Fig.2 Sample waveform and spectrogram of a U-initial nasal [n͊]

References

[1] Jones, D. (1924). spesimɛn kəriən. lə mɛːtrə fɛɔnetik ɔrgan də l asɔsjɑːsjɔ fɔnetik ɛ ːtɛrnasjɔnal, 14-15. Association Phonétique Internationale.

[2] Martin, S. E. (1951). Korean phonetics. *Language*, *27*(4), 519-533. Washington DC: Linguistic Society of America.

[3] Chen, M. & Clumeck, H. (1975). "Denasalization in Korean: A search for universals," In C. A. Ferguson, L. M. Hyman, and J. J. Ohala (eds.), *Nasalfest: Papers from a symposium on nasals and nasalization, 125-131*. Stanford, CA: Stanford University Linguistics Dept.

[4] Yoo, K. (2015). Domain-initial denasalisation in Busan Korean: a cross-generational case study. In *the Proceedings of the 18th International Congress of Phonetic Sciences*. University of Glasgow, Glasgow.

[5] Cho, T. & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29,* 155-190.

The association between electroglottograph amplitude and pitch contour in Taiwan Mandarin tone production

Chenhao Chiu

National Taiwan University (Taiwan)
chenhaochiu@ntu.edu.tw

Tonal languages employ different pitch profiles, either level or contour tones, to contrast lexical meanings. Distinct pitch profiles are determined by the frequency, stability, and variation of laryngeal vibrations [1,2]. To measure the laryngeal vibration, electroglottography (EGG) is one of the most commonly used methods for its characteristics of noninvasive nature [3,4]. While research attention has mainly focused on the analyses of close and open quotients of EGG pulses [5,6] and to the derivative of EGG [7,8] in accounting for different phonation types or vocal efforts, the issue of how the amplitude of EGG pulses may reflect glottal activity as to pitch performances is understudied. The current study aims to address the latter issue by examining the tonal production of Taiwan Mandarin speakers.

Taiwan Mandarin is one of the tonal languages that includes both level (high-level Tone 1) and contour tones (rising Tone 2, dipping Tone 3, and falling Tone 4). In particular, the dipping of falling-rising contour in Tone 3 is usually realized as a low-falling tone in non-final position, which may be considered to be tone reduction [9]. To examine how glottal activity is associated with F0 contours across different lexical tones (five types: Tone 1, Tone 2, full Tone 3, reduced Tone 3, and Tone 4), four native Taiwan Mandarin speakers were recruited for EGG data collection. Only male participants were recruited based on the fact that physiologically larger larynx would yield better EGG signals with regards to amplitudes [2]. Speakers produced vowels [a, i, u] matched with the five types of the four lexical tones at their own pace (approximately 1 – 1.5 syllables/second) in modal voice, with average intensity ranging from 60 to 70 dB. Ten repetitions were produced for each vowel. To measure EGG amplitude, each EGG pulse was first segmented. For each EGG pulse, the amplitude was defined as the difference between the pulse maximus and minimus. Both F0 contours and EGG pulses were normalized across time and then analyzed using smoothing spline ANOVA [10].

Preliminary EGG results showed that for Tone 1, 2, and full Tone 3, the contours of EGG amplitudes matched the respective pitch contours: Level for Tone 1, rising contour for Tone 2, and dipping for full Tone 3. The results also showed that when the tones ended high (i.e., Tone 1, 2, and full Tone 3), more substantial contacts and opening of the glottis (i.e., larger EGG amplitudes) were observed, suggesting that more efforts were implemented towards the end of the syllable in order to compensate for the aerodynamic decay. For Tone 4, on the other hand, the EGG amplitude undergoes a level, steady state while the pitch is determined to be a falling contour. Similar to Tone 4, vowels matched with a reduced Tone 3 were also associated with low F0 and low EGG amplitude at the end of the production, with reduced Tone 3 induced more creakiness towards the end. It is also noticed that while reduced Tone 3 also showed decreasing EGG amplitude, it occupied a higher and larger amplitude envelope than full Tone 3. Crucially, the result suggests that the ‘reduced’ Tone 3 in Mandarin may not necessarily be a reduced tone given more substantial glottal activity was observed throughout the production. Taken together the asymmetrical behaviour between pitch and EGG amplitude of reduced Tone 3 and Tone 4, the low and low-falling pitch in these two tones ought to be manifested by other glottal constriction or movement (e.g., vertical movement). Finally, the aforementioned patterns accounts for the most numbers of [a] while tokens of the other two vowels are more variable, suggesting that the tongue being low and back in [a] would force the pitch contour to be determined mostly by the dynamic of laryngeal movement. A more general implication is that the production of lexical tones may require more global coordination and such coordination may vary across different contexts depending on the degrees of freedom available in the laryngeal and supralaryngeal areas.

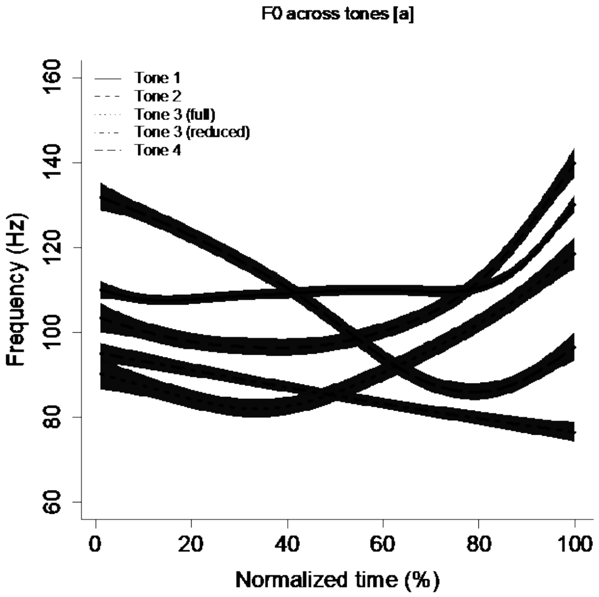


Fig.1 Exemplar F0 contours across four lexical tones for vowel [a]

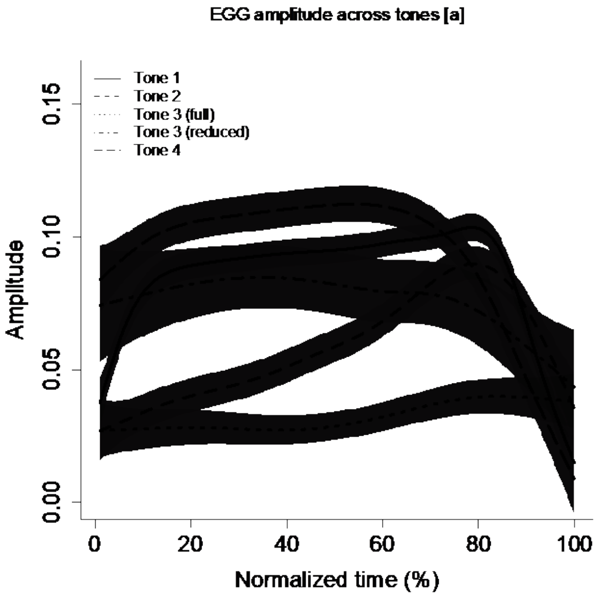


Fig.2 Exemplar EGG amplitudes across four lexical tones for vowel [a]

References

[1] Sawashima, M., Guy, T., and Harris, K. (1969) Laryngeal muscle activity during vocal pitch and intensity changes. *Haskins Laboratories Status Report, SR-19/20*, 211 – 220.

[2] Brunelle, M., Nguyên, D. D., & Nguyen, K. H. (2010). A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica*, *67*(3), 147-169.

[3] Haji, T., Horiguchi, S., Baer, T., & Gould, W. J. (1986). Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation. *The Journal of the Acoustical Society of America*, *80*(1), 58-62.

[4] Rothenberg, M. (1992). A multichannel electroglottograph. *Journal of Voice*, *6*(1), 36-43.

[5] Rothenberg, M., & Mahshie, J. J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech, Language, and Hearing Research*, *31*(3), 338-351.

[6] Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *The Journal of the Acoustical Society of America*, *127*(2), 1047-1058.

[7] Henrich, N., d'Alessandro, C., Doval, B., & Castellengo, M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *The Journal of the Acoustical Society of America*, *115*(3), 1321-1332.

[8] Awan, S. N., Krauss, A. R., & Herbst, C. T. (2015). An examination of the relationship between electroglottographic contact quotient, electroglottographic decontacting phase profile, and acoustical spectral moments. *Journal of Voice*, *29*(5), 519-529.

[9] Zhang, J. and Y. Lai (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27.1: 153-201.

[10] Derrick, D., & Schultz, B. (2013, June). Acoustic correlates of flaps in North American English. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, No. 1, p. 060260). ASA.

Prosodic strengthening effects on morpheme boundaries in Korean: A preliminary study

Jiyoung Lee[1], Sahyang Kim[2], and Taehong Cho[1]

[1]Hanyang University, [2]Hongik University
ljy1004@hanyang.ac.kr, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Studies on the interplay of phonetics and morphology have shown that the low-level phonetic realization may reflect its underlying morphological information [1,2,3]. Regarding the phonetic signature of morphological structure, for example, [napi] ('butterfly') in Korean which consists of a single morpheme shows more stable phonetic realization than [nap-i] ('lead+Nom.') which is composed of two morphemes, although both have the same segmental sequence [1]. This implies a morphological conditioning of phonetic realization, and is consistent with a view of Articulatory Phonology [4] that each lexical item has specified timing relations among articulatory gestures and thus a segmental string of a single lexical item is likely to be realized more stably in terms of the relations of articulatory gestures compared to the same string derived due to morphological concatenation [1]. The phonetic realization, however, is also known to be fine-tuned according to prosodic structure, which is largely considered to have a dual function of prosodic boundary marking and prominence marking (e.g., [5,6,7]). It is therefore reasonable to assume that the two higher-order linguistic structures interact with each other in determining the surface phonetic realization. But relatively little has been explored about the purported interaction. The present articulatory study, therefore, investigates how the interaction between the morphological structure and the higher-order prosodic structure is reflected on the low-level phonetic timing relations among articulatory gestures by using an Electromagnetic Articulography (EMA, AG501, Carstens Electronics). It specifically questions how the apparently homophonic segmental strings of differential morphological compositions (e.g., tautomorphemic /papi/ vs. heteromorphemic /pap-i/) in Korean are realized in terms of intergestural timing between the consonantal and the vocalic gestures that form the second syllable (/pi/ vs. /p-i/). The morphological effect was tested in conjunction with two prosodic factors: Focus (focused vs. unfocused) and Domain-Initial Strengthening (DIS, Intonational Phrase-initially vs. Intonational Phrase-medially). Given that focus is expected to enhance the contrast of underlying phonological structures [8], it is hypothesized that focus enhances the contrast of the underlying morphological structures. Furthermore, given that DIS is likely to heighten the phonetic clarity especially of the first syllable [6], it is hypothesized that the underlying coda /p/ in /pap-i/ is more likely to be subject to an influence of DIS than the underlying onset /p/ of the second syllable in /papi/, presumably because the former would show a more coda-like timing pattern than the latter would.

Eleven native speakers of Seoul Korean in their 20s produced four target words consisting of monomorphemic items ([papi], 'barbie' and [api] 'father,' informal), and bi morphemic items ([pap-i] 'meal + Nom. and [ap-i] 'pressure'+Nom.). Each target word was embedded in carrier sentences (1) where the focus was constructed by means of the morphologically contrastive context (e.g., [papi] vs. [pap-i]) and (2) where the target item occurred after the IP boundary or in the IP medial position (Table 1). Each speaker produced 240 sentences (4 target words * 2 focus * 2 boundaries * 15 repetitions). The articulatory measures included the time interval between the onset of the /p/ closing gesture (as reflected in Lip Aperture) and the onset of the /i/ gesture in the second syllable; and the %-sequence overlap between the two gestures.

We are currently in the process of analyzing the results. Preliminary results obtained from part of the collected data are shown in Figures 1 and 2. As shown in Figure 1, there appeared to be a clear distinction between tautomorphemic and heteromorphemic sequences in the time interval between the onsets of C and V gestures, so that the interval was shorter in the tautomorphemic sequence (e.g., [papi]) than in the heteromorphemic one (e.g., [pap-i]). This implied that the two gestures were more closely timed when they belonged to the same lexical item than when they belonged to different items. There was also a boundary (DIS) effect, such that the two gestures were timed more closely in the IP-initial than in the IP-medial (Wd-initial) condition. The results further indicated a possible interaction between Morphological Structure and Boundary, so that the difference in timing due to the morphological composition was more robust in the IP-initial than in the IP-medial condition. No evidence was observed either for the effect of Focus or for the interaction between Morphological Structure and Focus.

