2019 Vol.2 ISSN 2586-7997

HISPhonCog

Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2019

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

Linguistic and cognitive functions of prosody and higher-order linguistic structures in speech production and perception in native and non-native languages

May 24-25, 2019, Hanyang University, Seoul, Korea

Edited by Taehong Cho
Sahyang Kim
Jonny Jungyun Kim
Say Young Kim

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

HIPCS
Hanyang Institute for Phonetics and Cognitive Sciences of Language

HANYANG UNIVERSITY

PROGRAM AT A GLANCE

Day 0: May 23 (Thursday) 2019, Satellite Workshop	
09:30-10:00	*Onsite (pre) Registration*
10:00-13:00	**Workshop 1**: Sun-Ah Jun, Sahyang Kim & Taehong Cho Theoretical and Practical Issues on Korean ToBI (K-ToBI)
14:00-15:00	*Onsite (pre) Registration*
15:00-19:00	**Workshop 2**: Anne Cutler & Mark Antoniou Cognition and Bilingualism: Why there is debate about the cognitive advantage of being bilingual

Day 1: May 24 (Friday) 2019	
08:00-09:00	*Registrations (coffee & munch)*
09:00-09:10	**Opening remarks** (Univ. President: **Woo-seung Kim**; HIPCS Director: **Taehong Cho**)
General Session 1	Moderator: Mira Oh (Chonnam Nat. U.)
09:10-10:00	Invited Speaker 1: Mary Beckman (Ohio State University)
10:00-10:40	Invited Speaker 2: Sun-Ah Jun (UCLA)
10:40-11:00	*Coffee Break*
General Session 2	Moderator: Jeong-Im Han (Konkuk U.)
11:00-11:40	Invited Speaker 3: Michael Tyler (Western Sydney University & MARCS)
11:40-12:25	Oral Presentation (3 talks)
12:25-13:40	*Lunch*
General Session 3	Moderator: Minjung Son (Hannam U.)
13:40-14:20	Invited Speaker 4: Jason Shaw (Yale University)
14:20-15:05	Oral Presentation (3 talks)
Poster Session 1	
15:05-16:45	30 posters, *with coffee*
General Session 4	Moderator: Sahyang Kim (Hongik U.)
16:45-17:30	Oral Presentation (3 talks)
17:30-18:10	Invited Speaker 5: Cécile Fougeron (Paris 3, Sorbonne, CNRS)
18:20-20:30	*Banquet: HIT, 6th floor, Conference Venue*

Day 2: May 25 (Saturday) 2019	
08:30-09:00	Registrations (coffee & munch)
Special Session on Neuro-Cognitive Aspects of Prosody	Moderator: Say Young Kim (Hanyang U.)
09:10-09:50	Invited Speaker 6: Karsten Steinhauer (McGill University)
09:50-10:30	Invited Speaker 7: Ferenc Honbolygó (Hungarian Academy of Sciences)
10:30-10:50	*Coffee Break*
10:50-11:20	Invited Discussant: Holger Mitterer (University of Malta)
11:20-12:20	Oral Presentation (4 talks)
12:20-13:30	*Lunch*

General Session 5 Moderator: Hyesun Cho (Dankook U.)	
13:30-14:10	Invited Speaker 8: Jie Zhang (University of Kansas)
14:10-14:50	Invited Speaker 9: Edward Flemming (MIT)
Poster Session 2	
14:50-16:30	29 posters, *with coffee*
General Session 6 Moderator: Taehong Cho (Hanyang U.)	
16:30-17:10	Invited Speaker 10: Martine Grice (University of Cologne)
Discussion Session Moderator: Taehong Cho (Hanyang U.)	
17:10-17:25	Invited Discussant: Mary Beckman (Ohio State University)
17:25-17:40	Invited Discussant: Anne Cutler (Western Sydney University, MARCS, ARC Centre of Excellence)
17:40-18:10	General Discussion: Chaired by Taehong Cho (HIPCS, Hanyang U.)

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

Supported by
80th Anniversary Fund from Hanyang University
Global Research Fund from Hanyang University
Global Research Network (GRN) Fund from National Research Foundation of Korea
HIPCS (Hanyang Institute for Phonetics and Cognitive Sciences of Language)

The Organizing Committee

Chair

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

Committee members

Sahyang Kim (Hongik University, Seoul)
Jonny Jungyun Kim (Hanyang University, Seoul)
Say Young Kim (Hanyang University, Seoul)

Graduate Student Staff

Jiyoung Lee (Hanyang University)
Jungyun Seo (Hanyang University)
Hongmei Li (Hanyang University)
Hyunjung Joo (Hanyang University)
Hana You (Hanyang University)

TABLE OF CONTENTS

PROGRAM IN DETAIL

Day 0: May 23 (Thursday) 2019, Satellite Workshop	
09:30-10:00	*Onsite (pre) Registration*
10:00-13:00	**Workshop 1: Sun-Ah Jun, Sahyang Kim & Taehong Cho** Theoretical and Practical Issues on Korean ToBI (K-ToBI)
14:00-15:00	*Onsite (pre) Registration*
15:00-19:00	**Workshop 2: Anne Cutler & Mark Antoniou** Cognition and Bilingualism: Why there is debate about the cognitive advantage of being bilingual

Day 1: May 24 (Friday) 2019	
08:00-09:00 09:00-09:10	*Registrations (coffee & munch)* **Opening remarks** (Univ. President: **Woo-seung Kim**; HIPCS Director: **Taehong Cho**)
General Session 1	Moderator: Mira Oh (Chonnam Nat. U.)
09:10-10:00	Invited Speaker 1: **Mary Beckman** (Ohio State University) Connecting prosody to duality of patterning in two modalities (p. 9)
10:00-10:40	Invited Speaker 2: **Sun-Ah Jun** (UCLA) Adaptation to prosody in anticipatory structural analysis during visual search (p. 11)
10:40-11:00	*Coffee Break*
General Session 2	Moderator: Jeong-Im Han (Konkuk U.)
11:00-11:40	Invited Speaker 3: **Michael Tyler** (Western Sydney University & MARCS) Assessing the (un)categorisation of non-native phones to native phonological categories (p. 13)
11:40-11:55	(Oral 01) **Xin Wang, Siyu Chen, Bronson Hui** (Macquarie U., U. Greenwich, Michigan State University) The role of lexical tone in bilingual language processing (p. 31)
11:55-12:10	(Oral 02) **Xinran Ren & Peggy Pik Ki Mok** (Chinese U. of Hong Kong) Production of Korean Accentual Phrases by Cantonese-speaking learners (p. 33)
12:10-12:25	(Oral 03) **Youngon Choi, Minji Nam, Reiko Mazuka, Hyun-Kyung Hwang, Minha Shin, Jisoo Kim, Sunho Jung, Jimin Beom, & Naoto Yamane** (Chung-Ang U., RIKEN Center for Brain Sci., Duke U.) Development of speech perception in Korean infants: Discriminating unusual sound contrasts with diachronic change (p. 35)
12:25-13:40	*Lunch*
General Session 3	Moderator: Minjung Son (Hannam U.)
13:40-14:20	Invited Speaker 4: **Jason Shaw** (Yale University) Common coordination patterns in prosodic and segmental domains (p. 15)
14:20-14:35	(Oral 04) **Miao Zhang** (U. at Buffalo, SUNY) The role of information structure and foot structure on the kinematics of unstressed syllables (p. 37)
14:35-14:50	(Oral 05) **Hongmei Li, Sahyang Kim & Taehong Cho** (HIPCS; Hanyang U.; Hongik U.) Language-specific prosodic structural modulation of coarticulation vowel nasalization

14:50-15:05	in #NV and CVN# in Mandarin Chinese (p. 39)
	(Oral 06) **Tabea Thies, Doris Muecke, Anja Lowit, Elke Kalbe, Julia Steffen & Michael T. Barbe** (U. of Cologne, U. of Strathclyde, U. Hospital Colonge) Cognitive skills and prominence production: Highlighting prominent elements in the speech of patients with Parkinson's disease (p. 41)
Poster Session 1	
15:05-16:45	30 posters (see below), *with coffee*
General Session 4	Moderator: Sahyang Kim (Hongik U.)
16:45-17:00	(Oral 07) **Shigeto Kawahara, Jason Shaw & Shinichiro Ishihara** (Keio U., Yale U., Lund U.) Assessing tonal specifications through simulation and classification: The case of post-wh accent in Japanese (p. 43)
17:00-17:15	(Oral 08) **Jason Bishop** (CUNY-College of Staten Island, CUNY-Graduate Center) Do individual differences in working memory capacity predict cross-speaker variation in planning scope? Some prosodic tests (p. 45)
17:15-17:30	(Oral 09) **Jeremy Steffman** (UCLA) Distal rhythmic structure and speech rate in perception: Exploring prosodic and rate-dependent influences (p. 47)
17:30-18:10	Invited Speaker 5: **Cécile Fougeron** (Paris 3, Sorbonne, CNRS) Prosodic position effects: Strengthening for withstanding? (p. 17)
18:20-20:30	*Banquet: HIT, 6th floor, Conference Venue*

Poster Presentation (Day 1, 15:05-16:45, May 24, Friday, 2019)

(P01) **Julien Eychenne & Yejin Shon** (Hankuk U. of Foreign Studies)
A cross-linguistic analysis of phonological neighborhood density (p. 61)

(P02) **Kevin Samejon** (Notre Dame of Marbel U.)
Vowel length of emphasized Cebuano adjectives (p. 63)

(P03) **Yen-Chen Hao & Kenneth de Jong** (U. of Tennessee; Indiana U. Bloomington)
English obstruent perception by native Mandarin, Korean, and English speakers (p. 65)

(P04) **Young Hwang, Samson Lotven & Kelly Berkson** (Indiana U. Bloomington)
Pitch accent and the three-way laryngeal contrast in North Kyungsang Korean (p. 67)

(P05) **Rory Turnbull** (U. of Hawaiʻi at Mānoa)
Choices in abstract phonological analysis have direct consequences for psycholinguistic predictions: the case of phonological neighbourhood networks (p. 69)

(P06) **Haerim Hwang** (U. of Hawai'i at Mānoa)
Second language perception of English stops by Korean-speaking child learners: Effects of position and lexical knowledge (p. 71)

(P07) **Sang-Im Lee-Kim** (National Chiao Tung U.)
Effects of orthographic input on L2 production: the case of Korean-speaking learners of Mandarin Chinese (p. 73)

(P08) **Waan-Rur Lu & Sang-Im Lee-Kim** (National Chiao Tung U.)
Enhancement of sibilant contrasts in near merger during word processing by Min-Mandarin bi-dialectal speakers (p. 75)

(P09) **Nancy Hall, Bianca Godinez & Megan Walsh** (California State U., Long Beach)
Experimental evidence for perceptual hypercorrection in American r-dissimilation (p. 77)

(P10) **Jessica Chin & Mark Antoniou** (MARCS, Western Sydney U.)
The influence of native intonational and tonal categories on nonnative tone learning (p. 79)

Day 2: May 25 (Saturday) 2019	
08:30-09:00	*Registrations (coffee & munch)*
Special Session on Neuro-Cognitive Aspects of Prosody Moderator: Say Young Kim (Hanyang U.)	
09:10-09:50	Invited Speaker 6: **Karsten Steinhauer** (McGill University) Brain signatures of prosodic processing (p. 19)
09:50-10:30	Invited Speaker 7: **Ferenc Honbolygó** (Hungarian Academy of Sciences) Language specific representations in word stress perception: ERP evidence (p. 21)
10:30-10:50	*Coffee Break*
10:50-11:20	Invited Discussant: **Holger Mitterer** (University of Malta)
	Moderator: Jiyoun Choi (Sookmyung Women's U.)
11:20-11:35	(Oral 10) **Rachida Ganga, Haoyan Ge, Marijn Struiksma, Virginia Yip & Aoju Chen** (Utrecht U., Open U. of Hong Kong, Chinese U. of Hong Kong) Prosodic processing in sentences with "only" in L1 and L2 English: An ERP study (p. 51)
11:35-11:50	(Oral 11) **Nelleke Jansen, Rachida Ganga, Marijn Struiksma and Aoju Chen** (Utrecht U.) Musicality influences the processing of accentuation in sentences with 'only' by L2 English learners (p. 53)
11:50-12:05	(Oral 12) **Chiung-Yu Chang and Feng-Fan Hsieh** (National Tsing Hua U.) Neural responses to tonal and segmental gaps in Mandarin Chinese (p. 55)
12:05-12:20	(Oral 13) **Stephen Politzer-Ahles and Suyeon Im** (Hong Kong Poly Tech. U.) No word-specific mismatch negativity effect in Mandarin speakers (p. 57)
12:20-13:30	*Lunch*
General Session 5 Moderator: Hyesun Cho (Dankook U.)	
13:30-14:10	Invited Speaker 8: **Jie Zhang** (University of Kansas) How tone sandhi helps us understand the processing of phonological alternation (p. 23)
14:10-14:50	Invited Speaker 9: **Edward Flemming** (MIT) A generative phonetic analysis of the timing of L- phrase accents in English (p. 25)
Poster Session 2	
14:50-16:30	29 posters (see below), *with coffee*
General Session 6 Moderator: Taehong Cho (Hanyang U.)	
16:30-17:10	Invited Speaker 10: **Martine Grice** (University of Cologne) The role of intonation in vowel insertion (p.27)
Discussion Session Moderator: Taehong Cho (Hanyang U.)	
17:10-17:25	Invited Discussant: **Mary Beckman** (Ohio State University)
17:25-17:40	Invited Discussant: **Anne Cutler** (Western Sydney University, MARCS, ARC Centre of Excellence)
17:40-18:10	General Discussion: Chaired by **Taehong Cho** (HIPCS, Hanyang U.)

Poster Presentation (Day 2, 14:50-16:30, May 25, Saturday, 2019)
(P31) **Rachel Albar & Hiyon Yoo** (Laboratoire de Linguistique Formelle, UMR7710, CNRS/ U. Paris Diderot) Do Japanese learners distinguish prosodic levels in French? (p. 123)

(P32) **Hanbo Yan, Yu-Fu Chien & Jie Zhang** (Shanghai International Studies U.; Fudan U.; U. of Kansas)
The representation of variable tone sandhi in Shanghai Chinese (p. 125)

(P33) **Si Chen, Eunjin Chun, Bei Li & Yike Yang** (Hong Kong Polytechnic U.)
Perturbation effects in Chongming Chinese with and without focus (p. 127)

(P34) **Chen Lan & Peggy Mok** (The Chinese U. of Hong Kong)
Prosodic cues in the perception of Cantonese sarcasm (p. 129)

(P35) **Eunkyung Sung, Youngeun Kim & Sooyeon Lee** (Cyber Hankuk U. of Foreign Studies; Hankuk U. of Foreign Studies)
The effects of sentential context on compensation for English assimilated speech by L2 listeners (p. 131)

(P36) **Shengyi Wu & Ara Cho** (Shaoxing U.; Fudan U.)
Effects of L1 and L2 experience on the perception of Korean three-way stop contrast: evidence from Chinese and Korean listeners (p. 133)

(P37) **Dong Jin Kim, Okgi Kim & Hanyong Park** (U. of Wisconsin-Milwaukee)
Prosodic realization of multiple accusative construction in Korean (p. 135)

(P38) **Jiyoung Jang, Sahyang Kim & Taehong Cho** (U. of California, Santa Barbara; Hongik U.; Hanyang U.)
Effects of prosodic structure on voice quality associated with Korean three-way stop contrast (p. 137)

(P39) **Anna Dannenberg & Stefan Werner** (U. of Helsinki; U. of Eastern Finland)
Prosodic and syntactic boundaries in spontaneous English and Finnish speech (p. 139)

(P40) **Wai Ling Law, Olga Dmitrieva & Alexander Francis** (The Chinese U. of Hong Kong; Purdue U.)
Interaction between lexical tones and stress is affected by individual variation in language attitudes and L2 experience (p. 141)

(P41) **Xiaolin Li & Peggy Mok** (The Chinese U. of Hong Kong U.)
The acquisition of Xiamen tone sandhi by children (p. 143)

(P42) **Changwei Zhang, Si Chen & Man Tak Leung** (The Hong Kong Polytechnic U.)
The Perception and Production of Cantonese Syllable-final Stops in Mandarin Speakers: An Analysis of Perceptibility Scale (p. 145)

(P43) **Mengtian Chen** (U. of Iowa)
Syntactic and length constraints on the prosodic phrasing of second language sentences (p. 147)

(P44) **Eunkyung Sung, Youngeun Kim & Sooyeon Lee** (Cyber Hankuk U. of Foreign Studies; Hankuk U. of Foreign Studies)
A comparison of perceptual compensation for Korean regressive and progressive assimilations by L2 learners (p. 149)

(P45) **Yanjun Wei, Mengtian Chen & Jianqin Wang** (Beijing Language and Culture U.; U. of Iowa)
The effects of syntactic constraints on prosodic chunking of L2 speech in sentence recall (p. 151)

(P46) **Yu-Yin Hsu, Ka-Wai Chan, Tsz-Shan Lo, Xia Wang & Anqi Xu** (The Hong Kong polytechnic U.; U. college London)
Sentence final particles and Wh-indeterminates in Beijing Mandarin (p. 153)

(P47) **Chaak-Ming Lau** (The Chinese U. of Hong Kong)
Perception and analysis of utterance-final lengthening of Cantonese maː33 (p. 155)

(P48) **Mei Ying Ki** (The Chinese U. of Hong Kong)
A phonetic study of Cantonese rising-falling intonation (p. 157)

(P49) **Daiki Hashimoto** (U. of Tokyo)
Cumulative usage effects in inflectional paradigm: Probability of being affixed (p. 159)

(P50) **Stephen Politzer-Ahles, Ka Keung Lee, Lei Pan & Jueyao Lin** (The Hong Kong Polytechnic U., City U. of Hong Kong)
Cross-linguistic differences in long-lag priming (p. 161)

Invited Talks

Connecting prosody to duality of patterning in two modalities

Mary E. Beckman

Ohio State University (USA)
beckman.2@osu.edu

This talk sketches an idea about what kind of models of spoken language prosody will be most useful in making cross-language comparisons, to be able to identify non-trivial universals and understand how speech might differ from the vocal communication systems of our closest non-human relatives. It is an idea that I have been trying to articulate (with varying degrees of failure) for more than three decades, starting with the book that grew out of my doctoral dissertation [1].