The %-Sequence Overlap measure, as shown in Figure 2, indicates that there was a tendency towards *more* overlap between the two gestures in the tautomorphemic sequence than in the heteromorphemic one. Interestingly, however, Focus and Boundary showed opposite trends. Focus induced reduction in %-Sequence Overlap while %-Sequence Overlap increased due to Boundary. It appeared that the two gestures became more distinct (less overlapped) under focus. The results also showed a possible Focus x Boundary interaction, which suggested that the morphological distinction was indeed enhanced under focus at least in terms of the degree of intergestural overlap.

Taken together, these qualitatively observed results, though very preliminary, implied that the identical segmental strings are indeed produced differently as a function of morphological structure, and its distinction is further heightened by prosodic strengthening effects, especially when the morphological contrast received focus. More data will be analyzed to test the above-mentioned hypotheses in a statistical meaningful way.

Table 1. An example set of carrier sentences with a target word [pap-i]. The target word is underlined, and it is in bold when it receives focus (in contrast with its tautomorphemic counterpart [papi]). '#' refers to a prosodic boundary which is either IP orr Wd.

Conditions		**Target-bearing sentences**
#=IP	Morphological Focus (MF)	[tʃikɨmjʌki **papi**rago s*ʌnni, # **pap-i**rago s*ʌnni] Right here, did you write down Barbie or meal? 지금 여기, **바비**라고 썼니, # **밥이**라고 썼니?
	No Focus (NoF)	[tʃikɨmjʌki papirago **ni**ka s*ʌnni, # pap-irago **tʃje**ka s*ʌnni] Right here, is it you or that person who wrote a meal? 지금 여기, 밥이라고 **니**가 썼니, # 밥이라고 **재**가 썼니?
#=Wd	Morphological Focus (MF)	[tʃikɨmjʌki wuri**papi**rago s*ʌnni, wuri#**pap-i**rago s*ʌnni] Right here, did you write down our Barbie or our meal? 지금 여기, 우리**바비**라고 썼니, 우리**밥이**라고 썼니?
	No Focus (NoF)	[tʃikɨmjʌki wuripapirago **ni**ka s*ʌnni, wuri#pap-irago **tʃje**ka s*ʌnni] Right here, is it you or that person who wrote our meal? 지금 여기, 우리밥이라고 **니**가 썼니, 우리밥이라고 **재**가 썼니?

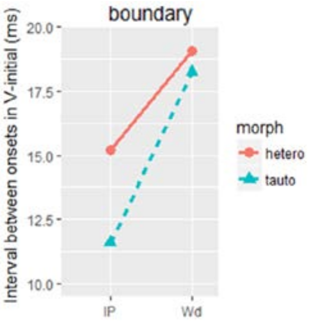


Figure 1. Interval between the onsets of the CV gestures in the second syllable.

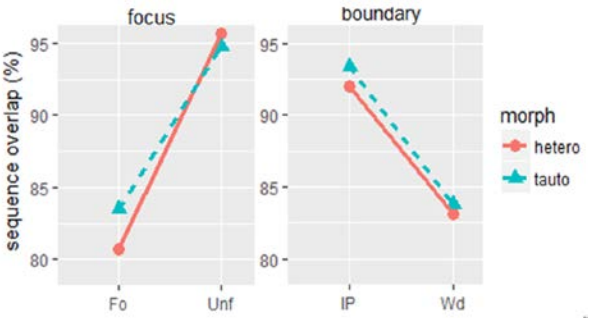


Figure 2. %-Sequence Overlap between C and V gestures in the second syllable.

References

[1] Cho, T. (2001). Effects of morpheme boundaries on intergestural timing: Evidence from Korean. Phonetica, 58(3), 129-162.

[2] Mousikou, P., Strycharczuk, P., Turk, A., Rastle, K., & Scobbie, J. M. (2015). Morphological effects on pronunciation. Proceedings of the 18th ICPhS, Glasgow (0816).

[3] Hedia S.B. & Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics*, 62, 34-49.

[4] Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*,49(3-4), 155-180.

[5] Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29(2), 155-190.

[6] Cho, T. (2006). Manifestation of Prosodic Structure in Articulation: Evidence from Lip Kinematics in English. *Laboratory Phonology 8*. (Berlin/New York: Mouton de Gruyter), 519-548.

[7] Cho, T., Kim, D. & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics*, 64,71-89.

[8] De Jong, K. J. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32, 493-516.

A preliminary study on Japanese phrase final lengthening in relation to prosodic structure: mora, lexical pitch accent and focus

Jungyun Seo[1], Sahyang Kim[2], Haruo Kubozono[3] & and Taehong Cho[1]

[1]Hanyang University (Korea), [2]Hongik University (Korea), [3]National Institute for Japanese Language and Linguistics (Japan)
jungyunlseo@gmail.com, sahyang@hongik.ac.kr, kubozono@ninjal.ac.jp, tcho@hanyang.ac.kr

Words are produced with longer duration in a phrase-final position than in a phrase-medial position in many languages. This phenomenon is called ‘phrase final lengthening’ or ‘preboundary lengthening’ (henceforth PBL) (e.g., [1, 2, 3, 4]), which is be in part due to the fact that movement of articulators tends to slow down over the course of time as the articulatory force is likely to weaken towards the end of a phrase (e.g.,[1]). This non-contrastive low-level phenomenon, however, can be modulated by the linguistic structure of a language. Different languages have different prosodic structures which may influence the fine-grained realization of PBL. In stress-timed languages like English and Greek, the scope of PBL may be extended to the stressed syllable even when it is not a final syllable (e.g., [4, 5, 6]). But this pattern is not expected to occur in Korean which is known as a syllable-timed language [7]. Based on the premise that the scope of PBL can be determined by language-specific structural factors, the goal of the present study is to build on this assumption by examining the distribution of PBL in Japanese, a mora-timed language. We focus on two linguistic structural aspects: the moraic structure and the lexically-defined pitch accent in conjunction of focus as a prominence marker.

The mora plays an important role in the speech rhythm in Japanese. Its structural importance has been demonstrated in various aspects: Japanese listeners segment what they hear into mora-sized chunks [8], and phonological evidence for moras has been found in blend errors and word and nick name formations (e.g., [9, 10]). Considering the importance of mora as an access code in speech perception [8], the question arises as to how the speech production could be modulated based on this unit. Therefore, whether the scope of PBL may be determined by the distribution of moras over the final word will be asked as one of the important research questions.

Furthermore, from the viewpoint that the prominence system in which pitch accent and focus are expected to interact is one dimension of the prosodic structure apart from the other boundary related aspect, this study will examine how the distribution of PBL is modulated by the prominence structure. As mentioned above, previous studies in English and Greek showed interactions between boundary and prominence—i.e., Stress in a non-final syllable may attract PBL (e.g., [3, 5, 11]). Pitch accent in Japanese is similar to lexical stress in English in that both may serve as a lexically specified prominence marker. But pitch accent in Japanese differs from stress, in that it is expressed primarily by F0 rather than amplitude or duration [12], which influences a higher-order intonational structure. Therefore, another question to be addressed is how (dis)similar the effect of pitch accent on PBL will be in comparison to that of stress observed in English and Greek. Through a manipulation of different prominence levels (lexically specified pitch accent and focus driven by the information structure), the PBL effect will be examined in connection with the realization of prominence.

In order to answer these questions, the disyllabic target words (CVCV (2μ), CVNCV (3μ), CVCVN (3μ), CVNCVN (4μ)) were chosen to vary systematically in the mora structure as shown in Table 1. The target words were further divided into two sets depending on the presence and the absence of pitch accent on the final syllable. The target words were inserted in carrier sentences so that they occurred in various prosodic conditions: Boundary (Phrase-final vs. Phrase-medial) and Focus (focused vs. unfocused). As can be seen in Table 2, target words (underlined) always occurred in the second sentence (‘B’) preceded by a question sentence (‘A’) which was used to induce an intended focus type for the target word. In total, twenty native speakers of Tokyo Japanese will take part in the experiment. We will measure the acoustic duration of each moraic unit (CV and N) of the target words to explore the scope of PBL. Preliminary results based on a subpart of the speakers will be presented at the conference.

Table 1. Bi-syllabic target words with different moraic structures in two pitch accent patterns (unaccented vs. initially-attenced) A set of nonsense words was included in order to make initially accented set segmentally comparable to the unaccented real words set.

Segmental composition		**Unaccented**		**Initially accented**	
Same consonants and vowels within a set	CVCV	TAKA	鷹	TA’KA	タカ
	CVNCV	TANKA	炭化	TA’NKA	タンカ
	CVCVN	TAKAN	多感	TA’KAN	タカン
	CVNCVN	TANKAN	短観	TA’NKAN	タンカン
Same consonants but different vowels	CVCV	SAKE	酒	SA’KO	迫
	CVNCV	SANKE	産気	SA’NKO	三個
	CVCVN	SAKEN	差遣 (さけん)	SA’KON	左近
	CVNCVN	SANKEN	三権，散見	SA’NKON	サンコン

Table 2. Carrier sentences. Target words are underlined and a contrastive focus falls either on the target word or on the word that precedes it.

Boundary	**Focus**		Example sentences with a target word “TAKA” (“taka”)
Phrase – final	**a.**	Focused (**contrastive** focus between SAKE vs. TAKA)	A. 今度もその**酒**、試しに使ってみる？(Kondomo sono***SAKE***, tamesinitsukattemiru?) B. いいえ、今度はその**鷹**、試しに使ってみる(Iie, kondowa sono***TAKA***, tamesinitsukattemiru) A. Do you try and use that **SAKE** this time again? B. No, this time I try and use that **TAKA**.
	b.	Unfocused (**contrastive** focus elsewhere KONO vs. SONO)	A. 今度も**この**酒、試しに使ってみる？(Kondomo **KONO**taka, tamesinitsukattemiru?) B. いいえ、今度は**その**鷹、試しに使ってみ (Iie, kondowa **SONO**taka, tamesinitsukattemiru) A. Do you try and use **THIS** sake this time again? B. No, this time I try and use **THAT** taka.
Phrase – medial	**c.**	Focused (**contrastive** focus between SAKE vs. TAKA)	A.これは、その**酒**と一緒に置きますか？(Korewa, sono**SAKE**to ittshoni okimasuka?) B. いいえ、今度はその**鷹**と一緒に置きます。(Iie, kondowa sono**TAKA**to ittshoni okimasu) A. Do you put this with that **SAKE**? B. No, this time I put it with that **TAKA**
	d.	Unfocused (**contrastive** focus elsewhere KONO vs. SONO)	A. これは、**この**鷹と一緒に置きますか？(Korewa, **KONO**takato ittshoni okimasuka) B. いいえ、今度は**その**鷹と一緒に置きます。(Iie, kondowa **SONO**takato ittshoni okimasu.) A. Do you put this with **THIS** taka? B. No, this time I put it with **THAT** taka.

References

[1] Cho, T. (2015). Language effects on timing at the segmental and suprasegmental levels. In M. A. Redford, (Ed.), The Handbook of Speech Production (pp. 505-529). Hoboken, NJ: Wiley-Blackwell.

[2] Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *the Journal of the Acoustical Society of America*, 89(1), 369-382.

[3] Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35(4), 445-472.

[4] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91(3), 1707-1717.

[5] Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of phonetics*, 55, 149-181.

[6] Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5), 1208-1221.

[7] Cho, T. (2017). Articulatory studies on preboundary lengthening in American English and Korean. A talk presented at the 3rd International Workshop on Dynamic Modeling. Cologne, Germany, 18-19, July 2017.

[8] Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of memory and language*, 32(2), 258-278.

[9] Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech*, 32(3), 249-278.

[10] Kubozono, H. (Ed.). (2015). *Handbook of Japanese phonetics and phonology* (Vol. 2). Walter de Gruyter GmbH & Co KG.

[11] Kim, S., Jang, J., & Cho, T. (2017). Articulatory characteristics of preboundary lengthening in interaction with prominence on tri-syllabic words in American English. *The Journal of the Acoustical Society of America*, 142(4), EL362-EL368.

[12] Beckman, M. E. (1986). Stress and non-stress accent (Vol. 7). Walter de Gruyter.

Acoustic Study of Vowels in Mizo

Wendy Lalhminghlui and Priyankoo Sarmah

Indian Institute of Technology Guwahati (India)
wendy@iitg.ac.in, priyankoo@iitg.ac.in

The present study is an acoustic analysis of vowels in Mizo. Mizo is an under-resourced language belonging to the Central Kukish Branch of the Tibeto- Burman language family spoken mainly in the state of Mizoram and its neighbouring states of North- East India, Chin State in Myanmar and in the Chittagong Hill tracts of Bangladesh. There are 0.7 million speakers of Mizo[1]. Previous studies reported that there are 5 vowels in Mizo namely, /a, i, u, ɛ, ɔ/ [2-5]. Each vowel has their corresponding long vowel [3, 4]. Mizo is a tonal language with four distinctive lexical tones namely, high, low, falling and rising tone [3]. Every syllable in Mizo has vowel nucleus and all the vowels can bear all the four tones of Mizo [4]. Since there is interaction between tones and vowel length in Mizo, the tone information can give a hint to the length of vowels in Mizo. For example, syllables with low tone and a coda mostly have short vowels. Therefore, this study aims to look into the duration of long and short vowels in Mizo. The present study is based on the recorded speech of 67 Mizo native speakers (34 female, 33 male) from Mizoram. There are 27 unique words (14 long vowels, 13 short vowels) with VC, CVC, CV: C syllables, having different tones. The vowel length is considered based on the existing literature. Each vowel is repeated three times by all the speakers in citation form resulting in a total of 5173 vowel tokens (2496 short, 2677 long). Since the formant frequencies help in identifying vowels, the present study considered the first two formants of vowels (F1 and F2) which are correlated to vowel height and frontness respectively. The acoustic analysis is done using Praat 5.3 [6]. Formant values are extracted at vowel midpoint for steady state formants. The F1 and F2 values are then plotted in NORM [7]. The temporal quality of vowels is also taken into consideration. The perceptual distance between all the five vowels are also calculated using Euclidean distance.

From these analysis, the result in the present study shows that there are five distinct vowels in Mizo. The plot of normalized values of the formants within one standard deviation of all the vowels of all the speakers in Figure 1 clearly depicts the existence of five vowels /a, ɛ, i, ɔ, u/ in Mizo for both long and short vowels.

Also, a one –way ANOVA test followed by Post-hoc Bonferroni test shows that all the vowels are statistically different from each other except in case of F2 of /ɔ/ and /u/ in short and long vowels. Hence, all the vowels are distinct in terms of their height and frontness quality except for the frontness quality of /ɔ/ and /u/ which can be interpreted as due to their backness. The comparison of F1 and F2 across the short and long vowels also shows that only the F1 of /i/ and F2 of /u/ are not significantly different.

The formants of male and female in short and long vowels plots are shown in Figure 2 and 3 respectively. As far as female and male difference in Mizo vowels is concerned, both short and long vowels show the same results where there is no significant difference of /ɔ/ in terms of F1 and F2 and /u/ in terms of F2.

This means that both female and male produce /ɔ/ with the same height and frontness, and /u/ with the same frontness. In terms of duration, statistical analysis has shown that there are no significant difference between the vowel durations of /a/ and /ɛ/, /a/ and /u/, /i/ and /ɔ/; and /u/ and /ɛ/ within short vowels. In long vowel, there is no significant difference in vowel durations between /a/ and /ɔ/, /ɛ/ and /i/ and /i/ and /u/. This suggests that there is no correlation between vowel quality and duration as far as Mizo is concerned. However, the duration of all the vowels is significantly distinct from each other across short and long vowels. The Euclidean distance shows that each vowel differs from one another perceptually. Both short and long vowels have shown the same results where /i/ and /u/ has the longest distance (346.08mels in short, 377.64mels in long) which is quite obvious since the two vowels are part of the peripheral vowels. /u/ and /ɔ/ has the shortest distance (72.71mels in short, 59.67mels in long) since both are back vowels. The same result is observed within male and female group. The study between male and female vowel is novel since the previous study did not consider gender as a factor of vowel variation in Mizo.

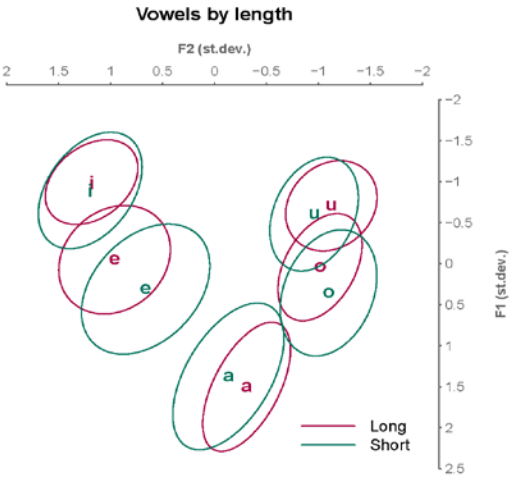
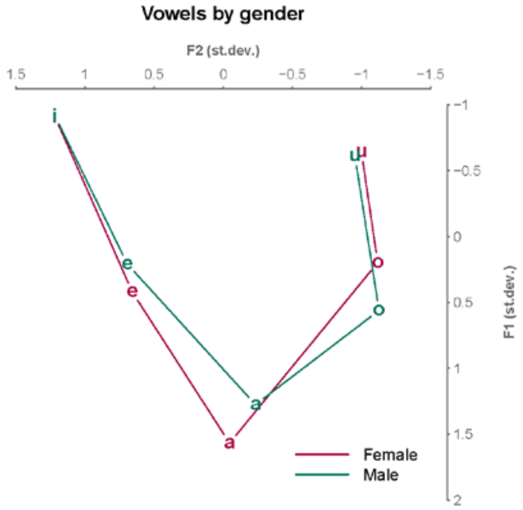


Fig. 1 Mizo vowel plot with normalized averaged F1 and F2 obtained from 67 speakers showing short and long distinction



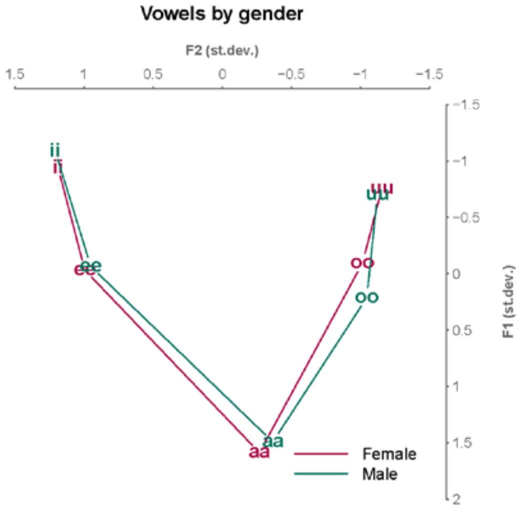


Fig.2 Mizo short vowels plot showing female and male distinction

Fig.3 Mizo long vowels plot showing female and male distinction

References

[1] http://www.censusindia.gov.in/

[2] Weidert, A. (1975). Componential Analysis of Lushai Phonology. John Benjamins Publishing Company, Amsterdam.

[3] Fanai, L. (1992). Some Aspects of the Lexical Phonology of Mizo and English: An Autosegmental Approach, PhD dissertation, CIEFL, Hyderabad, India.

[4] Chhangte, L. (1993). A Preliminary Grammar of the Mizo Language. Master's thesis, University of Texas, Arlington.

[5] Sarmah, P. & Wiltshire, C. (2010). A Preliminary Acoustic Study of Mizo Vowels and Tones. J. Acoust. Soc. Ind, 37, 121–129.

[6] Boersma, P. & Weenink, D. (2017). Praat, doing phonetics by computer [Computer program]. Version 5:9/10.

[7] Thomas, E.R. & Kendall, T. (2007). NORM : The vowel normalization and plotting suite. [Online resource]

Tactile cues might explain the verb-bias in Korean child-directed speech

Eon-Suk Ko[1], Kyungwoon On[2], Jinyoung Jo[2], Eunsol Kim[2],
Rana Abu-Zhaya[3], Amanda Seidl[3]

[1]Chosun University, [2]Seoul National University, [3]Purdue University
eonsuk@gmail.com

Korean learners' early vocabularies contain a greater ratio of verbs to nouns than English learners [1]. This difference might be due to Korean mothers' behavior either because of the structure of Korean (e.g., allowing for frequent elision of subjects and objects) or because Korean caregivers' interactions differ from English-speaking caregivers. Based on the recent proposal that touch might help infants' word learning [2], we test the hypothesis that Korean mothers might emphasize verbs rather than nouns with their touch.

We examined a multimodal corpus of 35 Korean mother-child dyads divided in three age groups: preverbal (6- to 9-month-olds, 12 dyads), early speech (12- to 16-month-olds, 11 dyads), and multi-word (25- to 28-month-olds, 12 dyads) stage. Each dyad played freely for about 50 minutes in a mock-apartment composed of a living room, a bedroom, and an eat-in kitchen and was recorded by 4 wall-mounted cameras and clip-on microphones.

To replicate claims concerning a verb- vs. noun-bias, we first analyzed the distribution of nouns and verbs in one-word utterances in age-matched data from 35 Korean mothers and 43 American mothers (Brent & Bates corpus). Since one-word utterances facilitate word learning [3], we hypothesized that Korean mothers would produce a higher proportion of verbs. A Chi-square test with Yates' continuity correction revealed that the proportion of nouns significantly differed by language ($\chi^2(1, N=7640)=1716.22$, $p<0.001$), confirming a verb-bias for Korean and a noun-bias for English (Choi, 2000).

To investigate mother's use of tactile cues as a function of part-of-speech, we selected 13 videos containing a 6-minute segment throughout which the dyads stayed on a mat. Audio was transcribed, and a forced-alignment tool kit was used to mark the boundaries between words, which were then manually fine-tuned and annotated for part-of-speech. The boundaries of touches were coded in ELAN using visual information as in [2].

We first examined touch-word alignment (Mean number of touches=16.4, SD=21.7) by comparing the actual data with random alignments generated by shuffling words. Alignment was defined as sequences in which touch+word boundaries co-occurred within a 250 ms interval. A paired Wilcoxon signed rank test showed that touch tends to align with words at a greater rate than by chance ($V=0$, $p<0.01$). We then examined the proportion of verbs and nouns aligning with touch. Our Chi-square test with Yates' continuity correction showed that the proportion of verbs co-occurring with touch was significantly higher than nouns ($\chi^2(1, N=2710)=6.68$, $p<0.01$).

Our results indicate that Korean mothers focus on teaching verbs over nouns by presenting them more often as isolated words as well as with well-aligned multimodal cues. The frequency of verbs and the co-occurrence of verbs with touch may contribute to explain the early acquisition of verbs in Korean learners. Further, the higher proportion of alignment in verb+touch units, over noun+touch units suggest that the verb-bias in Korean is not simply an artifact of grammatical properties, but reflects genuine differences between Korean and American mothers' word-teaching strategies.

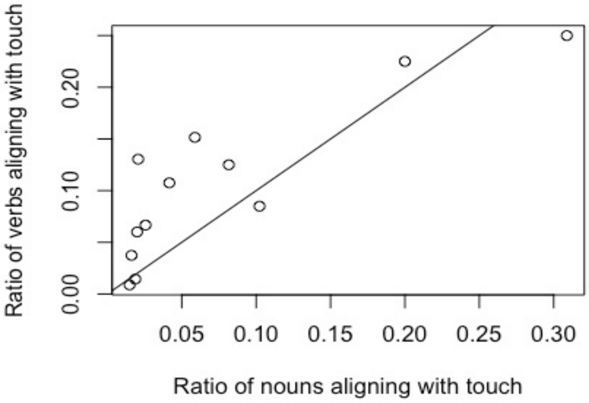


Fig.1 Co-occurrence ratio of nouns and verbs. Each data point represents the ratio of word+touch co-occurrence in a mother-child dyad

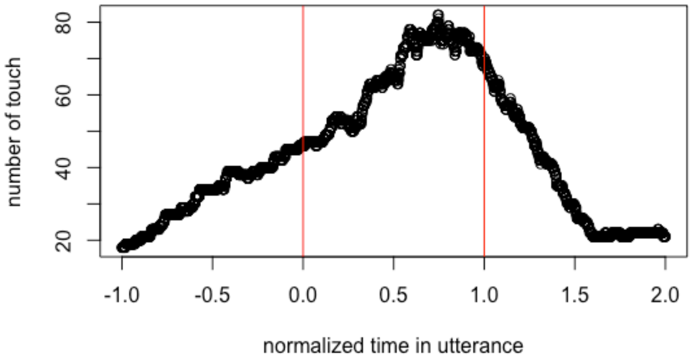


Fig.2 Number of touches over normalized time in utterance. Two red vertical lines represent the beginning and the end of utterances

References

[1] Choi, S. & Alison Gopnik (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language*, 22, pp 497-529 doi:10.1017/ S0305000900009934

[2] Abu Zhaya, R., Seidl, A., & Cristia, A. (2016). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues, *Journal of Child Language*, *44*, 1088-1116.