The idea has been difficult to articulate in large part because the alphabetic model that was foundational in the development of modern linguistics is so entrenched in our thinking about speech prosody. In laying out the idea, therefore, it is useful to look to a different modality, where the compositional structures that govern the alignment relationships among "gestures" (the basic units in the articulation of an utterance) are more transparently related to the alignment relationships among the objects of perception that are innate to the medium. That is, it is useful to incorporate observations and insights that have emerged in the last half century of research on the prosody of signed languages, such as American Sign Language [2, 3, 4] (ASL), Israeli Sign Language [5, 6] (ISL), and the newly emerging Al-Sayyid Bedouin Sign Language [7] (ABSL). I will review some of this work to make the first three of the following sequence of arguments:

(1) Every human language, signed as well as spoken, has a richly compositional syntactic system, with morphemic elements that can be combined and arranged to form potentially very complex and novel multi-layered sentences.

(2) Every human language also has a richly compositional prosodic system, with phonological elements that can be combined and arranged to form potentially very complex multi-layered utterances.

(3) Prosody is the "phonological patterning" counterpart to syntax in Hockett's specification of "duality of patterning" [8,9] (DoP) as the universal design feature that may differentiate human languages from the communication systems of other species. However, the universality of DoP is obscured if phoneme-sized segments are axiomatically assumed and described as the "building blocks" of morphemes, as in Martinet's "double articulation" [10], a theory of design features that is often (incorrectly) equated with DoP.

(4) The incorrect equation of Martinet's theory of "double articulation" with Hockett's theory of "duality of patterning" also obscures evidence of how DoP develops in ontogeny and of how DoP might have emerged in the phylogeny of our species.

The evidence that (1) is true of signed as well as spoken languages comes from studies of emerging languages such as ABSL and Nicaraguan Sign Language, where syntactically complex structures such as conditionals and also conventionalized agreement morphology have emerged within a single generation of early signers [11].

The evidence that (2) is true of signed as well as spoken languages also comes from emerging languages, and there is evidence of productive complexity for at least three levels of structure. First, manual gestures can be decomposed into specifications for the properties listed in (5), and a unit analogous to the syllable in spoken languages can be defined by coordinated changes in (at least one of) location, finger position, and orientation. Second, in all sign languages studied to date, content words are typically monosyllabic, but longer forms (e.g., disyllabic compound words) can be marked as prosodic words by phonological processes of reduction and/or spreading. Third, even the youngest sign languages have rich "visual intonation" systems by which non-manual gestures are aligned to the sequence of manual gestures so as to demarcate prosodic units comparable to the intonational phrases of spoken languages, as illustrated in Figs. 1 and 2.

(5) features in ASL	example of minimal pair or of sign with changing specification
handshape	[of dominant hand (H1) or of both hands (H1&H2) in some signs]
selected fingers	SCHOOL (all 5 fingers) vs. IM,POSSIBLE (thumb & pinky)
finger position	WHITE (all 5 fingers begin spread and then close to touch)
orientation	STAY (both palms up) vs NOW (both palms down)
location	ONION (at side of right eye) vs. APPLE (at right of chin)
endpoint	SCHOOL (H1 moves down from neutral space to touch H2)
movement type	NUDE (1 large movement) vs. AVAILABLE (2 short movements)

The evidence for (3) stems from a close reading of relevant parts of [8,9], and the realization that the argument against DoP in ABSL [7] is actually an argument instead against Martinet's theory. (There will be no time to review the evidence for (4), which is part of the larger idea that has been developed in more detail elsewhere [12,13].)

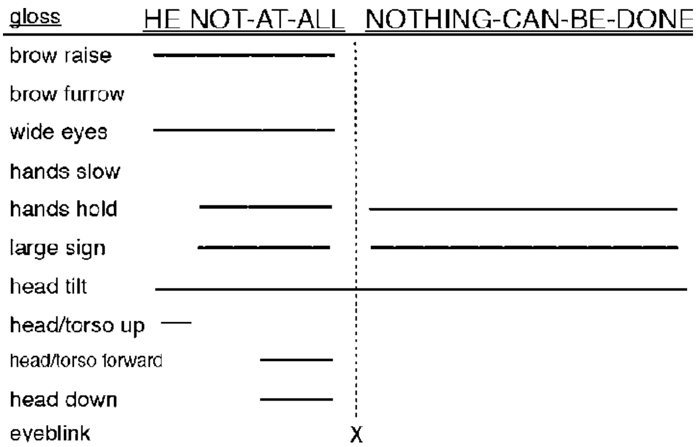


Fig.1 Display of the time-aligned tags for gestures of "visual intonation" of an ABSL utterance meaning 'If he says no, then nothing can be done.' (extract from Fig. 5 in [7]).

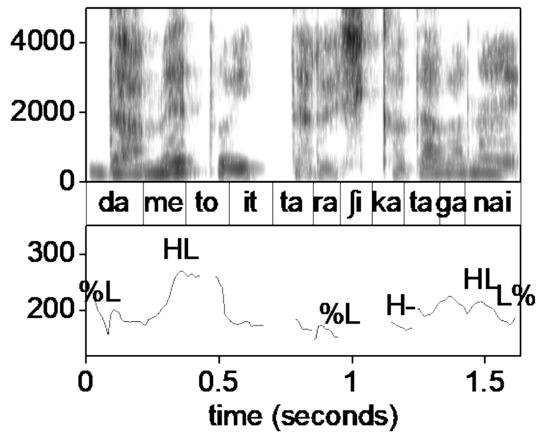


Fig.2 Display of a commonly used acoustic measure of "intonation" in an utterance of a Japanese translation of utterance in Fig. 1.

References

[1] Beckman, M. E. (1986). *Stress and non-stress accent*. Amsterdam : Foris Publications.

[2] Liddel, S. K., & Johnson, R. E. (2013). American Sign Language: The phonological base. *Sign Language Studies, 64*, 195-277.

[3] Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

[4] Wilbur, R. B. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Language and Speech, 52*(2/3), 245-285.

[5] Meir, I. (1998). Syntactic-semantic interaction in Israeli Sign Language verbs: The case of backwards verbs. *Sign Language and Linguistics, 1*(1), 3-37.

[6] Dachkovsky, S., & Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Language and Speech, 52*(2/3), 287-314.

[7] Sandler, W., Meir, I., Dachkovsky, S., Padden, C., & Aronoff, M. (2011). The emergence of complexity in prosody and syntax. *Lingua, 121*(13), 2014-2033.

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

[9] Hockett, C. F. (1961). Linguistic elements and their relations. *Language, 37*(1), 29-53.

[10] Martinet, A. (1957). Arbitraire linguistique et double articulation. *Cahiers Ferdinand de Saussure, 15*, 105-116.

[11] Senghas, A., Kita, S., & Özyuörek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science, 305*(5691), 1779-1782.

[12] Edwards, J., Beckman, M. E., & Munson, B. (2015). Cross-language differences in acquisition. In M. A. Redford (ed.), *Handbook of speech production*, pp. 530-554. Chichester, UK: Wiley-Blackwell.

[13] Plummer, A. R., & Beckman, M. E. (2015). Framing a socio-indexical basis for the emergence and cultural transmission of phonological systems. *Journal of Phonetics, 53*, 66-78.

Adaptation to Prosody in Anticipatory Structural Analysis during Visual Search

Sun-Ah Jun

UCLA (USA)

jun@humnet.ucla.edu

A growing number of studies have shown that prosody can guide syntactic interpretation and prosodic information is considered during very early processing stages (Ito & Speer [1]; Snedeker & Trueswell [2]; Schafer et al. [3]; Weber et al., [4]). It is reported that listeners process prosodic cues rapidly enough to *anticipate* a likely speaker-intended referent. Recent work on sentence processing has also addressed how processing expectations change with varying linguistic input, a process known as *linguistic adaptation* (Chang et al. [5]; Fine et al. [6]; Norris et al. [7]) and that comprehenders can weigh different kinds of information according to their reliability (Tanenhaus et al. [8]). These studies raise the possibility that listeners also modulate the degree to which they use prosodic cues in structural analysis based on the reliability of prosody they experience in the experiment. To address these questions, the current study investigated (i) the anticipatory effect of prosody in the processing of globally ambiguous sentences and (ii) whether listeners place less weight on prosodic information in structural decision when filler items have 'uninformative' prosody. [This work is in collaboration with Chie Nakamura and Jesse Harris.]

Experiments: Eye movements from native speakers of English (N = 32 per experiment) were recorded in a visual-world eye-tracking experiment setting. In the stimuli, a prosodic boundary was located either before or after the NP that precedes the ambiguous PP, encouraging 'Modifier' interpretation, (1a, *see the tiger that has the binoculars*), or 'Instrument' interpretation, (1b, *see the tiger by using the binoculars*). In addition to the location of prosodic boundary, we also manipulated the final word, which is either plausible as an instrumental object, *binoculars* in (1), or implausible as in (2), *popcorn*. The visual scene in Fig.1a was presented with the sentences in (1) and that in Fig.1b with the sentences in (2). If prosody guides listeners' structural analysis, a garden-path effect is expected with Instrument prosody only when the Modifier interpretation is semantically plausible as in (2b). In Experiment 1, filler items had informative prosody (3a). In Experiment 2, filler items had 'uninformative' prosody (3b), in which the boundary was inappropriately located in the middle of the final NP and was realized as a L-L% (cueing phrase-final). Both experiments consisted of 24 target and 48 filler items.

Results: We analyzed the looks made to each object in the picture from the onset of 'with' until the mean onset of the final word. In Experiment 1, the analysis showed that participants looked significantly more at the instrument object with Instrument prosody than with Modifier prosody ($p<.05$) (Fig.2, left). Crucially, the effects were observed before participants heard the final word, reflecting anticipatory use of prosody in assigning a structural analysis. Also, for the looks to the target object during the final word, participants looked at *popcorn* less often when hearing the noun with Instrument prosody than with Modifier prosody, but no such difference between two types of prosody was observed with *binoculars*, demonstrating that participants anticipated an instrument object (*binoculars*) when presented with Instrument prosody in Implausible-instrument (*popcorn*) condition and were led down to garden-path when they heard 'popcorn'. In Experiment 2, there was no main effect of PROSODY on the proportion of looks to the instrument object during the anticipatory time window (Fig2, right). Furthermore, in the proportion of looks to the target object during the final word time window, there was an interaction between PROSODY and TRIAL BLOCK (1st half, 2nd half) in Exp.2, with a bigger effect of prosody in the 1st Block than in the 2nd block. The difference between the two blocks was greater in Exp.2 than it was in Exp.1. This demonstrates that the more exposed to uninformative prosody, the less the participants used prosodic information. In sum, the current study showed an anticipatory use of prosody, which provides support for sentence processing models that assume immediate use of prosodic information. The results further suggest that listeners track how informative prosodic cues are, adjusting the extent to which they use prosody in making anticipatory judgments during structural analysis.

(1) Target items in the Plausible instrument condition

a. Modifier prosody: The boy $_{L\text{-}H}$% will see $_{L\text{-}H}$% the tiger with the binoculars.

b. Instrument prosody: The boy $_{L\text{-}H}$% will see the tiger $_{L\text{-}H}$% with the binoculars.

(2) Target items in the Implausible instrument condition

a. Modifier prosody: The boy $_{L\text{-}H}$% will see $_{L\text{-}H}$% the tiger with the popcorn.

b. Instrument prosody: The boy $_{L\text{-}H}$% will see the tiger $_{L\text{-}H}$% with the popcorn.

(3) Filler items

a. Informative prosody: The boy $_{L\text{-}H}$% will touch $_{L\text{-}H}$% the necktie and the razor.

b. Uninformative prosody: The boy $_{L\text{-}H}$% will touch the necktie and the $_{L\text{-}L}$% razor.

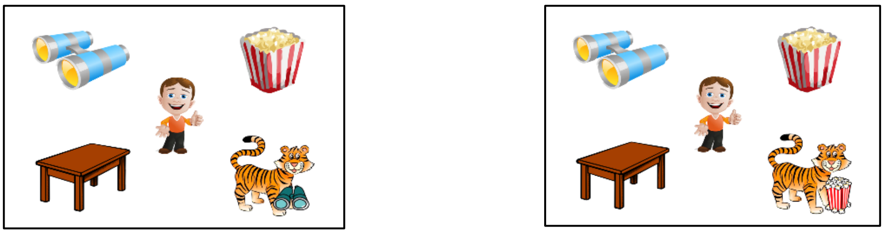


Fig.1 a. Visual scenes presented with (1a,b) b. Visual scene presented with (2a, b).

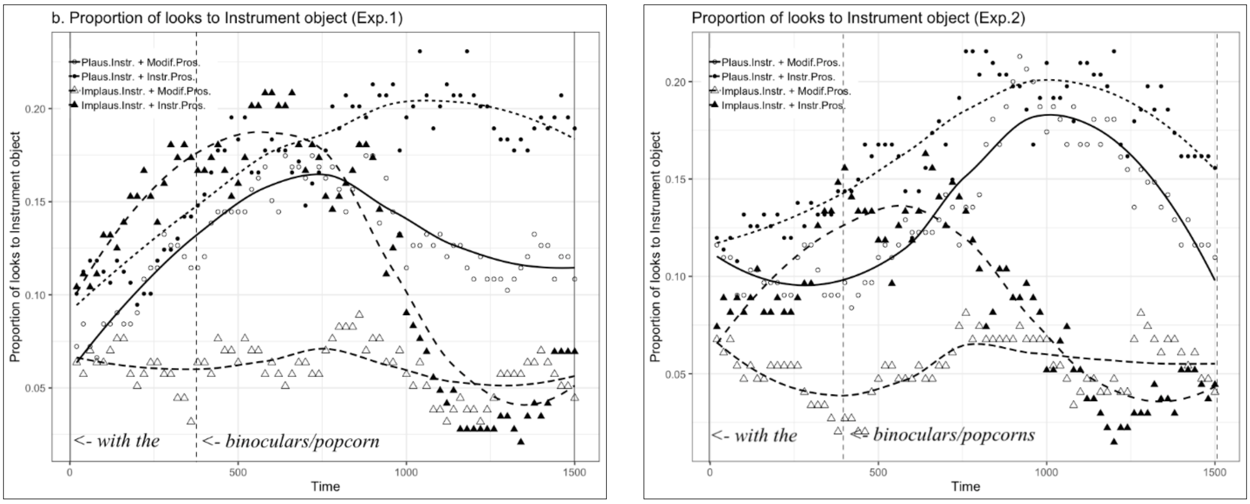


Fig.2 Proportion of looks to the Instrument object (e.g., binoculars) in Exp.1 (left) and in Exp.2 (right). Circles represent 'plausible' instrument nouns and triangles 'implausible' instrument nouns. Filled shapes are for Instrument prosody and empty shapes are for Modifier prosody.

References

[1] Ito, K. & Speer, S. (2008) Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory & Language, 58,541-573.*

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

[3] Schafer, A., Speer, S., Warren, P., & White, S. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169-182.

[4] Weber, A., Martine, G., Crocker, M. (2006) The role of prosody in the interpretation of structural ambiguity : A study of anticipatory eye movements. *Cognition*, 99, 63-72.

[5] Chang, F. Dell & Bock, K. (2006) Becoming syntactic. Psychological Review, 113, 234-272.

[6] Fine, A. B., Jaeger, T. F., Farmer, T. A., & Quan, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS ONE*, 8, e77661.

[7] Norris, D., McQueen, J. M. & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, 31(1), 4-18. doi:10.1080/23273798.2015.1081703.

[8] Tanenhaus, M. K., Kurumada, C., & Brown, M. (2015). Prosody and intention recognition. In L. Frazier and E. Gibson (Eds.), *Explicit and Implicit Prosody in Sentence Processing: Studies in Honor of Janet Dean Fodor*. pp.99-118. Springer International Publishing Switzerland.

Assessing the (un)categorisation of non-native phones to native phonological categories

Michael D. Tyler

Western Sydney University (Australia)
m.tyler@westernsydney.edu.au

Attunement to the native language shapes how consonants [1], vowels [2], and lexical tones [3, 4] are perceived. When acquiring their native language, children learn which phonetic differences signal a phonological contrast (*phonological distinctiveness*) and which differences constitute natural phonetic variability within a phonological category (*phonological constancy*) [5]. This results in rapid and efficient detection of native phonological contrasts, but it may inhibit accurate perception of non-native phonological contrasts. Natively tuned perception may help or hinder non-native speech perception depending on how the non-native phones/tones are assimilated to native phonological categories [1, 2, 6].

Pairs of non-native phones that are each assimilated to different native categories (a *two-category* assimilation) should be discriminated well because the listener detects a natively tuned phonological distinction [6]. If each non-native phone is assimilated to the same native category, with equal phonetic goodness of fit (a *single-category* assimilation), discrimination should be poor. In this case, the phonetic difference between the contrastive non-native phones may be one that is phonologically constant in the L1. That is, listeners may have learned to ignore a phonetic difference that is phonologically distinctive in the non-native language but phonologically constant in the native language. When non-native phones are assimilated to the same native category, but there is a difference in phonetic goodness of fit to the native phonological category (a *category-goodness* assimilation), the listener may benefit from this sensitivity and discriminate the contrast moderately well.

When a non-native phone is not assimilated as any single native phonological category it is uncategorised [6]. An uncategorised non-native phone may be weakly consistent with a single native category (*focalised*), weakly consistent with multiple native categories (*clustered*), or inconsistent with any native category (*dispersed*) [7]. The extent to which natively tuned perception facilitates discrimination of focalised and clustered non-native phones depends on the perceived phonological overlap [8]. For instance, if one clustered non-native phone is perceived as weakly consistent with a set of native phonological categories, and another non-native phone is weakly consistent with a different set of non-native phonological categories, then natively tuned perception should support discrimination. Conversely, if both phones are perceived as weakly consistent with the same set of native phonological categories then discrimination should be poor. When both non-native phones are perceived as dispersed, then the listener should be free of any influence from natively tuned perception, and discrimination should depend only on language-independent phonetic distance.

While most research on cross-language speech perception has focused on consonants and vowels, recent research has begun to consider whether the same principles apply to cross-language perception of lexical tone. Native tone-language listeners appear to perceive non-native tones in terms of their native tonal categories [9]. There is evidence of an influence of natively tuned perception on discrimination performance, but Reid et al. [9] argued that phonetic or acoustic similarity may need to be taken into account to explain the variability in performance. To some extent, non-tone-language listeners are able to categorise non-native lexical tones in terms of their native intonational categories (e.g., English *uncertainty*, *question [yes/no]*) [3, 4], but there is a generally poorer correspondence between categorisation and discrimination for intonational categories than for consonants, vowels, and lexical tones [3, 4, 9].

Perceptual assimilation is usually assessed using a forced-choice categorisation task, where listeners are presented with an auditory stimulus from the non-native language, and then assign it to the closest native phonological category using orthographical symbols and/or keywords [e.g., 2, 7, 8]. To determine whether a contrast is a category-goodness or single-category assimilation, it is necessary to ask listeners to rate the goodness of fit of the auditory stimulus to the chosen category label. Categorisation tasks give an excellent indication of how non-native phones are assimilated to the native language when there is high agreement within and between listeners. That is, the task works well when nearly all of the listeners categorise the non-native phone as the same native phonological category. With investigations expanding into stimuli that are likely to be uncategorised (e.g., vowels and lexical tones), and with category labels that go beyond consonants and vowels, it may be time for a fresh approach. I will argue that the forced-choice categorisation task is poorly suited to assessing cases where a non-native phone is uncategorised. For instance, if natively tuned perception can influence discrimination when a non-native phone is perceived as weakly consistent with multiple native phonological categories [8], then asking participants to select a single category label may not provide a true indication of perceptual assimilation. Furthermore, the use of arbitrary thresholds to determine categorisation is problematic [e.g., 8, 10], and it is not clear how to interpret category goodness ratings when a non-native phone is only weakly categorised. In this talk, I will take stock of the theoretical requirements that a test of perceptual assimilation must meet and sketch a possible best practice for determining the influence of the native language on the perception of non-native consonants, vowels, and lexical tones.

References

[1] Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109*(2), 775-794.

[2] Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*(1), 4-21.

[3] So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and speech, 53*(2), 273-293.

[4] So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition, 36*(2), 195-221.

[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*(5), 539-542.

[6] Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.

[7] Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America, 139*(1), EL1-EL5.

[8] Faris, M. M., Best, C. T., & Tyler, M. D. (2018). Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics, 70*, 1-19. doi:10.1016/j.wocn.2018.05.003

[9] Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Xu Rattanasone, N., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics, 77*(2), 571-591.

[10] Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America, 110*(1), 489-503.

Common coordination patterns in prosodic and segmental domains

Jason A. Shaw

Yale University
Jason.Shaw@yale.edu

In Articulatory Phonology, phonological representations take the form of temporally coordinated action units, known as gestures [1, 2]. Frequently observed coordination relations include synchronous (or in-phase) timing, whereby gestures start at the same time, and sequential (or anti-phase) timing, whereby the onset of one gesture is coordinated to the offset of another gesture [3]. In contrast to these coordination relations observed in the segmental domain, the pitch accents of intonation phonology have been argued to have target-based timing [4, 5]. On this view, pitch peaks (or troughs) of accentual tones, rather than the start or end of pitch control, are aligned to segments. Accordingly, pitch contours are derived from interpolation between L(ow) and H(igh) tone specifications, which are aligned to segments according to language-specific rules. It appears from this past work that segmental timing differs from prosodic timing in that segmental gestures (consonants and vowels) are coordinated with reference to gesture onsets and offsets while pitch accents (tones) are coordinated with respect to their targets (F0 maxima/minima). There has been some work trying to reconcile these differences, focusing on pitch patterns that can be modelled as tone gestures aligned according onsets (instead of pitch targets) [6]. In this talk, I argue based on data from Mandarin Chinese that consonant and vowel gestures can also show target alignment, offering another route to convergence between timing control in prosodic and segmental domains.

The key evidence comes from analysis of Electromagnetic Articulography (EMA) data from six speakers (three male) of Mandarin Chinese. Data was collected using the NDI Wave Speech production system. Sensors were attached to the tongue tip (TT), blade (TB), dorsum (TD), lips, jaw, nasion and mastoids. Lip Aperture (LA) was computed as the difference between the upper and lower lip sensors. Target items were a set of CV monosyllables that crossed all four lexical tones with two labial consonants {/m/,/p/} and three back rounded vowels {/o/,/u/,/ou/} yielding 24 items, which were repeated 6-12 times by each speaker producing a corpus of 949 tokens for analysis. Items were randomized with fillers and displayed one at a time on a monitor in pinyin.

To investigate the alignment of the vowel gesture relative to the consonant, a series of gesture landmarks were parsed from continuous kinematic trajectories: (1) **onset** of movement, (2) achievement of **target** (3) the **release** from target and (4) the **offset** of controlled movement. The alignment of consonants and vowels was assessed by computing C-V lag measures, subtracting consonantal landmarks from vowel landmarks. Fig. 1 compares three such lag measures. On average, the achievement of target of the vowel occurs around the offset of the consonant (right).

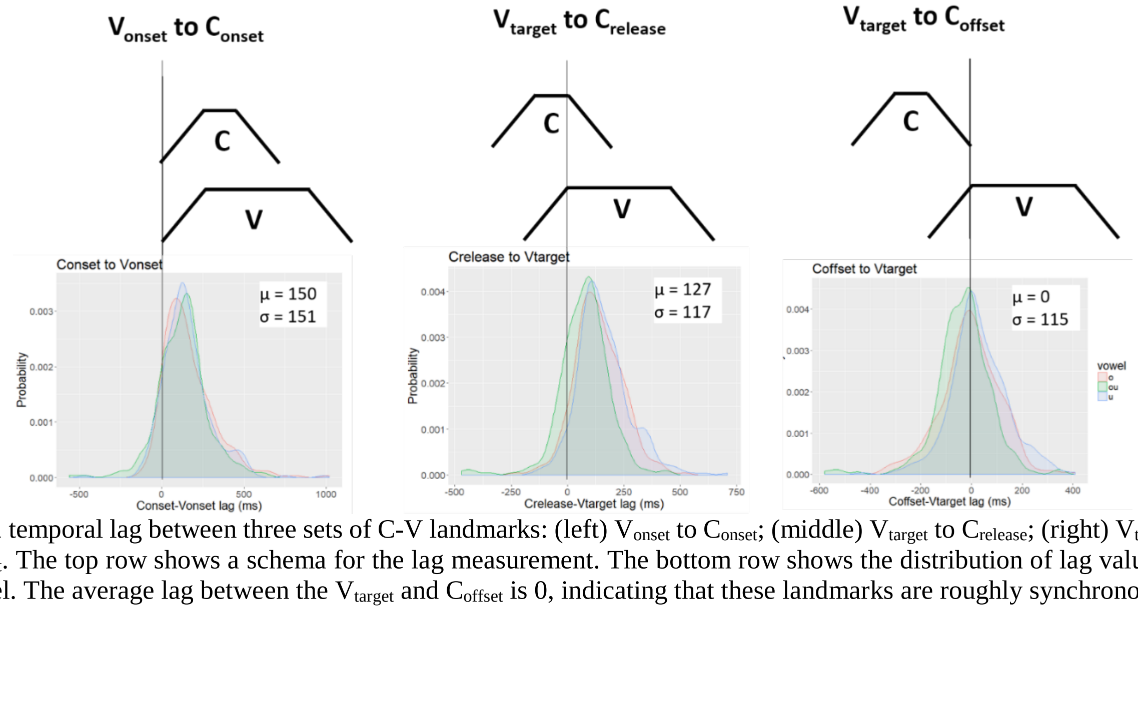


Fig.1 temporal lag between three sets of C-V landmarks: (left) V_{onset} to C_{onset}; (middle) V_{target} to $C_{release}$; (right) V_{target} to C_{offset}. The top row shows a schema for the lag measurement. The bottom row shows the distribution of lag values by vowel. The average lag between the V_{target} and C_{offset} is 0, indicating that these landmarks are roughly synchronous.

In contrast to the in-phase C-V pattern sometimes assumed for Mandarin Chinese [7], we found that the vowel gesture typically begins well after the consonant gesture; rather, it appears that the *vowel target* is more stably timed to the consonant than the *vowel onset*. Target-based timing makes an additional prediction that can be tested in the data. If the vowel target is the landmark that is temporally coordinated it would mean that the timing of the gesture onset is free to vary. Movement towards the target could start earlier or later in time depending, for example, on the distance of the tongue dorsum to its target. To investigate this possibility, we evaluated whether the lag between the consonant onset and vowel onset (leftmost schema in Fig. 1) is correlated with the spatial position of the tongue dorsum at movement onset. Fig 2 shows this relation. There is a significant negative correlation. The further the tongue dorsum is from the target, the earlier in time the vowel begins its movement. This pattern of spatially conditioned relative timing is not expected if gestures are timed according to their onsets and offsets [8]; however, the pattern is analagous to the finding in the intontation literature that the slope of F0 between pitch accents decreases with the number of syllables [9] or with speech rate [4].

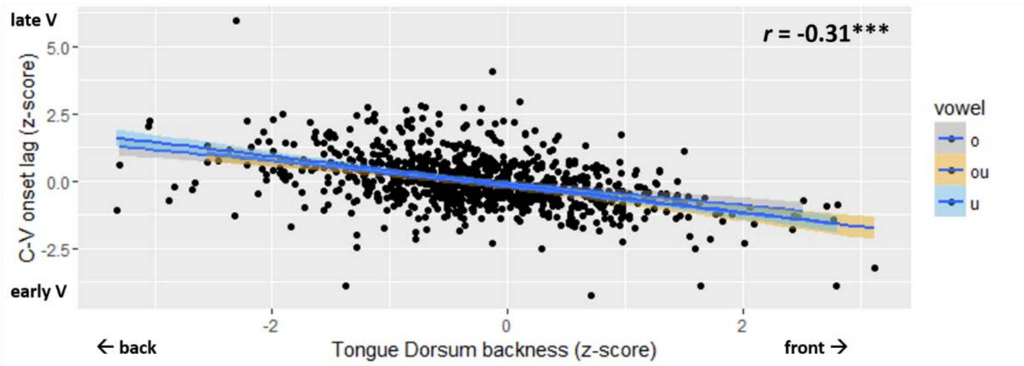


Fig.2 scatter plot of C-V lag (y-axis) and Tongue Dorsum backness (x-axis). The legend shows the Pinyin for the vowels, which correspond to: ‘o’ /uo/, ‘u’ /u/, ‘ou’ /ou/

Overall, the data suggest that C-V coordination in Mandarin is based on the alignment of the vowel target to the offset of the consonant. This pattern is in contrast to other languages which can be modelled with gestures timed either in-phase or anti-phase [3]. Target-based vowel alignment resembles tone alignment in the Autosegmental Metrical framework, offering a point of convergence between timing control in the segmental and prosodic domains.

References

[1] Gafos, A. & Goldstein, L. (2012). Articulatory representation and organization. In A. C. Cohn, C. Fougeron, & M. K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*. (pp. 220-231).
[2] Browman, C., & Goldstein, L. (1986). Towards an Articulatory Phonology. *Phonology Yearbook, 3*, 219-252.
[3] Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: a coupled oscillator mode. In F. Pellegrino, E. Marisco, & I. Chitoran (eds.), *Approaches to phonological complexiy*. (pp. 299-328). Berlin, New York: Mouton de Gruyter.
[4] Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant “segmental anchoring” of F0 movements under changes in speech rate. *The Journal of the Acoustical Society of America, 106*(3), 1543-1554.
[5] Arvaniti, A. Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of phonetics, 26*(1), 3-25.
[6] Katsika, A. Krivokapić, J. Mooshammer, C. Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics, 44*, 62-82.
[7] Xu, Y., & Wang, M. (2009). Organizing syllables into groups—Evidence from F< sub> 0</sub> and duration patterns in Mandarin. *Journal of phonetics, 37*(4), 502-520.
[8] Shaw, J. A., & Chen, W.-r. (submitted). Spatially-conditioned speech timing: evidence and implications. *Frontiers in Psychology: language science*, 1-17.
[9] Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology, 3*, 255-309.

Prosodic position effects: strengthening for withstanding?

Cécile Fougeron, Fanny Guitard-Ivent, Daria D'Alessandro

Laboratoire de Phonétique et Phonologie, CNRS/Université Sorbonne-Nouvelle (France)
{cecile.fougeron ; fanny.ivent ; daria.dalessandro}@sorbonne-nouvelle.fr

Prosodic position effects (PP effects) include variations in the phonetic make-up of segments according to their position relative to prosodic prominences and boundaries. Past results have suggested that prosodic position effects target the contrastive phonetic properties of segments under prominence or close to strong prosodic boundaries, with the idea of a strengthening of the segment properties as a way to enhance paradigmatic and/or syntagmatic contrasts in strong prosodic positions (see [1,2] for a review). Further results suggest that rather than an additional reinforcement of contrastive properties, the observed effects are the sign of a resistance against reduction in strong prosodic position. For instance, less tongue twister errors have been found in pitch accented or phrase initial words ([3]) and less consonantal distortions are made after strong prosodic boundaries in some dysarthrias ([4]). Reduced overlap has also been described in prominent position and across strong prosodic boundaries in different structural configurations with targets and triggers straddling a boundary (#) (e.g. V#C [5]; C#V [6]; V#(C)V [7]; C#C [8]), while controversial results have been reported for coarticulation between adjacent segments at domain edges. Reduced overlap at the beginning of strong prosodic constituent has been found by [9] for #CC sequences, and by [10] for #NV sequence, while no effects are reported in [11] for #CC or in [6] and [8] for #CV. Reduction of coarticulation between segments in strong prosodic position can be interpreted as a way to reinforce syntagmatic contrasts between less overlapping segments, but also paradigmatic contrasts if reduced coarticulation make the segment more distinctive.