[3] Brent, M. R. & Siskind, J. M (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, 31-44.

Mandarin tone training: considering lexical access

Bei Yang

University of Wisconsin – Madison
byang25@wisc.edu

Modern standard Chinese is a tone language in which tones can distinguish lexical meanings. There are two difficulties for second language tone acquisition. The first is that English-speaking learners' L1 is a non-tone language. The second difficulty lies in understanding the importance of lexical meaning during the production process. Lexical tones are processed physically and phonologically, together with the lexical meaning. However, L2 tones are learned as separate from lexical learning. There are two major training methods used to improve non-native speakers' tone production. One is perception training [11] and the other is audio-visual training [4, 9, 10]. However, both methods focused on the linguistic forms of tones, yet ignored lexical meaning.

In order to let learners experience tone contours, the current study employed audio-visual training to present tone contours visually. Meanwhile, in order to connect form to meaning, focus on form design was added, so that learners could access lexical meaning and encode phonological form to develop tone production process. The tone development is discussed after training through Levelt's production model.

According to the Levelt's model [3, 5, 6], there are several processing stages involved in lexical access in speech production: conceptual preparation, lexical selection, morphological encoding, phonological encoding, phonetic encoding, articulating and self-monitoring. The previous studies [8] have shown that major units processed at some levels of phonological encoding are different in English and Chinese. During phonological decoding, the tonally unspecified syllable is the explicit unit "at the first phonological level" [8:282] in Mandarin, whereas the segment is prepared at the same level in English; but the syllable is the major unit in phonological encoding in both Chinese and English. However, in Mandarin, tone is activated and processed mainly at a later stage of phonological encoding, which is similar to the activation of stress in English [3, 6, 8].

In the current study, 10 American learners of Chinese received a Focus on Form training, 9 received a Focus on FormS training, and 9 received no training. All participants took the assessment test five times: a pretest, a post-test, and three retention tests. Two audio-visual training paradigms were designed according to FonF and FonFS, respectively, for five days. Focus on Form training employs the picture-naming task which is a conceptually driven task. According to Levelt's model, a picture can activate lexical concept, lexical selection, and then encode phonological form and phonetic form. Focus on Forms training employs the reading-aloud task, in which phonological encoding is directly activated by *Pinyin* (a transcription of pronunciations in Mandarin Chinese which is not a written system). The previous studies about picture naming and reading aloud (e.g. [7]) showed that semantic activation of the target stimulus is required in picture naming; in contrast, reading aloud of a printed word, in principle, is on the basis of sublexical information (non-semantic). As for a language with transparent orthography, such as Italian [2], has found that semantic-conceptual effects for picture naming, and word-form effect (orthographic-phonological) for reading aloud. *Pinyin* is much more transparent than Italian. Therefore, the *Pinyin* reading-aloud process is from phonological encoding to articulation.

Instead of mono-syllabic words, this study examines lexical tones carried by disyllabic words. There are 20 tone combination patterns in total, according to the four lexical tones and the neutral tone that is carried by the second syllable of a disyllabic word. In total 120 disyllabic words were selected from the first-year Chinese textbook based on the 20 tone combination patterns. In the tone training programs, 80 words were used. In the assessments, 20 trained words and 40 untrained words were used. The participants had learned all the words before they participated in this study. Besides being categorized according to the tone combination patterns, the words were also categorized based on meaning. There were 17 meaning-based categories, including Chinese foods, clothing, and so on.

Five native speakers of Chinese were recruited to evaluate the five assessment tests of 28 participants. Qualtrics was employed to conduct the evaluation. The evaluation scores were analyzed in the framework of generalized linear mixed effects models [1] with fixed effects of group effects (FonF, FonFS, Control), training effect (trained vs. untrained), tone combination pattern effects and their interactions, and rater and subject random effects using SAS version 9.3 (SAS Institute, Cary, NC) PROC MIXED. For pairwise comparisons, the Tukey method was used to report the adjusted P-values and confidence intervals. Cronbach's alpha for 5 raters was .9505. The Shrout-Fleiss interclass correlation (ICC) for inter-rater reliability was .7772, indicating the raters were consistent in how they rated the speakers.

The results show that both training methods are significantly effective. However, Focus on Form training is significantly more efficient than the other. Trained words could be improved after two trainings, yet the improvement extended to untrained words only in the FonF group. Compatible tone patterns, i.e. in these patterns, the endpoint of the first tone contour is compatible to the starting point of the second, were improved significantly only after FonF training, while the incompatible tone patterns were improved significantly in both training groups yet the retention was better after FonFS training.

These results confirm that audio-visual training is effective to develop L2 tone contours, and reveal how FonF training helps learners process tones more native-like at the late stage of phonological encoding only after their tone production reaches a certain proficiency level. Since the first level of phonological encoding is relevant to the tonally-unspecified syllable, which indicates that the tonally-unspecified syllable directly connects to meaning; after that, the tone is processed based on the syllable. Stress in English that is processed at the same level of tones, is not required to distinguish lexical meanings generally. The way that native speakers of English process stress influences learners' tone process. English speaking learners of Chinese do not connect tonal process to lexical meaning directly, yet the tonally-unspecified syllable is. This explains why learners' tone production becomes much weaker when they concentrate on meaning.

References

[1] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.

[2] Bates, E., Burani, C., D'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. *Memory & Cognitio*n, 29(7), 986-999.

[3] Cholin, J., Schiller, N. O., & Levelt, W. J. (2004). The preparation of syllables in speech production. Journal of Memory and Language, 50(1), 47-61.

[4] Chun, D., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *Journal of Second Language Pronunciation*, 1, 86-114.

[5] Levelt, W. J. (1989). *Speaking: From intention to articulation (Vol. 1).* MIT press.

[6] Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(01), 1-75.

[7] Mousikou, P., & Rastle, K. (2015). Lexical frequency effects on articulation: a comparison of picture naming and reading aloud. *Frontiers in psychology*, 6, 1571.

[8] O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, 115(2), 282-302.

[9] Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *The Journal of the Acoustical Society of America*, 113, 1033–1043.

[10] Wang, X. (2008). Training for learning Mandarin tones. In F. Zhang & B. Barber (Eds.), *Handbook of research on computer-enhanced language acquisition and learning* (pp. 259–274). IGI Global.

[11] Wang, X. (2012). Auditory and visual training on Mandarin tones: A pilot study on phrases and sentences. *International Journal of Computer-Assisted Language Learning and Teaching*, 2, 16-29.

Prosodically-conditioned phonetic cue use in production of Korean aspirated vs. lenis stops

Jiyoun Choi[1], Jiyoung Lee[1], Sahyang Kim[2] & Taehong Cho[1]

[1]Hanyang University (Korea), [2]Hongik University (Korea)
jiychoi@hanyang.ac.kr, ljy1004@hanyang.ac.kr, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Korean aspirated and lenis stops used to be distinguished primarily by VOT [1]. But recent studies have shown that the two stops are currently merging in VOT and now distinguished primarily by F0, with high F0 for the aspirated and low F0 for the lenis, especially in young speakers' speech [2-4]. Based on this, it has been argued that the F0 cue becomes a lexical tonal contrast [2-4]. Evidence supporting this claim, however, comes only from productions at prosodic phrase-initial position, where the VOT cue for the contrast comes to be virtually redundant as the intonational phonology distinguishes between the aspirated and lenis stops by categorically assigning tones (high for the aspirated and low for the lenis). Thus, the currently reported change may indicate speakers' tendency to take into account the redundancy that arises at phrase-initial positions, presumably driven by the principle of effort minimization, thus relying primarily on the structurally reinforced F0 cue. This opens up the possibility that the observed phenomenon is accounted for by a position-specific differential use of existing cues, VOT vs. F0. Therefore, to better understand the nature of the purported tonogenetic sound change phenomenon, it is necessary to examine to what extent the primacy of F0 prevails in phrase-initial vs. medial positions where the intonational phonology does not provide contrastive tones to the aspirated vs. lenis stops.

To this end, we examined aspirated and lenis stops in both phrase-initial and medial positions. Note that in phrase-medial positions, lenis stops could be produced as voiced due to the Lenis Stop Voicing rule, implying that VOT could become a decisive cue (i.e., voiced for the lenis vs. voiceless for the aspirated) in phrase-medial positions. Nonetheless, if high/low F0 is a lexical contrast, a substantial F0 distance for the aspirated-lenis contrast should still be observed in the phrase-medial positions. We further examine this issue by exploring how the phonetic cues, especially the F0 cue, would be modulated in hyperarticulated speech. It has been well documented that phonological features are enhanced when they are focused (thus hyperaticulated) [5]; for instance, a Chinese lexical high tone becomes higher whereas a low tone becomes lower under focus [6]. In this regard, a crucial question is whether Korean speakers would show such bidirectional enhancement patterns with raising F0 for the aspirated vs. lowering it for the lenis. We systematically compare young speakers' speech to that for older speakers to see whether the young speakers, who are assumed to lead the sound change [2-4,7], would show a larger F0 difference for the contrasts and/or would show different focus-induced enhancement patterns, as compared to older speakers.

We recorded 32 Seoul Korean speakers: 16 young (aged 18-24) and 16 older speakers (aged 55-62). They produced eight minimal pairs of disyllabic Korean words differing in word-initial aspirated vs. lenis stops, e.g., [thansik] 'sigh' vs. [tansik] 'fast'. The words were placed in carrier sentences where boundary (phrase initial vs. medial) and focus (focused vs. unfocused) were manipulated (Table 1). Speakers were presented with pre-recorded prime questions and then asked to answer with test sentences. VOT and F0 at the midpoint of the following vowels were measured. Linear mixed-effects models were used for analyses.

Results for VOT showed that in phrase-initial position, older speakers showed significant VOT differences between the stops for both focus conditions, whereas young speakers did not show any (Fig.1a). In phrase-medial position, VOT results substantially differed as a function of focus type (Fig1b). When focused, results were similar to those for the phrase-initial positon, such that only older speakers showed significant VOT differences: this similarity can be explained by Korean focus realizations where a focused item is produced phrase-initially. That is, our target stops that were originally set up for being phrase-medial were actually produced phrase-initially when receiving focus. When it was unfocused, phrase-medial VOT was significantly longer for the aspirated than lenis and this was not further modulated by speakers' generation. These results thus reveal that the currently observed VOT merger is advanced for young speakers and it seems to be limited to their phrase-initial speech (including phrase-medial, focused one). Results for F0 showed that there were no significant differences between young and older groups in all conditions, showing comparable F0 weighting for the young and older speakers across different prosodic contexts (Fig1.c,d). Moreover, the young and older speakers raised F0 under focus for both aspirated and lenis stops across phrase positions, which does not seem to support the view that high/low F0 is a phonological feature.

It can therefore be assumed that the currently observed phenomenon may not be the sound change involving the emergence of a tonal contrast. Rather, it may be better understood as a prosodically-conditioned differential use of phonetic cues (VOT vs. F0), which relates to the economical way of using the cues by speakers (especially young speakers), i.e., dropping the VOT cue in phrase-initial positions, which is in fact redundant in the prosodic positions.