In this presentation, we will present recent results further supporting that prosodic strengthening can be interpreted as a way to withstand phonetic reduction in strong prosodic positions. Arguments are based on different studies on French vowels, using various methodologies, looking at the way PP effects modulate (a) coarticulation in post-boundary sequences, (b) duration dependent phonetic reduction, (c) phonetic variability across repetitions, and (d) acoustic discriminability of French vowels.

Study1 looks at variation in overlap in a post-boundary sequence ([12,13]). Anticipatory and carryover C-to-V coarticulatory effects are tested, and the structural relationship between the target and the trigger is manipulated (heterosyllabic V_1.C sequence vs. tautosyllabic V_1C and CV_1 sequences). Contextual effects of C, either alveolar (C_{ALV}=/t, d, z, s, l, n/) or uvular (C_{UV}=/R/), on the acoustic of V_1 =/a/ vowel are examined according to the prosodic position of the target vowel: in an Intonational Phrase initial position (IPi) or in a word-medial position (Wm). For IPi anticipatory coarticulation, the target /a/ vowel is sitting at the edge of the IP constituent (##V_1C and ##V_1.C), while for carryover coarticulation (##CV_1), the target is in the initial syllable of the domain (##CV_1). Speech material is extracted from two large corpora of natural French allowing for the study of 7000 tokens of /a/. Coarticulation is measured as F1 and F2 changes according to the context (C_{ALV} vs. C_{UV}). In all types of sequences, expected contextual effects (e.g. lower F1 in C_{ALV} vs. C_{UV} context) and prosodic position effects (higher F1 in IPi vs. Wm) are found. Crucially, this last effect appears to be stronger for V_1C and V_1.C, where /a/ is clearly more open in IPi, and weaker when the target /a/ is not in absolute initial position (i.e. CV_1). More interestingly, this study reveals some interactions between prosodic position and contextual effect showing *less coarticulation in IPi position for V_1C and V_1.C sequences* (but not for CV_1), with a greater reduction of coarticulation for V_1.C sequences. These results suggest that vowels immediately following an IP boundary withstand overlap with following consonants. Furthermore, the modulation of coarticulation by prosodic position informs us on the time window (and/or encoding unit) over which prosody adjusts the coordination between segments: this tuning does operate on sequences with less tightly specified coupling patterns such as heterosyllabic V.C, nucleus+coda VC sequence, but not on a tightly coordinated tautosyllabic CV sequences.

Study 2 looks at duration dependent spectral reduction according to prosodic position. Formant frequency of /a/ vowels in a controlled consonantal context (/p_p/ and /p_s/) are examined according to the duration of the vowel in two domain final positions: in Intonational Phrase final position (IPf) vs. in a word final but IP medial position (Wf); and in two domain initial positions: in Intonational Phrase initial position (IPi) vs. in a word initial but IP medial position (Wi). In IPf position, the vowel is at the domain edge (CV#), while in initial position, the vowel is the second element of an initial syllable (#CV). Read productions of the test sentences are collected for four speakers, for a total of 180 tokens per position. Expected duration-dependent reduction of F1 is found overall, with shorter vowels being less open. Crucially, *the relationship* *between F1 height and duration is found to depend on prosodic position*. While the F1 height of IP medial vowels, in both word-final and word-initial positions, is linked to its duration (r=.63 and r=.5, respectively), this relationship does not hold for vowels close to IP boundaries. Indeed, IP final (CV#) vowels are globally lengthened and have a higher F1 than Wf. Their distribution in both spectral and temporal dimensions overlap that of Wf vowels, but contrary to Wf vowels the two dimensions are not related (r=.17). Vowels in the initial syllable of IP (IPi #CV) are not lengthened compared to Wi position and their F1 is not correlated with their duration (r=.16). This results show that the spatio-temporal tradeoff responsible for vowel reduction in speech is affected by prosodic position. Vowels close to domain edges seem to be protected against duration-dependent phonetic reduction and show stable spectro-temporal specifications.

Study 3 follows-up on the idea that the resistance of prosodically strengthened segments may translate into more stable phonetic targets, and tests for this stability across multiple repetitions. Token-to-token variability of the vowel /a/ in the four prosodic positions described in study 2 above (IPf, IPi, Wf, Wi) is tested in the productions of four speakers so-far. Forty-five repetitions of the test sentences have been collected over 5 recording sessions over two weeks, with 9 repetitions of each condition per session. Variability is measured with a pair-wise variability index (PVI) computing the average differences in F1 between pairs of successive repetitions divided by the mean frequency of the pairs. For initial positions, PVI values vary according to speakers (p=.001) but no effect of prosodic position (IPi vs. Wi) nor interaction between speakers and position is found. For final positions, on the other hand, the speaker dependent PVI values (p=.01) interact with prosodic position (p=.005): *for two of the four speakers, IP final vowels show less token-to-token variability than W final ones*. More speakers remain to be analyzed to confirm whether vowels targets are indeed more stable in IP final positions.

Last arguments supporting the view that PP effect contribute to withstand phonetic reduction comes from a 4th study looking at the acoustic discriminability among vowels within the French oral vowels system ([14]). Discriminability is assessed on the base of classification results of two types of classifiers: a linear discriminant analysis (LDA) based on the four formants frequencies, and a deep convolutional neural network (CNN) based on spectrogram pictures. Classifiers were trained using a set of 4500 vowels extracted from a large read speech corpus and the test set includes 720 exemplars of /i, y, e, ɛ, a, x, u, o, ɔ/ (with /x/=/ø, œ/) produced either in intonational phrase initial (IPi) or word initial (Wi) position. Results show that PP effects translate into a *better discriminability of vowels* (overall better classification rate) in IPi than in Wi with the two methods. More crucially, among the dimensions showing a better discriminability in IPi, *less confusions are found between peripheral and central vowels*. These results suggest that PP effects do protect vowels against centralization and phonetic contrast reduction in IPi position.

Taken together these results support the view that information about prosodic phrasing is available at the moment in the planning process when low-level adjustments of phonetic targets are implemented.

References

[1] Cho, T. (2011). Laboratory phonology. In *The Continuum Companion to Phonology* (pp. 343-368).

[2] Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass, 10*(3), 120-141.

[3] Croot, K, Au, C., & Harper, A. (2010). Prosodic structure and tongue twister errors. In *Papers in Laboratory Phonology* 10.

[4] Antolík, T., & Fougeron, C. (2013). Consonant distortions in dysarthria due to Parkinson's disease, Amyotrophic Lateral Sclerosis and Cerebellar Ataxia. *Interspeech*.

[5] Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Acoustic results. JASA, 113(1), 516-531.

[6] Cho, T., Yoon, Y., & Kim, S. (2014). Effects of prosodic boundary and syllable structure on the temporal realization of CV gestures. *J. of Phonetics, 44*, 96-109.

[7] Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *J. Of Phonentic., 32*(2), 141-176.

[8] Meynadier, Y. (2003). Interaction entre prosodie et (co)articulation linguopalatale en français. PhD diss. U Aix-Marseille.

[9] Bombien, L., Mooshammer, C., Hoole, P., Rathcke, T., & Kühnert, B. (2017). Articulatory strengthening in initial German /kl/ clusters under prosodic variation. *ICPhS*, 457-460.

[10] Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned finetuning of coarticulatory vowel nasalization in English. *J. of Phonetics, 64*.

[11] Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics—The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Laboratory Phonology, 10*, 31-59.

[12] Guitard-Ivent, F. (2018). Coarticulation C-à-V en français: interaction avec le type de voyelle, la position prosodique et le style de parole. PhD diss, U. Sorbonne-Nouvelle.

[13] Guitard-Ivent, F. & Fougeron, C. (2017). Domain-Initial strengthening as reduced coarticulation, *PaPE*, Cologne.

[14] Guitard-Ivent, F., Chignoli, Fougeron C., (submitted). Are IP initial vowels acoustically more distinct? Results from LDA and CNN classifications. *Interspeech*.

Brain Signatures of Prosodic Processing

Karsten Steinhauer

McGill University (Montreal, Canada)
karsten.steinhauer@mcgill.ca

The virtually unlimited temporal resolution of event-related brain potentials (**ERPs**) can help advance our understanding of prosodic processing in real time and its critical role in online sentence comprehension.

Prosody often guides the syntactic analysis of sentences. This has been shown in structurally ambiguous 'garden path' sentences such as Early/Late Closure ambiguities (examples from **[1]**):

1a) Late Closure: *When a bear is approaching <u>the people</u> the dogs come running.*

1b) Early Closure: *When a bear is approaching <u>the people</u> come running.*

In reading studies, (1b) has been found to be more difficult to process, as the syntactic parser tends to initially attach the NP *'the people'* to the preceding verb (as required in 1a). However, in speech a prosodic boundary before this NP can prevent these difficulties while rendering sentence (1a) more difficult to process. Behavioral studies of these structures had to rely on quite unnatural tasks until the mid 1990s, producing inconsistent results. ERPs were found to better reflect both the processing of prosodic boundaries as well as their immediate impact on sentence interpretation in real-time – *independent* of task requirements **[2]**. Most importantly, this work identified a novel positive-going ERP waveform labeled *closure positive shift* (**CPS**) that was found to reliably mark the online processing of prosodic boundaries in listeners (**Figure 1**). Since its discovery, the CPS has been replicated in more than ten languages (e.g., Chinese, Dutch, English, French, German, Japanese, Korean and Swedish) and has been used to examine a number of issues related to prosodic sentence processing. For example, developmental research has shown that a CPS is reliably elicited only around age 3 years, potentially suggesting sufficient syntactic knowledge has to be acquired first **[3-4]**. On the other hand, older people (> 75 years) display a CPS similar to that in young adults, but they integrate prosodic, lexical and syntactic information in different ways **[5]**.

De-lexicalized speech and musical phrasing. Some studies observed the CPS in de-lexicalized (e.g., low-pass filtered or 'hummed') speech, indicating that this ERP component is in fact triggered by prosodic phrasing and not by lexical or syntactic processes **[6-7]**. This finding also raises issues about the domain-specificity of the cognitive processes underlying the CPS. In 2005, a study reported a 'music-CPS' at phrase boundaries in melodies **[8]**, but both its timing and the apparent finding that only trained musicians elicited it cast doubt on its equivalence with the prosodic CPS in speech. More recently, Glushko and colleagues demonstrated another CPS in music that strongly resembles the one at speech boundaries **[9]**, whereas the 2005 effect may reflect a different cognitive process. This new finding strongly suggests that phrase boundary processing in speech and music may – at least to some extent – rely on similar neurocognitive mechanisms.

Implicit/covert prosody in silent reading and punctuation. Prosodic representations also seem to be activated during silent reading **[10]**, thus suggesting another application for the CPS in an otherwise extremely difficult-to-study domain. To date, a number of ERP studies in a range of languages have shown that prosodic phrasing in silent reading elicits a CPS component similar to that in speech processing. The CPS for subvocal prosodic phrasing was observed irrespective of whether the boundary was induced by commas **[6, 11-12]**, by mapping a previously heard prosodic contour on the written text **[13]**, or by manipulating the length of noun phrases – as shown in a reading study in Korean **[14]**. However, the reliability of punctuation cues in inducing prosodic phrasing partly depends on the reliability with which readers use punctuation rules **[6, 15]**.

Categorical boundary perception? Annotation systems such as ToBI often assume categorical boundary perception (e.g., an intermediate versus an intonational phrase boundary), while some researchers have argued for a continuum and for relative boundary strength perception. A CPS study from our lab manipulated the strengths of competing boundaries in sentences with early versus late closure ambiguities and found behavioral and ERP evidence for a continuum **[16]** (see **Figure 2**).

Prosodic processing in second language learners. Prosody is usually not part of the curriculum in second language (L2) instruction, and relatively little is known about L2 learners' use of prosodic information in sentence processing. Recent ERP work suggests that both L1 background and language proficiency shape the integration of prosodic and syntactic cues, and that, importantly, even English native speakers' ERPs are influenced by their English proficiency level [17].

After a general introduction to CPS research, my talk will focus on (a) implicit prosody in Korean garden-path sentences, (b) categorical versus continuous boundary perception, and (c) prosody in L2 learners.

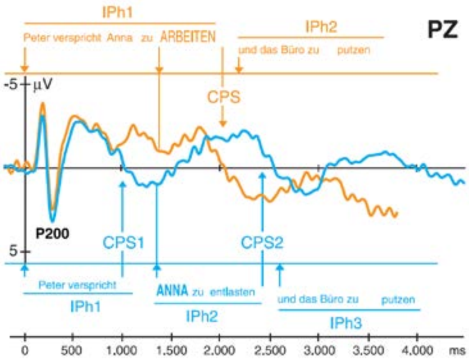


Fig.1 CPS components in ERPs, time-locked to sentence onset (Steinhauer et al., 1999)

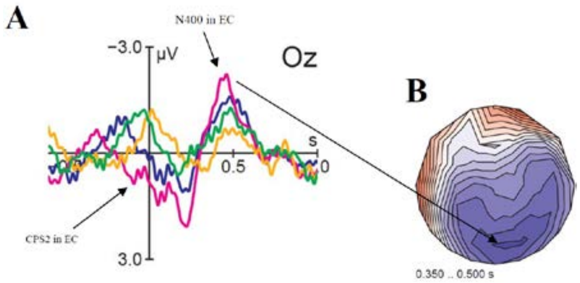


Fig.2 Effects of an incompatible late boundary of increasing strength (yellow=1, pink=4) in Early Closure sentences. The stronger the boundary, the larger the CPS, and the larger the N400 reflecting processing difficulties.

References

[1] Pauker, E., Itzhak, I., Baum, S.R., Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *Journal of Cognitive Neuroscience, 23 (10)*, 2731-2751.

[2] Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech. *Nature Neuroscience, 2*, 191-196.

[3] Männel, C., & Friederici, A.D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *Journal of Cognitive Neuroscience, 21*, 1988-2006.

[4] Männel, C., & Friederici, A.D. (2011). Intonational phrase structure processing at different stages of syntax acquisition: ERP studies in 2-, 3-, and 6-year-old children. *Developmental Science, 14,* 786-798.

[5] Steinhauer, K., Abada, S.H., Pauker, E., Itzhak, I. & Baum, S.R. (2010). Prosody-syntax interactions in aging: Event-related potentials reveal dissociations between on-line and off-line measures. *Neuroscience Letters, 472,* 133-138.

[6] Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker of prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*, 267-295.

[7] Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven processing: An event-related potential study. *Journal of Cognitive Neuroscience, 17*, 407-421.

[8] Knoesche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., et al. (2005). Perception of phrase structure in music. *Human Brain Mapping, 24*, 259–273.

[9] Glushko, A., Steinhauer, K., DePriest, J., & Koelsch, S. (2016). Neurophysiological correlates of musical and prosodic phrasing: shared processing mechanisms and effects of musical expertise. *PloS one*, *11*(5), e0155300.

[10] Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27,* 285-319.

[11] Liu, B.L., Wang, Z.N., & Jin, Z. (2010). The effects of punctuations in Chinese sentence comprehension: An ERP study. *Journal of Neurolinguistics, 23*, 66-80.

[12] Drury, J. E., Baum, S. R., Valeriote, H., & Steinhauer, K. (2016). Punctuation and implicit prosody in silent reading: an ERP study investigating English garden-path sentences. *Frontiers in Psychology*, *7*, 1375.

[13] Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain & Lang, 86*, 142-164.

[14] Hwang, H. & Steinhauer, K. (2011). Phrase length matters: The interplay between implicit prosody and syntax in Korean 'garden path' sentences. *Journal of Cognitive Neuroscience, 23 (11)*, 3555-3575.

[15] Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D.J. (2008). Sentence processing in the visual and auditory modality: Do comma and prosodic break have parallel functions? *Brain Research, 1224*, 102-118.

[16] Steinhauer, K. (2011). Combining Behavioral Measures and Brain Potentials to Study Categorical Prosodic Boundary Perception and Relative Boundary Strength. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII), Hongkong (China),* p. 1898-1901.

[17] Nickels, S., & Steinhauer, K. (2018). Prosody–syntax integration in a second language: Contrasting event-related potentials from German and Chinese learners of English using linear mixed effect models. *Second Language Research*, *34*(1), 9-37.

Language specific representations in word stress perception: ERP evidence

Ferenc Honbolygó

Brain Imaging Centre, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest (Hungary)
honbolygo.ferenc@ttk.mta.hu

The speech signal consists of two types of acoustical information: segmental information, being the property of speech sounds, and suprasegmental or prosodic information, being the property of several successive speech sounds. During speech perception, both of these are extracted from the speech input simultaneously and are encoded in separate memory representations. Thus, when listeners recognize speech, they are processing a prosodically determined variant [1]. We can identify several different types of prosodic information, the most important ones being length, rhythm, intonation and stress [2]. In the present talk, I focus on the processing of stress.

Stress, similarly to other prosodic features, is a multidimensional information, comprising several different acoustic changes, such as intensity and fundamental frequency (f0). Stress is a relative emphasis given to certain syllables within words or to certain words in sentences [3]. Word stress plays either a culminative or demarcative role, that is emphasizing or separating certain parts of the speech stream, thus potentially contributing to the segmentation of continuous speech into words [4]. Languages differ considerably in the use of word stress: in the position of the stressed syllable within multisyllabic words (initial, final, penultimate, etc.); in the variability of the stressed syllable's position (free or fixed); and whether stress can distinguish lexical meaning (contrastive or non-contrastive). Fixed-stress languages (like Hungarian) mandatorily assign syllable stress to a specific position within a word, and stress is non-contrastive. Therefore, it can be assumed that stress processing demonstrates language specificity.

In this talk, I will present event-related brain potential (ERP) results related to the processing of word stress patterns from three different studies, using meaningless pseudowords [5], meaningful words [6] and pseudowords spoken by foreign speakers. The basic premise of the studies was that words stress is processed in relation to long-term representation, which are pre-lexical and language specific. Consequently, we can assume that both words and pseudowords are processed similarly, and foreign words are processed differently from native words. In order to study these assumptions, we applied the passive oddball paradigm to elicit the Mismatch Negativity (MMN) ERP component, a fronto-centrally negative waveform appearing to the pre-attentive detection of violation of simple or complex regularities [7].

In the experiments, Hungarian participants heard disyllabic words (bɑˈbɑ, meaning 'baby'), pseudowords (bɛˈbɛ) and pseudowords pronounced either by a Hungarian or a German speaker (beːˈbeː). Stress could be either on the first (legal stress) or on the second (illegal stress) syllable. The experimental design in all three experiments was similar. We used two conditions: in the first, stimuli with the legal stress were standards and stimuli with illegal stress were deviants; in the second condition, the standards and the deviants were reversed. This allowed us to calculate the MMN component by subtracting ERPs to the standard from the ERPs to the deviant using physically identical stimuli.

Results showed that the pseudoword deviant having an illegal stress pattern elicited two consecutive ERP components that were considered as MMN, whereas the deviant having a legal stress pattern did not elicit MMN. Moreover, pseudowords with a legal stress pattern elicited very similar ERP responses irrespective of their role in the oddball sequence, i.e., if they were standards or deviants, demonstrating that their processing relied on long-term instead of short-term (stimulus sequence related) memory traces (see Fig.1.). Meaningful words elicited similar ERP deflections, but the lexical status slightly modulated the processing of stress patterns. This modulation was different for the legal and illegal stress patterns (see Fig.2.). Finally, the comparison of pseudowords with native and non-native pronunciation showed that all pseudowords in the deviant position elicited an Early Differentiating Negativity (EDN) and a Mismatch Negativity (MMN) component, except for the Hungarian pseudowords stressed on the first syllable. This suggests that

Hungarian listeners did not process the native legal stress pattern as deviant, but the same stress pattern with a non-native accent was processed as deviant.

In conclusion, our data show that word stress pattern change is processed preattentively by the human brain, and this is slightly modulated by the lexical status of the word. Words and pseudowords with legal stress pattern do not initiate an error detection mechanism, as indicated by the lack of MMN component, but only if they are native sounding. We suggest that this is an evidence that word stress is processed based on prelexical language specific stress representations.

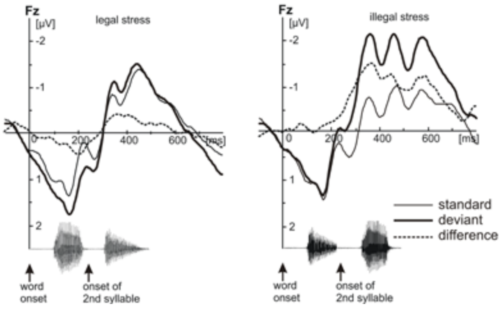


Fig.1 ERPs elicited by the pseudoword with legal and illegal stress pattern in the standard and deviant positions. The figures depict ERPs to the same stimulus in two different positions, and the difference wave obtained by subtracting the ERPs to the same pseudoword in the deviant and standard positions.

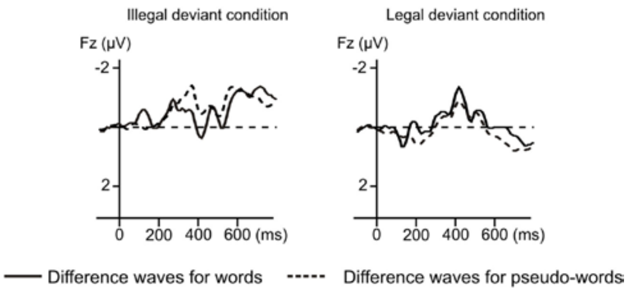


Fig.2 Difference wave ERPs for words and pseudowords.

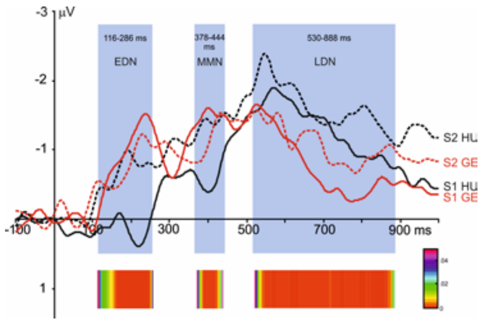


Fig.3 Difference wave ERPs cross-linguistic study. The results of the TANOVA analysis are depicted below the ERPs. Colored intervals show the significant Language x Stress position interaction. S1: stimuli with stress on the first syllable; S2: stimuli with stress on the second syllable; HU: Hungarian stimuli; GE: German stimuli. EDN: Early Differentiating Negativity; LDN: Late Discriminative Negativity.

References

[1] Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Lang. Speech, 40*(2), 141–201.

[2] Fox, A. (2000). *Prosodic features and prosodic structure: the phonology of suprasegmentals*. Oxford University Press, USA.

[3] Kager, R. (2007). Feet and metrical stress. In P. de Lacy (eds.), *The Cambridge Handbook of Phonology* (pp. 195–228). Cambridge University Press, Cambridge.

[4] Cutler, A., & Norris, D. (1988). The Role of Strong Syllables in Segmentation for Lexical Access. *J. Exp. Psychol. Hum. Percept. Perform., 14*(1), 113–121.

[5] Honbolygó F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *Int. J. Psychophysiol., 87*(2), 165–172.

[6] Garami, L., Ragó, A., Honbolygó, F., & Csépe, V. (2017). Lexical influence on stress processing in a fixed-stress language. *Int. J. Psychophysiol., 117*, 10–16.

[7] Winkler, I. Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci., 13*, (12), 532–540.

How Tone Sandhi Helps Us Understand the Processing of Phonological Alternation

Jie Zhang

University of Kansas (USA)
zhang@ku.edu

The understanding of how speakers internalize patterns of morpheme alternation (e.g., *dog[z]*, *cat[s]*, *bus[ɨz]*) is one of the fundamental goals of phonological research. The generative approach to the phenomenon is to assume that the different allomorphs of a morpheme are united by one underlying representation, and an input-output mapping mechanism, as part of the competence grammar, maps the underlying representation to the different surface representations in the appropriate contexts. Although phonological theories along this line do not make the claim that speakers perform the mapping in on-line production, they do make predictions about the different levels of representations that speakers are able to access in the production and perception processing of words that involve alternation, and some psycholinguistic research based on priming has shown that abstract underlying representations of morphemes play a role in spoken word recognition [1-5]. However, experimental work that investigates the processing of alternation remains relatively scarce, and current models of speech production [6-8] and spoken word recognition [9-12] are largely agnostic regarding where alternation fits in the model.

In this talk, I discuss a series of work on the productivity and processing of tonal alternation (commonly known as tone sandhi) patterns in dialects of Chinese and show that (a) the richness of the attested tone sandhi patterns provides a fertile testing ground for the processing of alternation, and (b) how an alternation is learned and processed on-line by speakers is strongly influenced by the phonological nature of the alternation. In particular, I focus on the comparison between tone sandhi patterns that can be driven by surface-true phonotactics and those that involve a circular chain shift and hence cannot.

One set of comparisons comes from tone sandhi patterns in which the nonfinal syllables of the tone sandhi domain undergo paradigmatic substitution, as illustrated by the Mandarin, Hailu Hakka, and Taiwanese Southern Min patterns in (1). Wug test results showed that the phonotactically motivated substitution patterns in Mandarin [13] and Hailu Hakka [14] are generally productive, while substitution patterns that involve a circular chain-shift in Taiwanese Southern Min lack full productivity and often categorically fail to apply to nonce words [15, among others]. Auditorily primed lexical decision tasks showed that disyllabic words undergoing tone sandhi in Mandarin [16] and Hailu Hakka [14] are more strongly primed by the first syllable of the word carrying the underlying tone, while tone sandhi words in Taiwanese Southern Min are more strongly primed by the sandhi tone [17]. Moreover, the priming effect of the underlying tone in Mandarin and Hailu Hakka does not interact with the frequency of the target word, while the priming effect of the sandhi tone in Taiwanese Southern Min is regulated by frequency.

(1a) Mandarin: 213 35 / __ 213
(1b) Hailu Hakka: 13 33 / __ X
(1c) Taiwanese Southern Min: 51 55 33 24 / __ X
↖ ↙
21

Another set of comparison comes from tone sandhi patterns in which the tone on the initial syllable of a tone sandhi domain is spread over the domain, either directly or after substitution, as illustrated by the Shanghai Wu and Wuxi Wu patterns in (2). Wug test results showed that the tone spreading pattern in Shanghai Wu is generally productive [18], while the substitution aspect of the tone sandhi in Wuxi Wu, which involves a circular chain shift, is not [19].

(2a) Shanghai Wu: 51-X 55-31
(2b) Wuxi Wu: 53-X 43-34

Although the studies reported here only focus on one property of phonological alternation — whether the alternation is driven by surface-true phonotactic, the results suggest that different types of phonological alternations may be learned by speakers in different ways and involve different processing mechanisms on-line. When accompanied by additional experimental paradigms as well as replication studies of similar phenomena, these studies have the potential to inform both models of lexical access and the nature of phonological grammar. It is also interesting to observe that the relation between the productivity of phonological alternation and how the alternation is processed has a striking parallel to the well-studied relation between morphological irregularity/semantic compositionality and morphological processing [e.g., 20]. It is hoped that the vast psycholinguistic and neurolinguistic literature on the latter can provide guidance for our investigation and modelling of phonological alternation.

References

[1] Lahiri, Aditi, Allard Jongman, and Joan Sereno (1990). The pronominal clitic [dər] in Dutch: A theoretical and experimental approach. In Geert E. Booij and Jaap van Marle (eds.), *Yearbook of Morphology, Vol. 3*, pp. 115-127. Dordrecht: Foris Publications.

[2] Gaskell, M. Gareth, and William D. Marslen-Wilson (1996). Phonological variation and inference in lexical access. Journal of Experimental Psychology: *Human Perception and Performance* 22: 144-58.

[3] Luce, Paul A., Connor T. McLennan, and Jan Charles-Luce (2001). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In J. Bowers and C. Marsolek (eds.), *Rethinking implicit memory*, pp. 197-214. Oxford: Oxford University Press.

[4] Gow, David W. Jr. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 45: 133-159.

[5] Gow, David, W. Jr. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28: 163-79.

[6] Levelt, William J. M. (1999). Models of word production. *Trends in Cognitive Science* 3(6): 223-232.

[7] Levelt, William J. M., Ardi Roelofs, and Antje S. Meyer (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22(1): 1-38.

[8] Indefrey, P. and William J. M. Levelt (2004). The spatial and temporal signatures of word production components. *Cognition* 92(1-2): 101-144.

[9] McClelland, James L. and Jeffrey L. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology* 18(1): 1-86.

[10] Marslen-Wilson, William D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25(1-2): 71-102.

[11] Norris, Dennis (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52(3): 189-234.

[12] Norris, Dennis and James M. McQueen (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review* 115(2): 357-395.

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

[14] Zhang, Jie, Hanbo Yan, Yuwen Lai, and Shaoren Lyu (2016). An experimental investigation of positionally conditioned tone sandhi in Hailu Hakka. Poster presented at the *15th Conference on Laboratory Phonology*, Ithaca, NY.

[15] Zhang, Jie, Yuwen Lai, and Craig Sailor (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua* 121.2: 181-206.

[16] Chien, Yu-Fu, Joan Sereno, and Jie Zhang (2016). Priming the representation of Mandarin tone 3 sandhi words. *Language, Cognition and Neuroscience* 31.2: 179-189.

[17] Chien, Yu-Fu, Joan Sereno, and Jie Zhang (2017). What's in a word: Observing the contribution of underlying and surface representations. *Language and Speech* 60.4: 643-657.

[18] Zhang, Jie and Yuanliang Meng (2016). Structure-dependent tone sandhi in real and nonce words in Shanghai Wu. *Journal of Phonetics* 54.1: 169-201.

[19] Yan, Hanbo and Jie Zhang (2016). Pattern substitution in Wuxi tone sandhi and its implication for phonological learning. *International Journal of Chinese Linguistics* 3.1: 1-45.

[20] Lemine, Alina, Eva Smolka, Jon A. Duñabeitia, and Christos Pliatskas (in press). Morphological processing in the brain: The good (inflection), the bad (derivation), and the ugly (compounding). *Cortex*. Accessible on-line at https://doi.org/10.1016/j.cortex.2018.08.016.

A Generative Phonetic Analysis of the timing of L- Phrase Accents in English

Edward Flemming

Massachusetts Institute of Technology (USA)
flemming@mit.edu

This paper builds on Pierrehumbert's foundational work on the generative phonetics of English intonation [1] through an analysis of the timing of the English low phrase accent (L-) in H*L-L% and H*L-H% melodies, developing a model of the realization of tones that has three key elements:

(i) F0 trajectories are modelled as the response of a dynamical system to a control signal that consists of a sequence of step functions connected by linear ramps.
(ii) The mapping from tones to F0 trajectory is derived by optimization with respect to conflicting constraints.
(iii) Variation in realization is derived by formulating the grammar as a Maximum Entropy model

Identifying elbows

Studying the timing of L- in H*L- sequences is challenging because L- is realized as an 'elbow' in the F0 trajectory – i.e. a point of inflection rather than a local maximum or minimum – when there is sufficient time between the H* peak and the end of the phrase (fig. 1). Algorithms for locating F0 elbows have been proposed [2, 3], but they are difficult to evaluate because there is no independent evidence concerning the true locations of elbows.

The approach adopted here is to develop a more explicit model of the interpolation between tonal targets. This makes it possible to infer the location of targets by fitting the model to the observed F0 contour. Specifically, the fall from the H* peak is modelled as the response of a cascade of four identical first order linear dynamical systems to a step function input, and the L- plateau is modelled as the response of the same system to a linear input (fig. 2). This type of model has been used to model articulator movements [4], and fits F0 contours much better than the more familiar critically damped second order dynamical systems [5] because it allows for variation in the timing of the acceleration peak.

The timing of L- elbows

This analytical method is used to test two hypotheses concerning the timing of L- elbows: (a) L- occurs at a fixed interval after H*, (b) L- is aligned to the end of the nuclear-accented word [1]. The data were recordings collected for a previous study [6]. The materials consist of 25 two-word phrases read by 15 speakers in a context designed to elicit an H*L-H% melody with H* on the first word (although some utterances were produced with H*L-L%). The number and length of the syllables following the primary stress in the first word were systematically varied (e.g. *á<u>lien</u>* vs. *pá<u>limony</u>*) to test if the interval from the H* peak to L- elbow varies to align L- to the word boundary or not.

The results do not support either hypothesis: L- is not aligned to the word boundary, but there is a significant tendency for L- to occur earlier when the interval between H* and the word boundary is shorter (fig. 3) (cf. [6]). This pattern of realization is analysed as a compromise between two constraints, one enforcing a target duration for the fall from H* to L-, and a second requiring the fall to be completed before the end of the word. The second constraint has lower weight, so the H*-L- duration only decreases by a small proportion of a decrease in duration to the word boundary.

Maximum Entropy phonetic grammar

The variation in L- timing within and between speakers is more striking than the effect of variation due to timing of the word boundary. Variation in phonetic realization can be modelled by adapting the widely used MaxEnt approach to phonological variation [7]. In MaxEnt grammar, candidates are assigned a numerical wellformedness score, H, equal to their summed constraint violations, and the probability of a candidate is proportional to e^{-H}. If the cost of violating a phonetic constraint is proportional to the square of the deviation from the constraint target, then the MaxEnt model derives a normal distribution over the space of realizations, with its mean at the lowest cost realization, and variance dependent on the summed constraint weights.