Table 1 Examples for a target /thasan/ (#: phrase boundary; in bold: focus)

IP-initial	Focus	Q	ipʌne p*opɨn tanʌnɨn, [$_{IP}$ **tasan** ape nonayo]?
			*“Shall I place this card in front of **tasan** (fecundity)?”*
		A	aniyo. ipʌnen, [$_{IP}$ **t^{h}asan** ape noayo].
			*“No, put it in front of **t^{h}asan** (calculation) this time.”*
	No focus	Q	ipʌne p*opɨn tanʌnɨn, [$_{IP}$ t^{h}asan **ape** nonayo]?
			*“Shall I place this card **in front of** t^{h}asan (calculation)?”*
		A	aniyo. ipʌnen, [$_{IP}$ t^{h}asan **tye** noayo].
			*“No, put it **in back of** t^{h}asan (calculation) this time.”*
IP-medial	Focus	Q	ipʌne p*opɨn tanʌnɨn, [$_{IP}$ nolan **tasan** ape nonayo]?
			*“Shall I place this card in front of yellow **tasan** (fecundity)?”*
		A	aniyo. ipʌnen, [$_{IP}$ nolan **t^{h}asan** ape noayo].
			*“No, put it in front of yellow **t^{h}asan** (calculation) this time.”*
	No focus	Q	ipʌne p*opɨn tanʌnɨn, [$_{IP}$ nolan t^{h}asan **ape** nonayo]?
			*“Shall I place this card **in front of** yellow t^{h}asan (calculation)?”*
		A	aniyo. ipʌnen, [$_{IP}$ nolan t^{h}asan **tye** noayo].
			*“No, put it **in back of** yellow t^{h}asan (calculation) this time.”*

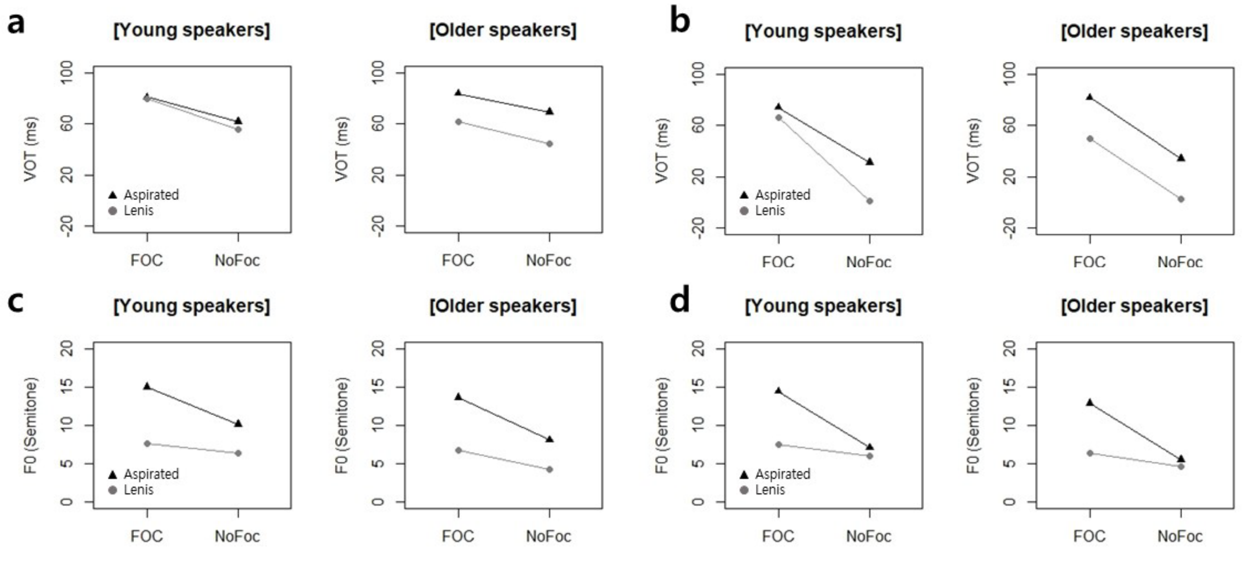


Fig.1 Aspirated vs. lenis contrast as a function of speaker group and focus type in terms of VOT at (a) phrase-initial and (b) medial position, and in terms of F0 at (c) phrase-initial and (d) medial position

References

[1] Cho, T., Jun, S.-A, & Ladefoged, P. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics,* 30, 193-228.

[2] Kang, K.-H, & Guion, S.G. 2008. Clear speech production of Korean stops : changing phonetic targets and enhancement strategies. *Journal of Acoustical Society of America*, 124, 3909-3917.

[3] Kang, Y. 2014. Voice Onset Time merger and development of tonal contrast in Seoul Korean stops : a corpus study. *Journal of Phonetics*, 45, 77-90.

[4] Silva, D.J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporarly Korean. *Phonolgy*, 23, 287-308.

[5] de Jong, K.J. 2004. Stress, lexical focus, and segmental focus in English : patterns of variation in vowel duration. *Journal of Phonetics*, 32, 493-516.

[6] Chen, Y., & Gussenhoven, C. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 36, 724-746.

[7] Bang, H.-Y, Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. 2018. The emergence, progress, and impact of sound change in progress in Seoul Korean: implications for mechanisms of tonogenesis. *Journal of Phonetics*, 66, 120-144.

[8] Jun, S.-A. 1993. The Phonetics and Phonology of Korean Prosody. Ph.D. Ohio State University.

Effects of prosodic structure on nasal coarticulation in Korean and L2 English

Jiyoung Jang[1], Sahyang Kim[2], and Taehong Cho[1]

[1]Hanyang University (Korea), [2]Hongik University (Korea)
jiyoungljang@gmail.com, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Previous studies on phonetics-prosody interface has well-established that fine-grained phonetic details of segments are conditioned by the higher-order prosodic structure [1]. A recent study on nasal coarticulation in English [1] has also shown that the non-contrastive coarticulatory process, which is assumed to arise as an inevitable result of a biomechanical process in connected speech [2], is modulated by prosodic structure. For example, under prominence (pitch-accent), nasal duration was lengthened but the vowel resisted the coarticulatory influence (less nasalized), demonstrating phonological enhancement of [nasality] and [orality], respectively. Phrase-initially, nasal consonant was shortened (more consonant-like) and vowel was less nasalized (more vowel-like) showing structural CV contrast enhancement. Phrase-finally, nasal was lengthened due to preboundary lengthening and vowel was more vulnerable to the coarticulatory influence (more nasalized). As detailed prosodic modulation is assumed to be specified in a language by making reference to other linguistic system of the language, it remains to be seen to what extent the prosodic modulation of this inevitable coarticulatory process is considered to be universal or language-specific.

Korean is a language that has a distinct prosodic system. While English realizes prominence through nuclear pitch accent on a lexical-stressed syllable, Korean is an edge-based language which tends to realized prominence by demarcating the edge of a prosodic unit [3]. The importance of boundary marking in Korean may lead to a prediction that greater domain-initial strengthening would come about. Also, it needs to be tested how the reported word-initial denasalization in Korean [4] would interact with this process. Alternatively, since coarticulation is biomechanically-driven in human speech, the two languages may show similar manifestation in this low-level process. To address the issues, we investigate fine-grained prosodic modulation of nasal coarticulation process in Korean, and discuss the results in relation to cross-linguistic similarities and differences by comparing them with the existing data in English. In addition, we also examine the prosodic conditioning of nasal coarticulatory process in L2 English. Together with the findings in Korean, the results will allow us to understand to what extent L2 phonetics-prosody interface is affected by native language experience, universally applicable tendency, or L2 learning.

Twelve native Korean speakers who are L2 learners of English participated. To test whether specific patterns are learned, the speakers were divided into two groups (advanced, intermediate). Example sentences for Korean CVN# are given in Table 1 (Korean #NVC materials were constructed similarly; see [1] for English materials). Acoustic nasal duration and energy were measured and normalized A1-P0 for indication of vowel nasalization (degree of nasal coarticulation) was calculated at various vowel points (relative: 25, 50, 75% points away from the nasal; absolute: 20, 40, 60ms points away from the nasal).

Results are summarized in Fig. 1. In Korean, prominence reduced vowel nasalization under focus regardless of nasal duration in both CVN# and #NVC, showing similar effects as in English. There was no systematic effect of boundary in domain-initial vowel (#NVC), but the nasal duration and energy was reduced phrase-initially, suggesting structural CV enhancement. Also, in #NVC, overall nasality at the initial portion of the vowel (adjacent to #N) was significantly lower than the later portion (which maybe in support for the denasalization process in Korean). In general, nasal coarticulatory process in Korean shows similar enhancement patterns for both prominence (phonological [orality] enhancement) and boundary (structural CV enhancement) as in English, even though the detailed strategies are different to some extent. In L2 English, prominence effects were similar to those in L1 English: while nasal duration was lengthened, vowels resisted the nasal influence. The longer nasal duration, which was shorter in L1 Korean, indicates that L2 learners acquired English-specific modulation. As for (pre)boundary effect, learners reduced the vowel nasality in phrase-initial position (as native English speakers) at the initial portion of the vowel (25% for relative and 20, 40ms for absolute points). Also, they did not show extreme denasalization which was detected in their L1 production. These results suggest that the learners acquired English-like pattern for #NVC. For phrase-final position (CVN#), there was no effect of coarticulatory vulnerability on the vowel (as in L1 English and Korean), but the nasal duration showed preboundary lengthening. Group did not show significant main effect, suggesting that modulation for the non-contrastive coarticulation is learned in the earlier stage of L2 acquisition.

In all, the present study implies that prosodic strengthening on the non-contrastive nasal coarticulatory process generally shows cross-linguistically similar enhancement patterns. It also has further implications on the theories of L2 phonetics which has been focusing on L1 influence on L2 in the segmental level that L2 production can be more fully understood by considering the phonetics-prosody interface into account.

Table 1 Example sentences in Korean CVN#. Target words are underlined and focused words are in bold.

IP-final	**Focused**	[ani]. IP [k^{h}ɯnap*a **pam**]. IP [twɛs.sʌ]? “*No. It’s big uncle’s night. Did you get it right?*” (as an answer to “This time, is the card big uncle’s **rice** ([pap])?”)
	Unfocused	[ani]. IP [**k^{h}ɯn**ap*a pam]. IP [twɛs.sʌ]? “*No. It’s* ***big*** *uncle’s night. Did you get it right?*” (as an anwer to “This time, is the card **little** uncle’s night?”)
Wd-final	Focused	[ani]. IP [k^{h}ɯnap*a **pam** twie]. IP [twɛs.sʌ]? “*No. Right to big uncle’s* ***night****. Did you get it right?*” (as an answer to “This time, do I place the card right to big uncle’s **rice**?”)
	Unfocused	[ani]. IP [**k^{h}ɯn**ap*a pam twie]. IP [twɛs.sʌ]? “*No. Right to* ***big*** *uncle’s night. Did you get it right?*” (as an answer to “This time, do I place the card right to **little** uncle’s night?”)

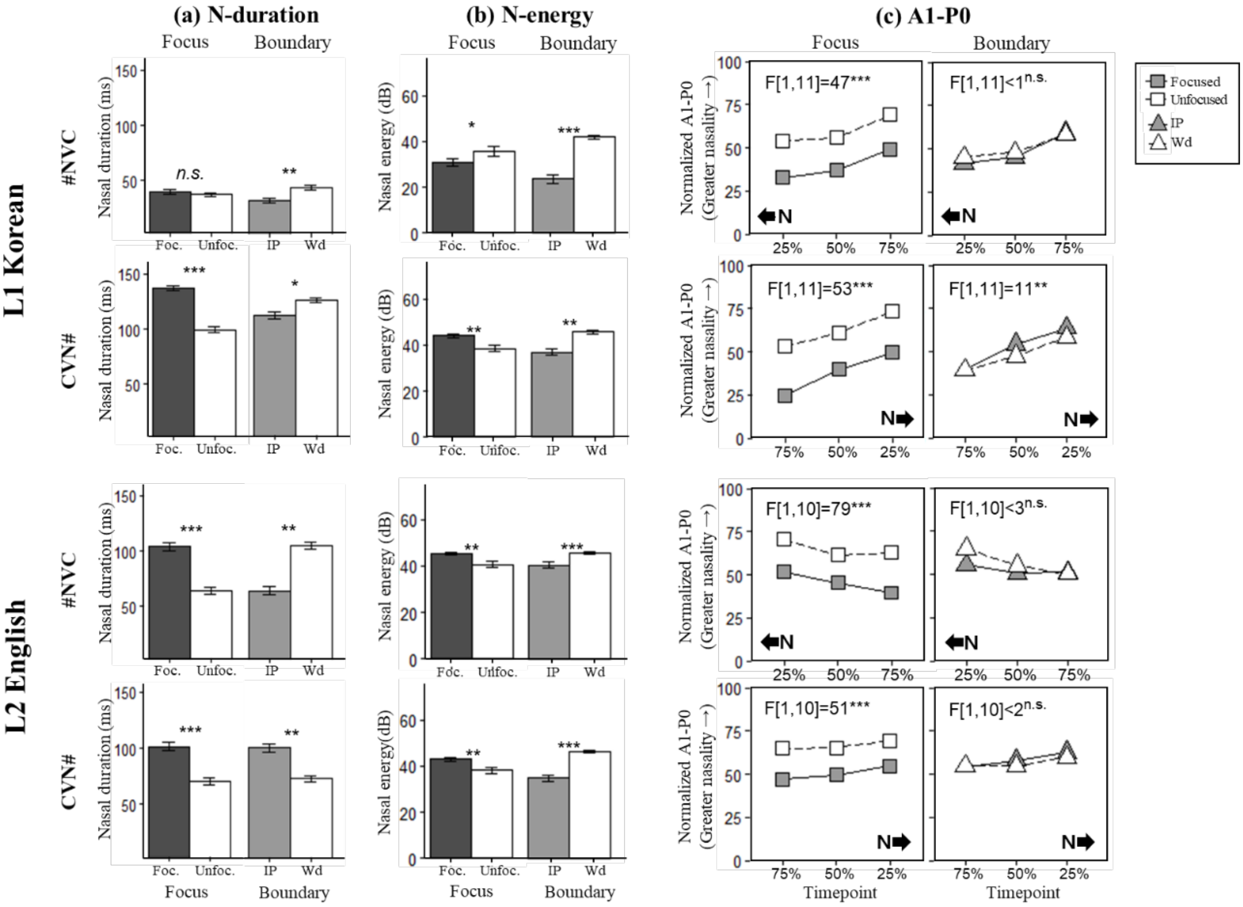


Fig. 1 (a) Mean nasal duration (ms). (b) Mean nasal energy (dB). (c) Mean normalized A1-P0 at various vowel points (relative 25, 50, 75% points from the nasal). (***, **, *, and n.s. refer to p<.001, p<.01, p<.05, and p>.09, respectively.)

References

[1] Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics*, 64, 71-89.

[2] Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. *Coarticulation: Theory, data and techniques*, 7-30.

[3] Jun, Sun-Ah. (2005). Prosodic Typology. *Prosodic Typology: The Phonology of Intonation and Phrasing*, 430-458. Oxford University Press.

[4] Yoshida, K. (2008). Phonetic implementation of Korean denasalization and its variation related to prosody. *IULC Working Papers*, 8(1).

Constituent edges and final vowel lengthening in bilingual Drehu

Catalina Torres[1 2], Janet Fletcher[1 2], Gillian Wigglesworth[1 2]

[1]University of Melbourne (Australia), [2]ARC Centre of Excellence for the Dynamics of Language (Australia)
catalinat@student.unimelb.edu.au, j.fletcher@unimelb.edu.au, g.wigglesworth@unimelb.edu.au

This study focuses on final vowel lengthening at the edges of prosodic constituents in Drehu, an under-described Oceanic language spoken in Lifou, New Caledonia. French and Drehu have been in contact for over a decade and today only a few monolingual speakers of Drehu remain in Lifou. In this paper, we seek to investigate duration patterns of word final vowels in Drehu. Our second goal is to gain better insights into speech production processes influencing bilingual Lifou French. It is found that in Drehu, at the right edge of constituents, final vowel lengthening is not used extensively, similar to languages with a quantity system like Northern Finnish [1].

Drehu has a syllabic structure allowing for V, CV, CVV, CVC combinations and a seven vowels system, which all include a length distinction [2]. Further, Drehu was impressionistically described as a language with lexical prominence marking [3]. There are however no phonetic studies describing prominence marking, prosodic constituents or a prosodic hierarchy in Drehu. This paper represents a first attempt in the study of the phonetics of the prosodic structure of the language. A previous study on boundary marking at the right edge of the AP (Accentual phrase) and ip (intermediate phrase) in Lifou French, shows that bilinguals mainly rely on pitch span expansion and not on final syllable lengthening to distinguish between prosodic breaks [4]. In Standard French, final syllable and vowel lengthening have been described as strong correlates distinguishing AP- from ip- boundaries [5]. Although in Lifou French final syllable lengthening occurs, it does not represent a major factor distinguishing an AP from an ip boundary. Hence, it is hypothesised this originates from cross-linguistic influence from Drehu. Similarly to [5], we looked at word final vowels at constituent edges of polysyllabic words, all nouns, containing only short vowels and ending on CV.

Four female speakers (aged 29-44) were recorded performing a word insertion task displayed on a pc screen. Speakers responded to a sociolinguistic questionnaire and reported they acquired both languages during early childhood (starting with either language no later than at 6 years), and use both regularly. A total of 37 words, ending on a vowel (/a/, /e/, /ɛ/, /i/, /o/, /u/) was selected for the analysis (12x 2 syllables, 15x 3 syllables, 10x 4 syllables, and 1x 5 syllables). Words were inserted in sentence initial, medial, and final position.

Utterances were force aligned in a language independent grapheme to phoneme conversion [6]. Phoneme alignment was manually corrected, all target tokens were labelled in Praat [7], and whenever a pause was inserted (potentially indicating a prosodic break), this was also noted. Measurements for the duration of all vowels were taken and then fitted into a linear mixed effects model in order to investigate if the position in the carrier phrase and the insertion of the pause had an effect on duration.

We employed the step function in order to find the best model and conducted a bonferroni post hoc test to confirm significance of results. The model included 196 observations, vowel, and position of the target token within the carrier phrase as fixed factors, and speaker, plus target word as random factors. Preliminary results for two speakers show there is no significant difference at sentence initial, medial or final position but a significant difference between these three positions and the insertion of a pause at sentence medial position ($p<0.0001$), (see Fig. 1). Presumably prominence marking and boundary marking coincide at this edge but the reason for this difference cannot yet be concluded and must be further investigated.

Studies on bilingual prosody and rhythm have prioritized better studied or genetically closely related language-pairs like Spanish and English or Spanish and Catalan [8,9]. This study emphasizes the importance of implementing laboratory phonological methods in the field. It could be established that in Drehu final vowel lengthening is used similarly to languages like Northern Finish, hence in a restricted way. This knowledge helps understanding why in Lifou French final syllable lengthening is not a significant factor distinguishing between prosodic breaks. The specificities of the length distinction in Drehu and how it can influence lengthening in Lifou French must be further investigated. Future research will therefore examine the realization of long versus short vowels in Drehu.

Carrier phrases with token in (1) sentence initial, (2) medial and (3) final positions. Examples are given in the current Drehu orthography.

(1) ___ la ëjen qene drehu
___ ART name language drehu (X we call it in Drehu)

(2) Ame la ___ tre ka lolo
PRES1 ART ___ PRES2 STAT beautiful (This X is beautiful)

(3) Ngöne la qene drehu kola hape ___
In ART language Drehu PREDIC say___ (In Drehu we say X)

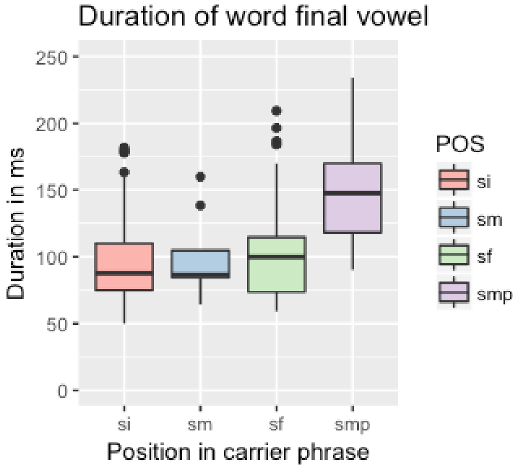


Fig.1 Duration of word final vowels in ms. Positions : (si) sentence initial, (sm) sentence medial, (sf) sentence final, and (smp) sentence medial with following pause

References

[1] Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of phonetics*, *37*(1), 29-45.

[2] Moyse-Faurie, C. (1983). Le drehu, langue de Lifou (Îles Loyauté). Phonologie, morphologie, syntaxe. *Langues et Cultures du Pacifique Ivry*, (3), 1-212.

[3] Tryon, D. T. (1968). Dehu grammar.

[4] Torres, C., Fletcher, J., Wigglesworth, G. (2018). Acoustic correlates of the French Accentual Phrase in Lifou (New Caledonia). In *Proc. Speech Prosody*.

[5] D'Imperio, M., & Michelas, A. (2014). Pitch scaling and the internal structuring of the Intonation Phrase in French. *Phonology*, *31*(1), 95-122.

[6] Reichel, U. D. (2012). PermA and Balloon: Tools for string alignment and text processing. In *Proc. Interspeech*.

[7] Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 05)[Computer program]. Retrieved May 1, 2009.

[8] Schmidt, E., & Post, B. (2015). The development of prosodic features and their contribution to rhythm production in simultaneous bilinguals. *Language and speech*, *58*(1), 24-47.

[9] Simonet, M. (2011). Intonational convergence in language contact: Utterance-final F0 contours in Catalan–Spanish early bilinguals. *Journal of the International Phonetic Association*, *41*(2), 157-184.

Consonant and Vowel Co-articulation in Beijing Mandarin*

Wai-Sum Lee

City University of Hong Kong (Hong Kong)
w.s.lee@cityu.edu.hk

There have been a number of studies of consonant and vowel co-articulation in syllable production, through analysis of formant frequencies ([1, 2, 3]) and/or articulator movements ([4, 5, 6, 7, 8]) at the consonant-vowel transition. The present study investigates the co-articulation of the intervocalic plosive consonants with the neighboring vowels in Beijing Mandarin (BM). It determines the variations in the co-articulation strength between the plosives [p t k] and the preceding and the following vowels in (C)V_1CV_2 bisyllables, by examining the tongue positions at the V_1-C and C-V_2 transitions.

The test materials used for the investigation included 21 (C)V_1CV_2 bisyllabic words in BM (Table 1), where the intervocalic C = [p t k] and the neighboring V_1 and V_2 = [i a u]. Ten repetitions of each test bisyllable were elicited from native BM speakers, male and female, in their early twenties. EMA AG500 was used to record synchronized audio signals and articulatory actions during the test bisyllables at the three different points on the tongue: tongue front (TF), tongue middle (TM), and tongue back (TB). By making reference to the waveforms and spectrograms of the bisyllables, temporal points of the vowel were selected, specifically (i) the mid-point and the offset of V_1 for determining the V_1-C co-articulation and (ii) the onset and the mid-point of V_2 for determining the C-V_2 co-articulation. A large or small change in the Euclidean distance (ED) between the positions of the points on the tongue for two temporal points of the vowel is taken to indicate a respective weak or strong co-articulation strength (CS) between the plosive and the neighboring vowel.

Articulatory data elicited from three male BM speakers were analyzed. Summary results for a male BM speaker based on the variations in ED (Table 2) are presented as follows. In general, the ED data of the three BM speakers are similar.

(1) The difference in ED at any one of the three tongue points (TF, TM, TB) between the temporal mid-point and the offset of V_1 is small, where V_1 = [i] or [u] is followed by [p] (0.70-2.09 mm). This indicates a strong CS between the high vowel [i] or [u] and the following [p]. The CS between [a] (V_1) and the following [p] decreases, as there is an increase in the difference in ED between the mid-point and the offset of [a] (3.61-3.87 mm). The CS is also stronger between V_2 and the preceding [p] when V_2 = [i] or [u] than when V_2 = [a]. This is evidenced by a smaller difference in ED between the onset and the mid-point of [i] or [u] (1.08-1.58 mm) than between the onset and the mid-point of [a] (3.01-3.14 mm). The CSs are similar between the following [p] and V_1 and between the preceding [p] and V_2.

(2) The CS is weak between [t] and the neighboring V_1 or V_2. Between the mid-point and the offset of V_1, the difference in ED is larger at the tongue point TF (4.28-18.52 mm) than at the other two tongue points, TM (1.97-11.35 mm) and TB (0.82-10.61 mm). Between the onset and the mid-point of V_2, the difference in ED is also larger at TF (7.81-19.74 mm) than at TM (2.55-12.82 mm) and TB (1.42-12.70 mm). The data suggest TF is less compatible than TM and TB with the tongue gesture for articulation of a vowel. The differences in ED of any one of the three tongue points are smaller when V_1/V_2 = [i] (0.82-7.81 mm) than when V_1/V_2 = [a] (10.61-19.74 mm) or [u] (9.30-16.58 mm). The data indicate that the CS is stronger between the alveolar [t] and the front [i] than between [t] and the low [a] or the back [u]. There is no large difference in ED between V_1-[t] and [t]-V_2, an indication that the CSs between V_1 and the following [t] and between the preceding [t] and V_2 are similar.