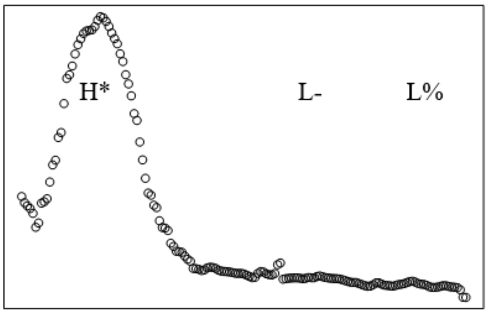


Fig.1 F0 track of an English H*L-L% melody produced on the phrase 'alien annihilator'

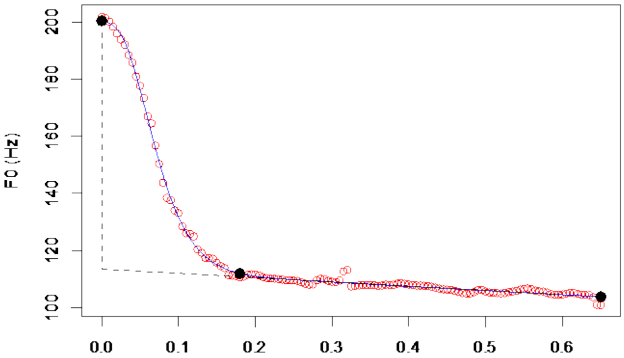


Fig.2 F0 measurements (red), model F0 trajectory (solid line), and input to the production model (dashed lines) for the utterance shown in fig. 1.

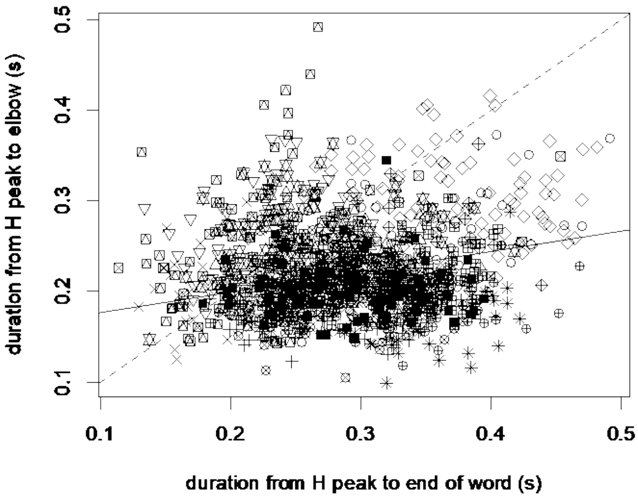


Fig.3 Plot of the relationship between duration from H* to the L- elbow and durationship from H* to end of word. The solid line shows the best-fitting linear relationship, and the dashed line shows the expected trend if L- is aligned to the end of the word. Speakers are plotted with different symbols.

References

[1] Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. Thesis, MIT, Cambridge, MA.

[2] del Giudice, A., Shosted, R., Davidson, K., Salihie, M., & Arvaniti, A. (2007). Comparing methods for locating pitch "elbows". *Proceedings of the 16th International Congress of Phonetic Sciences*, 1117-1120.

[3] Reichel, U., & Salveste, N. (2015). Pitch elbow detection. *ESSV*, Katholische Universität Eichstätt-Ingolstadt, Germany, 25.–27. March 2015. Ed. Wirsching, G. Dresden: TUD Press, 143−149.

[4] Birkholz, P., Kröger, B.J., & Neuschaefer-Rube, C. (2011). Model-based reproduction of articulatory trajectories for Consonant-Vowel sequences. *IEEE Transactions on Audio, Speech, and Language Processing*, 19, 1422-1433.

[5] Fujisaki, H. & Hirose, K. (1984), Analysis of voice fundamental frequency contours for declarative sentences of Japanese, *Journal of the Acoustical Society of Japan*, 5, 233-242.

[6] Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2010). Turning points, tonal targets, and the English L- phrase accent. *Language and Cognitive Processes*, 25:7-9, 982-1023.

[7] Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In J. Spenader, A. Eriksson, and O. Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120.

The role of intonation in vowel insertion

Martine Grice

University of Cologne (Germany)
martine.grice@uni-koeln.de

In Italian, the pronunciation of loan words with a final consonant, such as <chat> or <facebook>, is often characterised by an additional word-final schwa. The insertion of this schwa is highly variable and dependent on a number of factors. These include speaker-specific preferences, metrical structure and the identity of the word-final consonant. Crucially, a considerable amount of variation is conditioned by intonation: A vowel is more likely to occur – and is acoustically more prominent – if the intonation is complex or rising than if it is falling (Grice, Savino & Roettger, 2018).

Another language in which vowels can be inserted, word-finally and word-medially, is Tashlhiyt Berber (Ridouane & Fougeron, 2011, Grice, Ridouane & Roettger, 2015). This language is known for its long consonantal, even voiceless, sequences but also for a high degree of variability in the insertion of vocalic elements. Here too, schwa insertion is not only affected by speaker-specific patterns and properties of adjacent consonants, but also by intonation, with it being more likely to surface in positions in which communicatively relevant tonal movements are realised.

In both of these languages, the insertion of a vowel creates a segmental environment with high periodic energy and a rich harmonic structure, which is optimal for realising intonation contours. The need to realise intonational tones – the *tune* – thus leads to the enhancement of the segmental makeup of the material bearing these tones – the *text*. The insertion of vowels is akin to the lengthening of vowels that are already present. This takes place in words ending in open syllables in Italian (Grice, Savino & Roettger, 2019), again when there is pressure to realise a complex or rising tune.

The other side of the coin is the preservation of tone-bearing material, such as the blocking of vowel devoicing and deletion in a number of languages (Gordon, 1998; Kilbourn-Ceron & Sonderegger, 2018) in similar tonal contexts, albeit some involving lexical rather than intonational tones. Inserting or preserving vocalic material serves to increase the likelihood of transmitting and retrieving intonational meaning (Roettger & Grice to appear), in line with findings that suggest that the interpretation of intonational contours is strongly impaired if conflicting requirements of tune and text lead not to enhancement of the text but to a truncation of the tune (Odé 2005; Rathcke 2013). Insofar as vowel insertion facilitates the production and perception of intonation, it is relevant for phonology and cannot be treated simply as a phonetic artefact.

References

[1] Grice, Martine, Timo Roettger & Rachid Ridouane. (2015). Tonal association in Tashlhiyt Berber: Evidence from polar questions and contrastive statements. *Phonology* 32(2): 241-266.

[2] Grice, Martine, Michelina Savino &Timo Roettger. (2018). Word final schwa is driven by intonation – The case of Bari Italian, *Journal of the Acoustical Society of America*, 143 (4): 2474–2486.

[3] Grice, Martine, Michelina Savino & Timo Roettger (2019) Tune-text negotiation: The effect of intonation on vowel duration, *Proc. International Congress of Phonetic Sciences*, Melbourne, August 2019.

[4] Gordon, M. (1998). The phonetics and phonology of non-modal vowels: a cross-linguistic perspective. In Annual Meeting of the Berkeley Linguistics Society (Vol. 24, No. 1, pp. 93-105).

[5] Kilbourn-Ceron, O., & Sonderegger, M. (2018). Boundary phenomena and variability in Japanese high vowel devoicing. Natural Language & Linguistic Theory, 36(1), 175-217.

[6] Odé, Ceclia (2005). Neutralisation or Truncation? The perception of two Russian pitch accents on utterance-final syllables, *Speech Communication*, 47 (1-2), 71-79.

[7] Ridouane, Rachid & Cecile Fougeron (2011) Schwa elements in Tashlhiyt word-initial clusters, Journal of the Association of Laboratory Phonology, 2 (2), 275-300.

[8] Roettger, Timo & Martine Grice (to appear). The tune drives the text - Competing information channels of speech shape phonological systems, *Language Dynamics and Change*.
[9] Rathcke, Tamara (2013). On the neutralizing status of truncation in intonation: A perception study of boundary tones in German and Russian, *Journal of Phonetics* 41 (3-4), 172-185.

Oral Presentations
(Day 1, General Sessions)

The role of lexical tone in bilingual language processing

Xin Wang[1], Siyu Chen[2] & Bronson Hui[3]

[1]*Macquarie University (Australia),* [2]*University of Greenwich (UK),* [3]*Michgan State University (USA)*
x.wang1@mq.edu.au

In the bilingualism literature, cross-language activation has been demonstrated to support the phenomenon that bilinguals can't simply shut off the language not-in-use when they are processing the other language. Cross-language activation can be driven by two sources of information: phonological overlap or automatic/implicit translation. For instance, when Russian-English bilinguals were listening to the word 'marker' in English only, their Russian word 'marku' was activated as well [1]. This cross-language activation was driven by between-language phonological overlap (i.e., 'marker' in English and 'marku' in Russian are phonologically overlapped to some extent). In addition, automatic/implicit translation was first demonstrated in ERP measures by Thierry and Wu (2007) [2]. In their study, Chinese-English bilinguals showed priming effects between 'ham' and 'train', which was absent in native English speakers. This effect was attributed to the shared Chinese character /huo/ in the translations of 'ham' and 'train'. The current project seeks to understand the role of lexical tones in cross-language activation through both mechanisms: phonological overlap and automatic/implicit translation.

In the literature of Mandarin spoken word recognition, lexical tones are mostly considered as important as segments and share similar time course during lexical activation [3]. However, little is known about the status of lexical tone in bilingual language processing. In a recent effort to investigate the role of lexical tone in automatic/implicit translation, Wang et al. (2017) demonstrated that Mandarin-English bilinguals were able to activate 'feather' when listening to 'rain' in English only, using the visual world paradigm [4]. This cross-language effect was attributed to the activation of the Mandarin translation (e.g., /yu3/) of 'rain', which subsequently activated 'feather' whose Mandarin translation is also /yu3/. Importantly, this effect was absent in 'fish' whose Mandarin translation is /yu2/. This contrast showed that lexical tones were critical in cross-language activation. That is, lexical tones were activated during the implicit access to L1 Mandarin when bilinguals were processing their L2 English only.

What about the role of lexical tone in cross-language activation when the input shares similar segmental information with the non-target language (i.e., phonological overlap)? First, we report results from two auditory lexical decision tasks where we instructed Mandarin-English bilinguals to decide whether the sound they heard was a real word or not in English. We manipulated two conditions: inter-lingual homophones (e.g., 'bay' sounds similar to Mandarin /bei4/) vs. non-inter-lingual homophones, as well as controlling other psycholinguistic lexical variables in Experiment 1. Stimuli were recorded by a native English speaker. We didn't find any difference between these two conditions for Mandarin-English bilingual listeners. This means that Mandarin lexicon was not activated with phonological overlap. In Experiment 2, we superimposed Mandarin tones onto English monosyllabic words, the same stimuli used in Experiment 1, we found significant delay in processing inter-lingual homophones by bilinguals. This suggests that Mandarin words were activated to induce lexical competition and this cross-language effect was driven by phonological overlap only with the presence of lexical tones.

However, tonal superimposition creates unnaturalness in the spoken stimuli and affects the comprehension behaviour of bilingual listeners in an unknown way. In Experiments 3 and 4, to avoid these methodological caveats, we only employed Mandarin words in Tone 4, which sound equivalent to English words pronounced naturally in a falling pitch. We created naturally produced spoken stimuli in English and Mandarin by natives but also controlled other psycholinguistic variables so that words in English and Mandarin were matched with each other for comparison. Similar to Experiment 1 and 2, we employed two auditory lexical decision tasks to compare the difference between inter-lingual homophones and non-inter-lingual homophones: one exclusively in English (Experiment 3, e.g., 'bay' as the inter-lingual homophone) and the other exclusively in

Mandarin (Experiment 4, e.g., ‘/bei4/ as the inter-lingual homophone). If bilingual listeners were sensitive to the pitch contours associated with syllables that are meaningful in Mandarin, we would observe inhibition for inter-lingual homophones due to lexical competition. However, we didn’t observe inhibition on the inter-lingual homophones in the English task. That is, the falling pitch in naturally produced English stimuli was not sufficient to induce lexical competition from Mandarin, even though they sound like Mandarin Tone 4. In the Mandarin task, we observed inhibition on the inter-lingual homophones as what we predicted as a result of lexical competition. That is, Mandarin /bei4/ activated English ‘bay’, but not the other way around. This asymmetry suggests that bilingual listeners were able to restrict lexical activation to English only in Experiment 3 when the acoustic alignment between tones and segments was not native-like. In Experiment 4, when bilinguals were processing their L1 Mandarin only, phonological overlap activated their L2 at the segmental level. This is consistent with the literature, as a result of cross-language lexical competition. Again, these results support the claim that lexical tones are critical in activating Mandarin lexicon during cross-language activation. We will discuss these results in relevant bilingual models.

References

[1] Spivey, M., & Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. Psychological Science, 10, 281-284.
[2] Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. Proceedings of National Academy of Science USA, 104, 12530–12535.
[3] Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. Journal of Memory and Language, 62(4), 407–420.
[4] Wang, X., Wang, J. & Malins, J. (2017) Do you hear ‘feather’ when listening to ‘rain’? Lexical tone activation during unconscious translation: Evidence from Mandarin-English bilinguals. Cognition, 169, 15-24. https://doi.org/10.1016/j.cognition.2017.07.013.

Production of Korean Accentual Phrases by Cantonese-speaking Learners

Xinran Ren[1] & Peggy Mok[2]

[12]*The Chinese University of Hong Kong (Hong Kong)*
renxinr@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

Cantonese is a tone language, in which the realization of intonation is restricted mainly to the sentence-final syllable [1]. Accordingly, native Cantonese speakers may have difficulty acquiring second language intonation with richer pitch variation in other parts of the utterance. For example, it has been suggested that Hong Kong English (HKE) intonation is formed in such a way that every syllable in an English utterance is assigned a Cantonese level tone to represent the stress patterns of each English word [2]. However, while English can be described as a stress accent language with the pitch accent anchored to a stressed syllable, Korean has a relatively fixed phrasal pitch accent, namely, TH-LH (T=H/L depending on the feature [stiff vocal cords]) in the accentual phrase (AP), a prosodic unit smaller than an intonation phrase (IP) but larger than a phonological word [3]. This study examined a set of non-native intonation patterns by Cantonese learners in the production of Korean accentual phrases. Several dimensions (prosodic phrasing, speaking rate, segment-tone interaction and surface tonal realization) were examined in 4 native Korean speakers and 17 Cantonese learners (beginning vs. intermediate) using a sentence reading experiment. We try to explore the developmental patterns of these dimensions in non-native intonation and whether they result from a universal acquisition process or prosodic transfer from L1.

40 simple sentences were adapted from the materials in [4]. These sentences all ended with 'giM-ci-rIR zo-a-hA-yo' meaning 'like kimchi' but differed in subject phrases. In order to test the word size effect and segment-tone interaction in the learners, these subject phrases varied among 3-7 syllables including a topic marker 'nIN', and half of the initial syllables started with a [+stiff vocal cords] segment while the other half started with a [-stiff vocal cords] one. A sample sentence is 'mi-na i-mo-nIN giM-ci-rIR zo-a-hA-yo' (Mina's untie likes kimchi). Phrasings and tonal patterns of the subject phrases were marked using Korean ToBI. The phrasing of APs and IPs was cued by boundary tone contours (usually LHa for APs and L%/HL% for IPs in declaratives) and types of pauses (with or without a strong subjective sense of pause) [3].

The results showed that the Cantonese learners produced AP boundary H tones consistently, but had various problems in realizing the non-boundary tones: as shown in Figure 1, comparing with native Korean speakers who produced a four-tone HHLH pattern for the 5-syllable AP, the non-native speech was featured by a relatively flatter contour, e.g., HH-Ha (middle-right), and by adding one more tone for the 5-syllable AP (bottom-right). This more-tone pattern has a high error rate of over 70% across both learner groups, and it mirrors HKE in which every syllable in English is assigned a Cantonese level tone. For the other dimensions, the two learner groups sometimes patterned together in terms of segment-tone interaction, speaking rate and the influence of speaking rate on AP length. However, the intermediate learners also showed significant improvement in that they produced longer APs and IPs and a higher AP-final H tone towards the native pattern.

This study confirmed several universal properties of non-native prosody at early stages, such as more intonational and accentual phrases (i.e., more truncated phrasing) and slower speech rates. L1-specific transfer was also observed on the surface tonal patterns, similar to Hong Kong English. Furthermore, our findings suggest that some phonological properties of intonation, e.g. boundary tone types, are acquired earlier than some phonetic ones such as surface tonal realization.