* This research is supported by an International Scholarly Exchange Research Grant (#9500023/RG003-P-14) from the Chiang Ching-kuo Foundation, Taiwan.

(3) The CS is stronger between [k] and the neighboring vowel [i] or [u] than [a]. For V_1 preceding [k], the difference in ED between the mid-point and the offset of V_1 is smaller when V_1 = [i] (2.06-5.09 mm) or [u] (2.33-4.97 mm) than when V_1 = [a] (6.93-13.15 mm). This is due to the similarity in articulation between the velar [k] and the high vowel [i] or [u], and the competition in articulation between [k] and the low vowel [a]. The ED between [u] in the V_2 position and the preceding [k] (3.52-4.47 mm) is similar to the ED between [u] in V_1 position and the following [k] (2.33-4.97 mm), which indicates the similarity in CS between V_1-[k] and between [k]-V_2.

(4) Based on the average ED for the three tongue points, the extent of variation in tongue position between V_1 and C is larger when C = [t] (9.35 mm) than when C = [k] (5.88 mm) and [p] (2.22 mm). Between C and V_2, a larger variation in the average ED is also observed when C = [t] (10.66 mm) than when C = [k] (4.07 mm) and [p] (1.90 mm). Thus, the descending order of CS between the intervocalic C and the neighboring V_1 and V_2 in BM (C)V_1CV_2 bisyllables is C = [p] > C = [k] > C = [t]. Furthermore, the small difference in CS between V_1-C and C-V_2 may suggest articulatory ambisyllabicity in the intervocalic consonant in BM bisyllables.

The results of this study will be discussed in connection with the findings about the CS in C-V and V-C co-articulation in other languages reported in previous studies.

Table 1 Test (C)V1CV2 bisyllables in BM, where the intervocalic C = [p t k] and V1 and V2 = [i a u] (* = non-occurring)

	With an intervocalic [p]			With an intervocalic [t]			With an intervocalic [k]		
V_1 V_2	[i]	[a]	[u]	[i]	[a]	[u]	[i]	[a]	[u]
[i]	[**ipi**]	[**ipa**]	[**ipu**]	[**iti**]	[p**ita**]	[**itu**]	*	*	[**iku**]
[a]	[f**ap**i]	[p**apa**]	[f**apu**]	[p**ati**]	[f**ata**]	[p**atu**]	*	*	[p**aku**]
[u]	[**upi**]	[k**upa**]	[**upu**]	[**uti**]	[**uta**]	[k**utu**]	*	*	[**uku**]

Table 2 Euclidean distances (in mm) of each of the three tongue points (TF, TM, TB) between the mid-point and the offset of V1 and between the onset and the mid-point of V2 in (C)V1CV2 bisyllables for a male BM speaker

	Neighboring with [p]			Neighboring with [t]			Neighboring with [k]		
V_1	TF	TM	TB	TF	TM	TB	TF	TM	TB
[i]	1.79	1.33	0.70	4.28	1.97	0.82	5.09	2.39	2.06
[a]	3.87	3.91	3.61	18.52	11.37	10.61	6.93	12.05	13.15
[u]	1.01	1.64	2.09	16.58	10.66	9.30	4.97	3.98	2.33
V_2	TF	TM	TB	TF	TM	TB	TF	TM	TB
[i]	1.23	1.26	1.08	7.81	2.55	1.42	--	--	--
[a]	3.14	3.01	3.03	19.74	12.82	12.70	--	--	--
[u]	1.33	1.58	1.42	15.61	11.75	11.57	3.52	4.47	4.22

References

[1] Krull, D. (1988). Consonant-vowel co-articulation in spontaneous speech and in reference words. *PERILUS*, *7*, 1-149.

[2] Sussman, H.M., Bessell, N., Dalston, E., & Majors, T. (1997). An investigation of stop place of articulation as a function of syllable position: a locus equation perspective. *JASA*, *101*, 2825-2838.

[3] Modarresi, G., Sussman, H., Lindblom, B., & Burlingame, E. (2004). Stop place coding: an acoustic study of CV, VC#, and C#V sequences. *Phonetica*, *61*, 2-21.

[4] Recasens, D. (1984). V-to-C co-articulation in Catalan VCV sequences: an articulatory and acoustical study. *Journal of Phonetics*, *12*, 61-73.

[5] Farnetani, E. (1990). V-C-V lingual co-articulation and its spatiotemporal domain. In W.J. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling* (pp. 93-130). The Netherlands: Kluwer Academic Publishers.

[6] Fowler, C.A. & Brancazio, L. (2000). Co-articulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel co-articulation. *Language and Speech*, *43*, 1-41.

[7] Recasens, D. & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *JASA*, *125*, 2288-2298.

[8] Iskarous, K., Fowler, C.A., & Whalen, D.H. (2010). Locus equations are an acoustic expression of articulator synergy. *JASA*, *128*, 2021-2032.

The roles of featural distance and type of featural change in word recognition: A pupillometry study

Katalin Tamási & W. Quin Yow

Humanities, Arts and Social Sciences, Singapore University of Technology and Design (Singapore)
katalin.tamasi;quin@sutd.edu.sg

Introduction. This study investigates whether mature lexical representations are encoded in units that are isomorphic to features. If that is the case, lexical activation is expected to be proportional to the degree of featural match between input and lexical representation (TRACE model: [1]). Support for this hypothesis comes from research that manipulate *featural distance* by introducing one vs. more featural changes to words [2–7]. These studies find that changing one feature affects word recognition to a lesser extent than changing more than one feature. Another line of research manipulates *type of featural change*, showing how individual featural changes may differ in terms of perceptual and functional relevance [8–11]. Our study combines the two approaches by investigating the roles of both featural distance and type of featural change in word recognition using a minimally demanding online method that requires no conscious response: the single-picture pupillometry paradigm [12].

Method. In each trial, a pictured referent and a corresponding auditory label are presented while the participant's pupillary response is recorded. By manipulating the number of featural changes in the label onset (0–3 changes, e.g., *baby ~ daby ~ taby ~ saby*), we test whether the pupil dilation is proportional to the featural distance between the heard label and the lexical representation. As the type of featural change is counterbalanced, the effects of place of articulation, manner of articulation, and voicing changes on word recognition can also be studied. To investigate the potential impact of language background (as pointed out by [9]), two populations were tested: German-speaking adults in Germany, who were presented with German labels, and English–Mandarin-speaking adults in Singapore, presented with English labels.

Results and discussion. Our results support the existence of a relationship between pupil dilation and featural distance. In both populations, correct forms elicit smaller pupillary responses than one- and two-featural changes, which in turn elicit smaller responses than three-featural changes (c.f., Fig. 1). This pattern indicates sensitivity to the degree of mispronunciation and, as such, corroborates previous work that found word recognition to be modulated by featural match ([2–7]), further advancing the notion that lexical representations are composed of feature-like units. The lack of difference between the one- and two-feature change conditions may stem from diversity among the types of featural change. Our results lend support to this possibility: In both populations, place change elicits a larger degree of pupil dilation relative to manner and voicing changes, indicating that place of articulation is the most perceptually salient and/or the most functionally relevant feature in word recognition. These results are consistent with previous research that found the place of articulation feature to be more perceptually and functionally relevant than the voicing feature in several languages (ENGLISH: [8]; DUTCH: [9]; FRENCH: [10, 11]), converging to show that features contribute to a differing extent to the buildup of lexical representations.

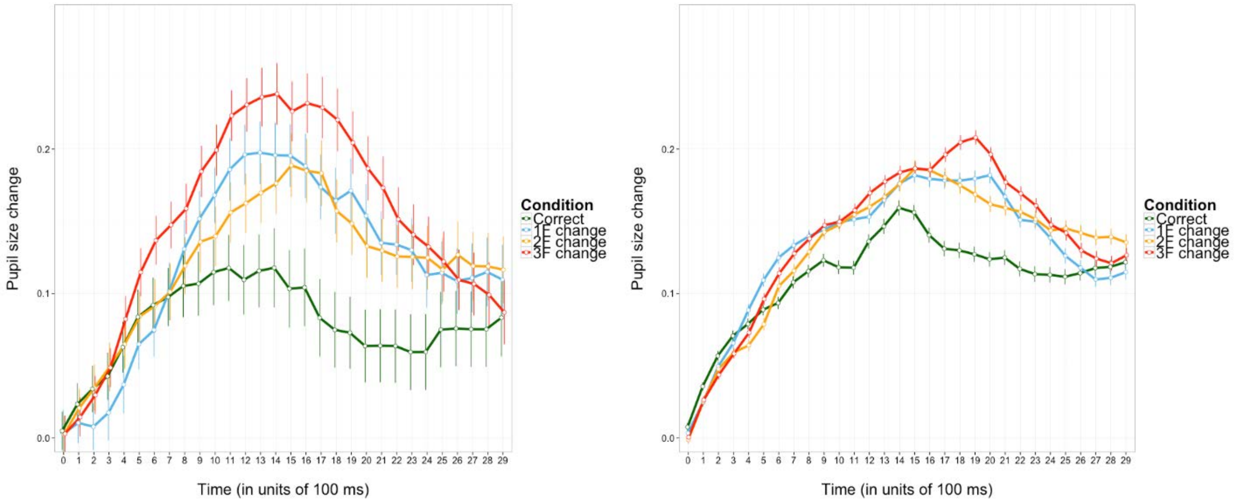


Fig.1 Corrected pupil size change (mm) over time in response to differing degrees of mispronunciation (error = 95% CI). Left: German participants (N = 24, gaze data sampling rate: 60 Hz), right: Singaporean English-Mandarin speaking participants (N = 39, gaze data sampling rate: 300 Hz). Time is zeroed on the label onset and is provided in units of 100 ms.

References

[1] McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18 (1)*, 1–86.

[2] Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language, 37 (4)*, 463–480.

[3] Lukatela, G., Eaton, T., Lee, C., & Turvey, M. (2001). Does visual word identification involve a sub-phonemic level? *Cognition, 78 (3)*, B41–B52.

[4] Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: HPP, 22 (6)*, 1376–1392.

[5] Milberg, W., Blumstein, S., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin Psychonomic Soc*, *26 (4)*, 305–308.

[6] Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: HPP*, *37 (2)*, 496–505.

[7] White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *J of Memory and Language, 68 (4)*, 362–378.

[8] Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America, 64 (1)*, 44–56.

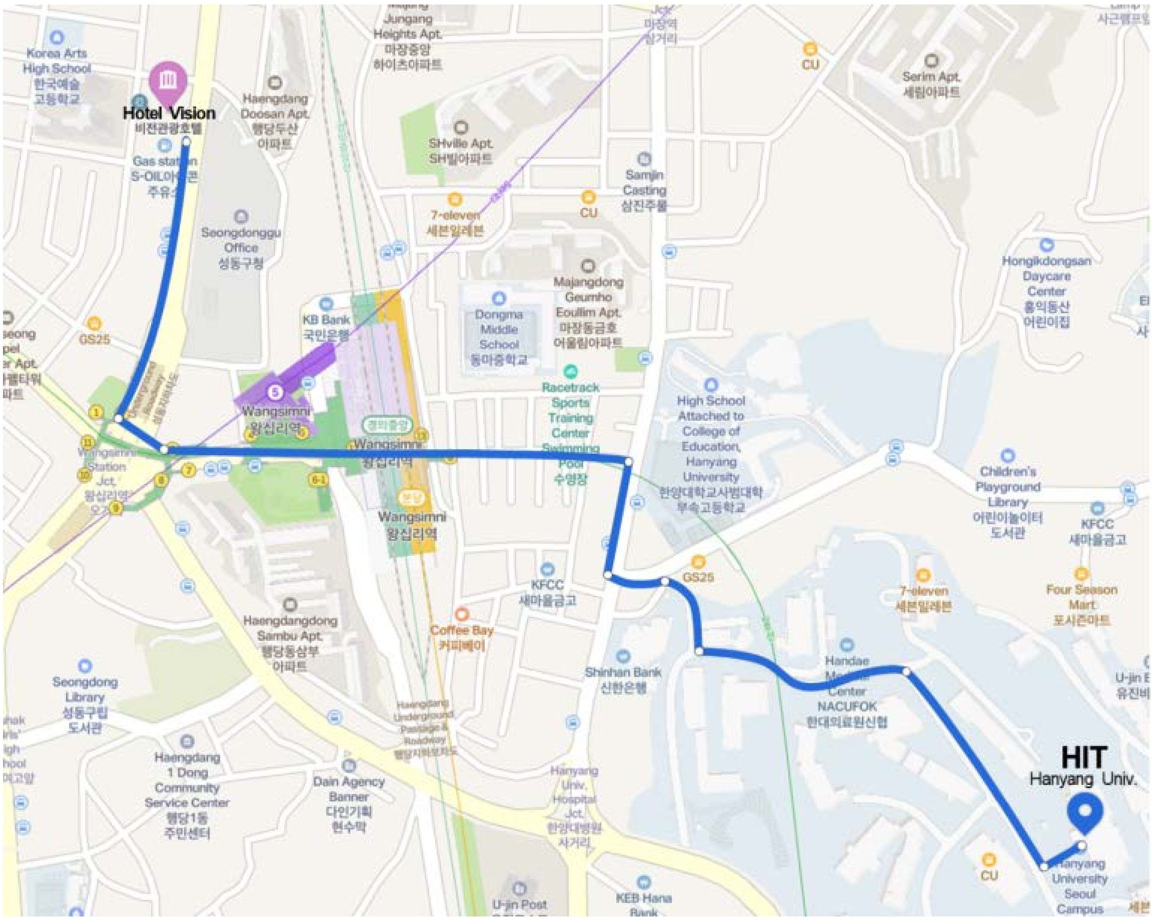
[9] Ernestus, M., & Mak, W. M. (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language, 90 (1-3)*, 378–392.