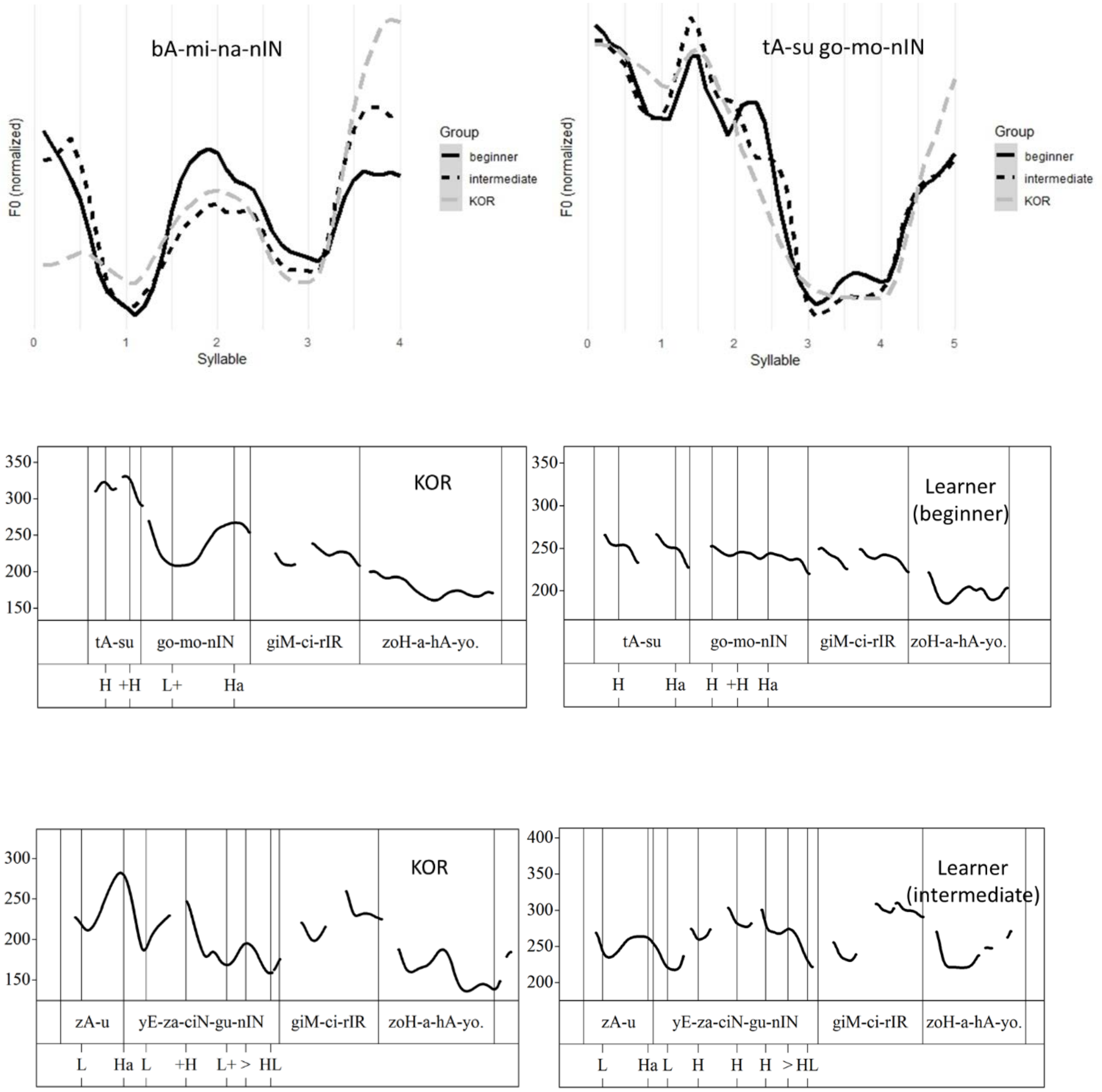


Fig.1 Pitch tracks of three different subject phrases (4-syllable vs. 5-syllable vs. 7-syllable): (top) the average f0 contours by syllables in different groups; (middle and bottom) single tokens from native Korean speakers (left) and learners (right).

References

[1] Chow, S. (2002). *A phonetic study of the intonation in Hong Kong Cantonese* (Doctoral dissertation, City University of Hong Kong).

[2] Luke, K. K. (2008). Stress and intonation in Hong Kong English. In *14th Conference of the International Association for World Englishes (IAWE)* (pp. 1-5). Hong Kong.

[3] Jun, S. A. (2005). Korean intonational phonology and prosodic transcription. In *Prosodic typology: The phonology of intonation and phrasing*. (pp. 201-229). Oxford University Press.

[4] Jun, S. A. & Oh, M. (2000). Acquisition of second language intonation. In (Ed.), *Sixth International Conference on Spoken Language Processing (ICSLP)* (pp. 73-76). Beijing, China.

Development of Speech Perception in Korean Infants: Discriminating Unusual Sound Contrasts with Diachronic Change

Youngon Choi[1], Minji Nam[1], Reiko Mazuka[2, 3], Hyun-Kyung Hwang[2], Minha Shin[1], Jisoo Kim[1], Sunho Jung[1], Jimin Beom[1], & Naoto Yamane[2]

[1]Chung-Ang University (Korea), [2]RIKEN Center for Brain Science (Japan), [3]Duke University (USA)
yochoi@cau.ac.kr, minjinamm@gmail.com, reiko.mazuka@riken.jp, hyunkyung.hwang@riken.jp, shinmh21@naver.com, cindyjisoo@naver.com, sho9517@gmail.com, beomjimin.psyche@gmail.com, yamanen@brain.riken.jp

Compared to other languages, phonetic categories that are utilized in Korean are relatively unusual: (1) stops and affricates contrast in three ways as fortis, lenis, and aspirated and they are all voiceless (e.g., for bilabial stops: /p*/, /p/, /p^h/, and for affricate: /c*/, /c/, /c^h/). Fricative contrast in Korean is also all voiceless unlike other languages although they do not contrast in three ways. (2) two of the three categories, lenis and aspirated, have also undergone a tonogenetic sound change in the past few decades such that voice onset time (VOT) is no longer the primary cue for distinction and instead, fundamental frequency (F0) has become the primary cue for their distinction [1]. These unique phonetic characteristics and diachronic change that has occurred in Korean present a unique opportunity to study the emergence of phonetic categories in young infants and the specific roles that infants' initial perceptual capacities and the phonetic input play in speech perception development. In the past several decades, much has been uncovered about the development of speech perception and yet, it still is crucial to investigate cases such as Korean to better understand how diverse phonetic categories emerge and what factors contribute to these development.

We have been investigating Korean infants' development of the ability to discriminate these sound categories, including three-way bilabial stop contrasts [2], two-way fricative sound contrast [3, 4], and three-way affricate contrasts (currently on-going). Using a modified visual habituation paradigm [5, 6], we have tested 4 through 12 months olds to examine the developmental patterns of these sound categories.

For fricative contrast, presented in /a/ vowel context, spoken by a female speaker using Infant-directed speech register such as fortis /s*a/ (mean aspiration duration = 3.6 (SD=1), mean noise duration = 170.07 (SD=30)) versus lenis /sa/ (mean aspiration duration = 70.4 (SD=19), mean noise duration = 123.27 (SD=10)), we observed that 7-9-month-olds (N = 23, 9 girls, mean age = 255 days, age range: 215~301 days) reliably discriminated the fricative pair, displaying significantly increased looks to the changed sound category trial ($F(1, 21) = 3.807$, $p = .016$) as compared with the unchanged sound trial. By comparison, the younger infants (N = 21, 7 girls, mean age = 165 days, age range: 138~209 days) did not show a clear sign of distinction ($F(1, 19) = 0.728$, $p = .404$). These results suggested that Korean infants begin to discriminate fortis-lenis fricative pair around 7 months of age and not before [3]. Interestingly, when the fricative contrast was presented in /i/ vowel context [4], 7-8-month-olds (N = 20, 8 girls) could not distinguish the contrast ($t(19) = .004$, $p = .997$). It was 8 1/2 to 9 months of age (N = 20, 10 girls) that could reliably discriminate the fricative pair ($t(19) = 2.236$, $p = .038$). This finding additionally showed that infants' abilities to discriminate the phonetic contrast may be somewhat limited and initially context-bound and slowly generalize to various phonetic contexts.

For bilabial stop contrasts, even 7-9-month-olds (N = 24, 11 girls, mean age = 255 days, age range: 212~296 days) did not show signs of clear discrimination for lenis-aspirated contrast (i.e., /pu/ versus /p^hu/) because their looking times did not differ by the change in the sound category ($F(1, 23) = 0.248$, $p = .623$). 10-12-month-olds (N = 27, 16 girls, mean age = 340 days, age range: 307~377 days), however, showed significant difference in their looking time in response to the sound category change ($F(1, 26) = 7.030$, $p = .013$), suggesting that it is after about 10 months of age that Korean infants begin to discriminate this pair [2]. By comparison, we observed that infants as young as 5-6 months (N = 24, 11 girls, mean age = 182 days, age range: 160~209 days) could discriminate the fortis-aspirated contrast ($t(23)= 2.592$, $p = .016$). For fortis-lenis contrast, it appeared taking time until Korean infants begin reliably distinguishing them. In our study, it was

8-9 months of age that infants (*N* = 20, 9 girls, mean age = 278 days) could clearly discriminate this pair but 7-8-month-olds (*N* = 20, 9 girls, mean age = 238 days) were not able to discriminate this contrast. Furthermore, mothers of 7-9-month-old infants showed individual differences in producing VOT differences across the three pairs and the individual differences in VOT particularly for fortis-lenis pairs showed a positive correlation ($r(20) = .507$, $p = .023$) with their infants' looking time differences between the same and change category trials. The effect of maternal phonetic differences accounted for about 17.6% of 7-9-month-olds' discrimination behavior even after controlling for infant's age, which showed a positive correlation with infants' discrimination ability [7]. These suggested that the larger the related acoustic cue was provided by the mothers, the better their infants were to distinguish the two sound categories.

Currently, we are testing infants between 4 and 6 months with fortis-aspirated affricate pair (i.e., /c*a/ versus /c^ha/). Although the results are preliminary (with 13 babies participated thus far), so far it appears that Korean infants begin to discriminate this pair around 5-6 months, similar to our observation on fortis-aspirated bilabial stop contrast.

Taken together, our findings demonstrate that Korean infants' learning of three-way contrasts may take longer than those whose language utilizes two way contrast. More importantly, the sound contrasts that have undergone a tonogenetic sound change were the latest for Korean infants to become able to discriminate. Our finding on the relationship between maternal phonetic input and infants' individual differences in their discrimination abilities suggest that this delay may have something to do with this on-going change. In another investigation, we observed that maternal input varied more widely than expected: some mothers reflect the tonogenetic change in their input when they speak to their infants but not to other adults. But other mothers do not fully reflect the diachronic change in their input to their babies. The mixed nature of the input due to this on-going diachronic change might delay young infants as they discover the relevant sound categories in their language. This, in turn, suggests that the acoustic properties of the input provided for infants play an important role in the development of speech perception.

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] Choi, Y., Nam, M., Shin, M., Yamane, N., & Mazuka, R. (2017). Emergence of Korean infants' abilit to discriminate the three-way stop contrasts: contribution of initial biases and nature of input. *Paper presented at International Association for the Study of Child Language meeting*. Lyon, France, July, 2017.

[3] Shin, M., Choi, Y., & Mazuka, R. (2018a). Development of fricative sound perception in Korean infants: The role of language experience and infants' initial sensitivity. *PloS one*, *13*(6), e0199045.

[4] Shin, M., Choi, Y., & Mazuka, R. (2018b). Development of Korean infants' fricative perception: comparison across different vowel contexts. *Poster presented at the International Congress on Infant Studies meeting*, Philadelphia, USA, July, 2018.

[5] Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381–382.

[6] Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology*, *46*(1), 106–119. https://doi.org/10.1037/a0016718

[7] Nam, M., Choi, Y., Shin, M., Kim, S., & Mazuka, R. (2018). The role of maternal phonetic input on their infants' speech perception ability. *Poster presented at the International Congress on Infant Studies meeting*, Philadelphia, USA, July, 2018.

The Role of Information Structure and Foot Structure on the Kinematics of Unstressed Syllables

MIAO ZHANG[1]

[1]*University at Buffalo, SUNY (USA)*
miaozhan@buffalo.edu

Speakers vary their production along a continuum of hypoarticulation and hyperarticulation under different contexts [1]. One of those contexts is information structure. A large body of research has explored the acoustic effect of accentual lengthening on syllables when the word is contrastively focused [2]–[5]. Articulatorily, in a study examining the role of information structure on speech articulation in German, Mücke & Grice [6] found that supralaryngeal articulation can be modified by information structure alone rather than mediated by accentuation. However, no research to date has looked into how the articulation of such syllables is influenced by focus. Moreover, no articulatory research to date has explored how post-lexical foot structure influences the articulation of unstressed syllables.

This study aims to explore how supralaryngeal articulation of unstressed syllables varies by information structure and post-lexical foot structure. Two questions are to be answered: i) Is articulation of unstressed syllables affected by focus? ii) Does the foot structure play a role in hyper/hypoarticulation? While the first question is motivated by an empirical gap in the literature, the second explores a prediction from research on the prosodic hierarchy. Namely, greater articulatory effort is involved at higher levels of the prosodic hierarchy [6], [7]. Thus footed syllables should involve greater articulatory effort.

Movement of consonantal opening gestures in word-initial, unstressed syllables was investigated via an experiment using Electromagnetic Articulography (EMA). Three articulators (the lower lip, tongue tip, and tongue body) were examined. The effect of focus was investigated by different question-answer pairs, and the effect of foot structure was investigated by altering the stress pattern of the preceding word (‘trochee’: σσ́#**σ**, “permit **Pa**tricia” [pɚ(ˈmɪt # **pə**)$_F$(ˈtɹɪ.ʃə)$_F$] vs. ‘dactyl’: σ́σ#**σ**, “limit **Pa**tricia” [(ˈlɪ.mɪt)$_F$ # **pə**(ˈtɹɪ.ʃə)$_F$]). Four English native speakers were recorded. Four kinematic measures were taken: opening gesture duration (from the constriction to maximum opening), maximum displacement (distance travelled from constriction to maximum opening), peak velocity, and time-to-peak velocity (starting from the constriction). These measures were extracted using MVIEW [8] in MATLAB.

The overall results across speakers are displayed in Fig 1 and Fig 2. After eliminating mispronounced trials and outliers, a total of 627 tokens were examined. Both a pooled data analysis and a by-speaker analysis were performed. Each kinematic measure was examined via four-way ANOVA. In the by-speaker analysis, a series of two-way ANOVAs were performed for each articulator by speakers. The pooled results are shown in Table 1 with by-speaker results shown in Tables 2.

The pooled analysis showed that the four kinematic measures were all affected by both focus and foot structure. Within the contrastive focus and trochaic foot condition compared to other conditions, unstressed syllables were produced with overall greater opening duration and maximum onset consonant displacement, faster but less stiff movement of articulators when compared with broad and background conditions. The interaction between focus and foot was not significant. Inter-speaker variation was found for information structure, whereas the effect of foot structure was more consistent across speakers. Speakers F1, F2 exhibited more articulatory strengthening across focus conditions, while speaker M1 and M2 exhibited less. F1, F2 showed more strengthening in contrastive/broad focus whereas M1 showed more in broad focus.

The results show that supralaryngeal articulation in unstressed syllables is influenced by information structure and metrical foot structure, as predicted by focus-induced modification and prosodic strengthening respectively. Post-lexically unfooted syllables involve less articulatory effort in comparison with footed syllables.