[10] Martin, A., & Peperkamp, S. (2015). Asymmetries in the exploitation of phonetic features for word recognition. *The Journal of the Acoustical Society of America, 137 (4)*, 307–313.

[11] Martin, A., & Peperkamp, S. (2017). Assessing the distinctiveness of phonological features in word recognition: Prelexical and lexical influences. *Journal of Phonetics*, *62*, 1–11.

[12] Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (2017). Pupillometry registers toddler’s sensitivity to degrees of mispronunciation. *J of Exp Child Psychology, 153*, 140–148.

FROM HOTEL VISION TO CONFERENCE VENUE



1. On Foot: You can walk from the hotel to Hanyang University. Please refer to the direction indicated in the map below.

- **Estimated time of travel**: 20 mins

2. By Bus

(1) Take Bus No. 2222 from Seongdonggu Office bus stop (성동구청; 1 min away from the hotel towards Wangsimni Station).

(2) After 3 stops, get off at Hanyang University bus stop (한양대정문).

(3) Walk to HIT.

- **Estimate time of travel**: 15 mins (5 mins for the bus ride)
- **Fee**: 1,200KRW/1.20USD

3. By Subway

(1) Take subway Line 2🟩 (green line) from Wangsimni Station (왕십리역).

(2) Get off at the next station, Hanyang University Station (한양대역). Exit from Exit 2. Walk to HIT.

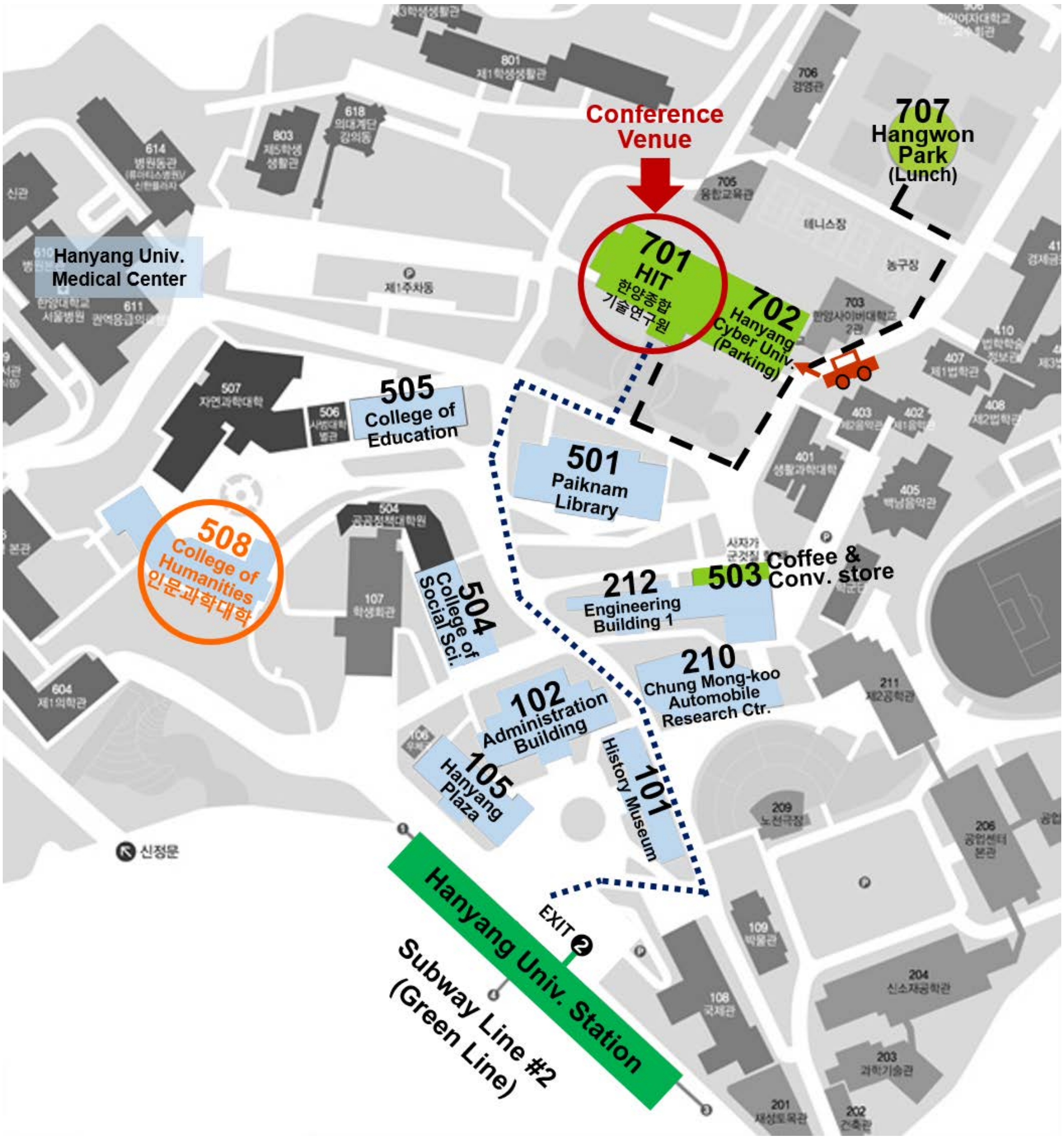
- **Estimated time of travel**: 15 mins (2 mins for the subway)
- **Fee**: 1,250KRW/1.20USD

4. By Taxi: Take a taxi in front of the hotel and ask the driver to go to HIT (pronounced H-I-T) in Hanyang University.

- **Estimated time of travel**: 10 mins
- **Fee**: 3,000 ~ 4,000KRW/3~4USD

*** Note that **minimum fare** for a taxi in Korea is 3,000KRW/3USD. You must pay the fare according to the taxi meter. In Seoul, you can get to most of the places at a rate less than 10,000KRW/10USD. Please be aware of this so that you do not get overcharged.

CAMPUS MAP AND AMENITIES



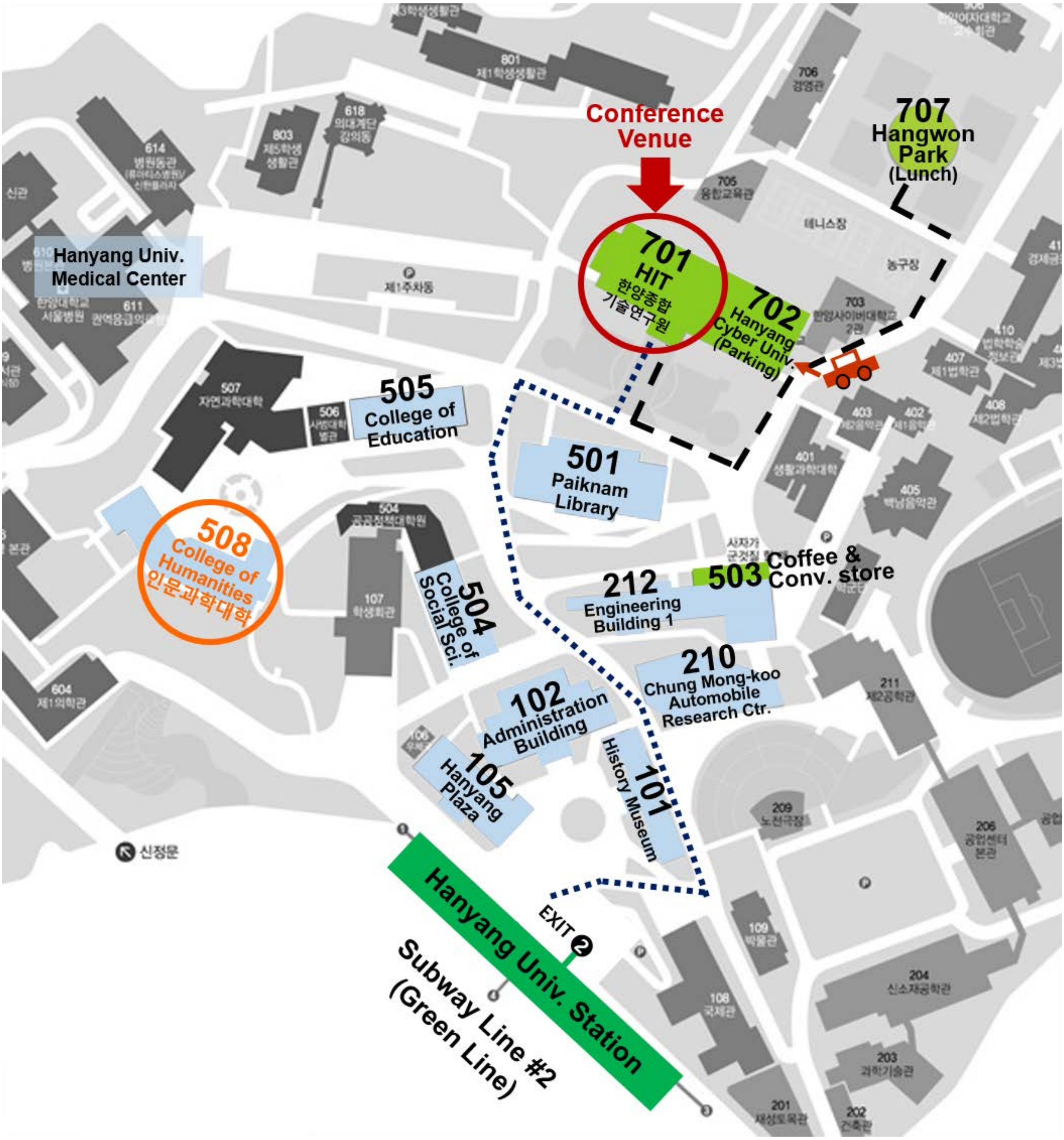
- ➢ The easiest way to locate yourself on campus is to find building numbers around you. Every building has its own building number written on the outside wall.
- ➢ Free parking tickets will be provided at the registration desk (up to 12 hours/day). You may park at any legal parking stall on campus marked by a white line, but the closest is the underground parking structure at the Hanyang Cyber University building (**Building #702**) on the B1&B2 levels. The entrance to #702 is indicated by a red arrow on the map. Since #702 is annexed to HIT, you can simply take the elevator from the parking levels to the 6th level.
- ➢ Lunch will be provided on both days at a cafeteria located at Hangwon Park (**Building #707**, B1 level).
- ➢ A nearby coffee shop and a convenience store (Building #503) can be found in front of the Engineering Building 1 (**Building** #212).
- ➢ Free Wi-Fi is available at the conference venue (**HIT, Building #701**). You can access to the network named "HYU-wlan(Free)" without a password. Note that this network may not be available at other places on campus.

HISPhonCog 2018: Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2018

May 18, 2018.
Hanyang University, Seoul, Korea

인쇄일	:	2018년 5월 18일
발행일	:	2018년 5월 18일
발행인	:	조태홍
발행처	:	한양 음성•언어인지과학연구소

Printing Date	:	May 18, 2018
Publishing Date	:	May 18, 2018
Publisher	:	Cho, Taehong
Publishing Place	:	Hanyang Institute for Phonetics and Cognitive Sciences of Language



ISSN 2586-7776