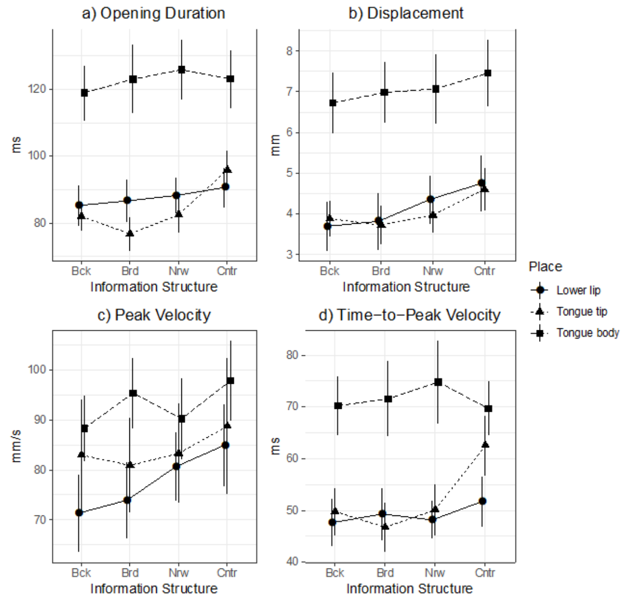


Fig 1. Means and 95% CIs for the kinematic measures by information structure

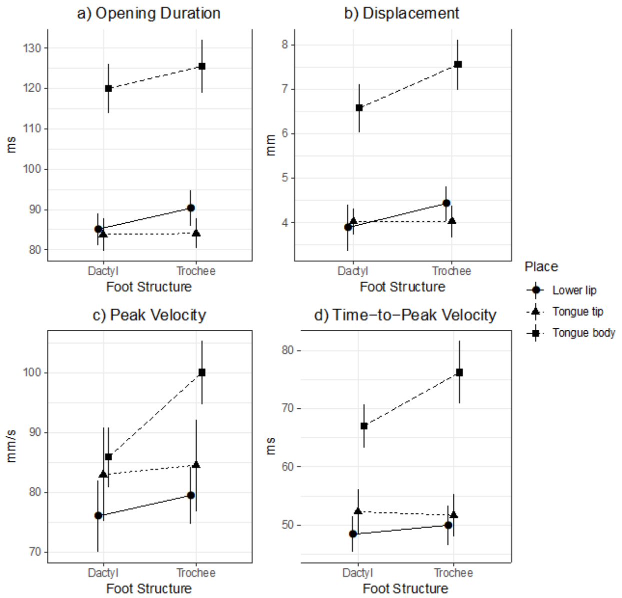


Fig 2. Means and 95% CIs for the kinematic measures by foot structure

Table 1. Results of overall statistical analysis

***: $p < .001$, **: $p < .01$, *: $p < .05$, only two-way interactions are reported

	Duration	*Displacement*	*Peak velocity*	*Time-to-peak velocity*
Main effects				
Info str. ($F(3, 531)$)	4.11**	7.23***	4.61**	3.21*
Place of Articulation ($F(2, 531)$)	177.185***	182.31***	21.97***	104.09***
Foot str. ($F(1, 531)$)	4.05*	10.96***	11.21***	5.38*
Speaker ($F(3, 531)$)	9.52***	71.08***	106.54***	28.12***
Interactions				
Info str. × Place ($F(6, 531)$)	1.90	0.49	1.02	3.22**
Place × Foot str. ($F(2, 531)$)	0.60	3.04*	4.08*	4.09*
Info str. × Foot str. ($F(3, 531)$)	0.68	1.42	0.78	1.76
Interactions with Speaker				
Info str. × Speaker ($F(9, 531)$)	2.45**	2.75*	1.95	1.68
Place × Speaker ($F(6, 531)$)	11.47***	30.64***	28.78***	12.89***
Foot str. × Speaker ($F(3, 531)$)	2.63*	0.81	1.58	7.21***

Table 2. Results of comparisons between information structures for each speaker.

**: $p < .017$; *: $p < .05$; Ø = background, B = broad focus, N = narrow focus, C = contrastive focus

Speaker	**C vs. ∅**				**N vs. ∅**				**B vs. ∅**			
	Longer	Larger	Faster	Less stiff	Longer	Larger	Faster	Less stiff	Longer	Larger	Faster	Less stiff
F1	/t/**	/t/**		/t/**								
F2		/p/**	/p/*			/p/**						
M1							/t/*		/k/**	/k/**	/t/*	
M2	/t/*											
Speaker	**C vs. B**				**C vs. N**				**N vs. B**			
	Longer	Larger	Faster	Less stiff	Longer	Larger	Faster	Less stiff	Longer	Larger	Faster	Less stiff
F1	/t/**	/t/**		/t/**	/t/**	/t/**		/t/**				
F2		/t/**	/t/*									
M1												
M2												

References

[1] Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. *Speech Production and Speech Modelling*, Dordrecht: Springer Netherlands, 403–439.

[2] Turk, A. E. and Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics, 25(1)*, 25–41.

[3] Turk, A. E. and White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics, 27(2)*, 171–206.

[4] Heldner, M. and Strangert, E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics, 29(3)*, 329–361.

[5] Cambier-Langeveld, T. and Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics, 27(3)*, 255–280.

[6] Mücke, D. and Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics, 44(1)*, 47–61.

[7] Cho, T. (2011). Laboratory Phonology. *The Continuum Companion to Phonology*, N. C. Kula, B. Botma, and K. Nasukawa, (Eds.) London/New York: Continuum, 343–368.

[8] Tiede, M. (last used on 2/21/2019). MVIEW: software for visualization and analysis of concurrently recorded movement data. New Haven, CT: Haskins Laboratories.

Prosodic-structural Modulation of Vowel Nasalization in Mandarin Chinese

Hongmei Li[1], Sahyang Kim[2] & Taehong Cho[1]

[1]*Hanyang University (Korea),* [2]*Hongik University (Korea)*
lhm82cn@hanyang.ac.kr, sahyang@gmail.com, tcho@hanyang.ac.kr

Previous studies on coarticulatory vowel nasalization [5,7] have shown that prosodic structure fine-tunes the seemingly low-level coarticulatory process. For example, in both American English and Korean, focus-related prominence elongates N’s duration but induces V’s coarticulatory resistance to nasalization in NVC and CVN, pertaining to an enhancement of each phoneme’s paradigmatic contrast. Furthermore, in the *domain-initial* position (#NVC), N duration and V-nasalization decreases, which has been interpreted as a syntagmatic CV contrast enhancement. In *domain-final* position (CVN#), both languages showed coarticulatory vulnerability with more V-nasalization, indicating articulatory weakening towards the end of a phrase which may result in loosening the velic elevation gesture. The cross-linguistic comparisons, however, revealed some differences in the patterns of prosodically-conditioned V-nasalization in reference to language-specific features.

The current study extends this cross-linguistic evidence for prosodic structural modulation of coarticulatory V-nasalization to Mandarin Chinese. We investigate how V-nasalization is acoustically manifested as a function of prosodic strengthening factors such as boundary strength and prominence in Mandarin and discuss how the coarticulatory effect in Mandarin speaks to the language’s linguistic contrast systems. Mandarin employs distributional restrictions on the occurrence of consonants in the coda position, where only /n/ and /ŋ/ are allowed with no oral counterparts [1,6]. In addition, the nasals’ oral constriction is often incomplete or substantially lenited [2,6] and many Chinese dialects have lost nasal codas such that VN has become nasalized V or just V [6]. A specific question then arises as to how Mandarin V-nasalization manifests itself in such a nasal lenition (VN) context and how it is conditioned by prosodic strengthening factors.

Two #NV words (/ma/, /na/; #NVC with C[+oral] is not allowed in Mandarin) and four CVN# words (/pan/, /tan/, /paŋ, /taŋ/) were produced in various prosodic boundary (Phrase-initial/final vs Phrase-medial (=Wd-initial/final)) and prominence (Focused vs Unfocused) contexts by 16 native speakers (8F, 8M). The lexical tone was controlled to be Tone 1. The example test sentences are given in Table 1. Acoustic N duration and A1-P0 were measured. For NV, A1-P0 values were measured at 25%, 50% and 75% points in the vowel for relative timepoints, and at 25ms, 50ms, 75ms away from the nasal consonant for absolute timepoints. As the coda N was often produced without clear acoustic evidence for its oral constriction for CVN, the entire VN-duration was divided into 10 portions by an increment of 10%, and 9 points were taken from 10% point (near the end of VN) to 90% point (near the vowel onset). Six absolute timepoints between 25ms (near the end of VN) to 150ms (near the vowel onset) were also taken with an increment of 25ms.

The results are summarized in Figs 1 and 2. In both #NV and CVN#, N-duration increased under prominence, enhancing N’s nasality, while V-nasalization decreased, enhancing V’s orality, consistent with the pattern of paradigmatic contrast enhancement found in English and Korean [5,7]. As for the boundary effects, N-duration and V-nasalization decreased in the IP-initial position, increasing N’s consonantality and V’s orality thus enhancing the syntagmatic CV contrast, similar to the findings of English [5]. In IP-final position (CVN#), while there was no clear-cut acoustic distinction between V and N, the VN duration showed phrase-final lengthening. In the absolute measurement, V’s nasality increased when V was closer to the end of VN. However, crucially, the degree of V-nasalization was not larger phrase-finally in the relative measurement. The overall lack of boundary effects stands in sharp contrast with phrase-final V-nasalization patterns found in American English and Korean which showed more nasalization in the IP-final than IP-medial position. The results are discussed in terms of distributional restrictions in Mandarin Chinese that allows only nasals in the coda which are often lenited. In conclusion, the observed cross-linguistic generalizability and language-specificity support the general view that the low-level phonetic process operates across languages, but it is fine-tuned by the language's internal structure such as the language-specific phonological and prosodic structure (e.g., [3,4,5,7,8]).

Table 1. Examples of CVN in carrier sentences. Target words are underlined and italicized. Focused words are in bold.

	IP condition	Wd condition (IP-medial)
Focused	A: [tsuɔ3 pjɛn1 ʂi4 **pa1** ma] Is the left one **EIGHT**? B: [pu2 ʂi4]. IP [tsuɔ3 pjɛn1 ʂi4 ***pan1***]. IP [ʂi4 pu]? No. The left one is ***CLASS***, isn't it?	A: [li2 tsaɪ4 ʂi4 **pa1** ʂɑŋ4 mjɛn4 ma]? Is the pear on the top of IS **EIGHT** B: [pu2 ʂi4]. IP [li2 tsaɪ4 ʂi4 ***pan1*** ʂɑŋ4 mjɛn4]. No. The pear is on the top of IS ***CLASS***.
Unfocused	A: [**tsuɔ3** pjɛn1 ʂi4 pan1 ma] Is the **LEFT** one CLASS? B: [pu2 ʂi4]. IP [**joʊ4** pjɛn1 ʂi4 *pan1*]. IP [ʂi4 pu]? No. The **RIGHT** one is *CLASS*, isn't it?	A: [**tʰɑʊ2** tsaɪ4 ʂi4 pan1 ʂɑŋ4 mjɛn4 ma]? Is the **PEACH** on the top of IS CLASS? B: [pu2 ʂi4]. IP [**li2** tsaɪ4 ʂi4 *pan1* ʂɑŋ4 mjɛn4]. No. The **PEAR** is on the top of IS *CLASS*.

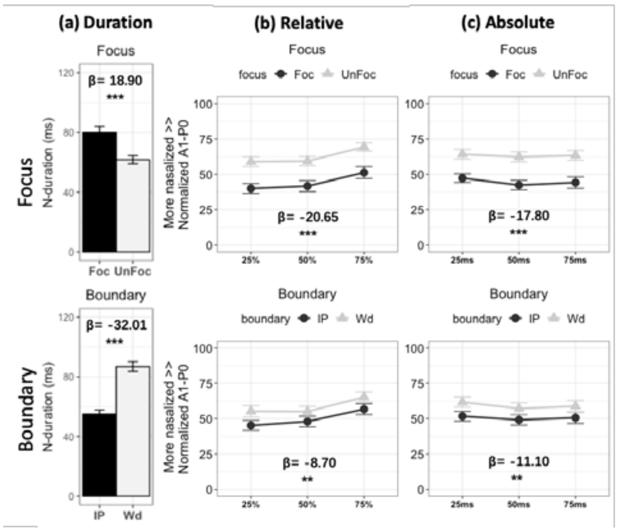


Fig. 1 N-duration and V-nasalization (A1-P0) in **#NV**

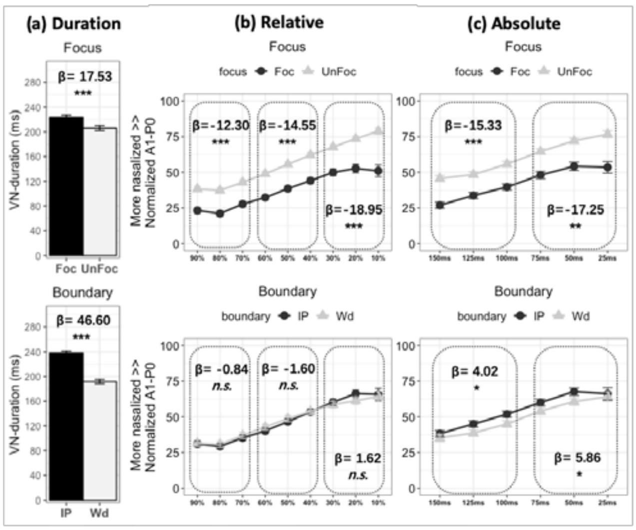


Fig. 2 VN-duration and V-nasalization (A1-P0) in **CVN#**

References

[1] Chao, Y.R. 1948. *Mandarin Primer*. Cambridge: Harvard University Press.

[2] Chen, M. Y. 2000. Acoustic analysis of simple vowels preceding a nasal in standard Chinese. *J. Phon.*, 28, 43-67.

[3] Cho, T. 2016. Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass*, 10(3), 120-141.

[4] Cho, T., James McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *J. Phon.*, 33(2) 121-157.

[5] Cho, T., Kim, D., Kim, S. 2017. Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *J. Phon.* 64, 71-89.

[6] Duanmu, San. 2000. *The Phonology of Standard Chinese*. New York: Oxford University Press.

[7] Jang, J., Kim, S., Cho, T. 2018. Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences. *J Acoust Soc Am.* 144, EL33-39.

[8] Kim, S., Kim, J., Cho, T. 2018. Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *J. Phon.*, 71, 65-80.

Cognitive skills and prominence production: Highlighting prominent elements in the speech of patients with Parkinson's disease

Tabea Thies[1,3], Doris Mücke[1], Anja Lowit[2], Elke Kalbe[3], Julia Steffen[3] & Michael T. Barbe[3]

[1]University of Cologne (Germany), [2]University of Strathclyde (United Kingdom),
[3]University Hospital Cologne (Germany)

tabea.thies@uni-koeln.de

Introduction: Patients with idiopathic parkinsonism suffer from a neurodegenerative disorder of the nervous system. Due to a progressive loss of dopaminergic cells in the substantia nigra patients with Parkinson's disease (PD) develop problems with motor and non-motor functions. On the motor level the dysfunctions defect the voluntary movements and lead to symptoms like: rigidity, resting tremor and bradykinesia [1]. On the level of cognition PD patients have problems with the executive functions, cognitive flexibility, working memory and control of attention [2]. Furthermore, the speech system gets affected which often leads to dysarthric speech. This hypokinetic dysarthria impacts the phonation, articulation and the respiratory system. The speech deficits include monoloudness, monopitch, reduced stress, imprecise articulation, variability of speech rate, disfluencies and voice tremor [3, 4, 5].

PD affects communication as well as other related functions such as cognition, but complex prosodic aspects such as focus marking are less well studied. Prominence marking in German requires changes in intonation and articulation [6]. Speakers use multiple cues in the phonetic domain to regulate prosodic marking [7], e.g. modulation of F0 and syllable duration. Furthermore, there is insufficient literature on the impact of cognition on these phonetic aspects of prominence marking. The aim of this study is to analyse how prosodic prominence correlates with motoric and cognitive abilities in patients with Parkinson's disease in the complex process of speech production.

Method: In the present study, we are investigating the prosodic marking strategies of PD patients and compare them to the productions of neurotypical speakers. Therefore, we investigate the production of target words in divergent focus structures, contrastive focus and background. We recorded 38 German speakers: 19 patients with idiopathic Parkinson, 13 males and 6 females, aged between 50 – 80, and 19 healthy aged and gender matched controls. For the patient group we assessed severity of the disorder, motoric activity level (UPDRS III, [8]) and speech intelligibility [9]. Further, all speakers were classified in terms of their level of cognition (TMT for executive functions and processing speed, [10], BTA for attention control [11], digit span for working memory [12]).

As speech material, we used a question-answer scenario presented on a computer screen to manipulate focal structure by means of contextualizing contexts. Nine target words were placed in either contrastive focus or background position in sentences such as <Die Fliege hat die grüne NAse berührt.> ("The fly has touched the green nose.") related to pictures on a computer screen. Target words were always disyllabic (CV.CV structure), containing one of the three long vowels /i:/, /a:/ or /o:/, in the accented syllable, such as <Nase> /na:z@/. In total, we recorded 1368 tokens (9 target words x 38 speaker x 2 focus structures x 2 adjectives). For acoustic measurements, we analysed the overall pitch range, the pitch height of rising pitch accents, syllable duration, vowel formants and intensity means.

Results: The results show that, in line with [13], PD patients with mild dysarthric symptoms can convey contrastive focus by increasing pitch, intensity and syllable duration. Furthermore, they adjust their articulatory movements in terms of formant changes and increase their vowel space in prominent positions. We did not find overall group differences between patients and controls comparing the phonetic parameters: pitch, syllable duration and intensity. However, patients have a smaller vowel space than the control participants. Regarding cognitive abilities, we found no differences in terms of attention control and working memory. But the groups differed significantly according to their processing speed and executive functioning.

Furthermore, we determined influencing factors affecting prominence marking in the patient's group like motoric and cognitive dysfunction. The more severe the executive dysfunction the more intensity and F0 get modulated. The motoric abilities influence the vowel articulation leading to a decrease in vowel space due to motoric impairment.

For the dynamic speech system this reflects abnormalities in the regulation mechanism of expressing prosodic prominence, constantly mediating between linguistic structure and the physical control system. There are some systems which are hyperarticulating (overmodulation of phonetic parameters) and some which are hypoarticulating (undermodulation of phonetic parameters) to produce a communicative output. We will discuss that the patients' system is out of balance, because there is too much effort in the glottal and subglottal system and less effort in the supraglottal system. This leads to an inefficient way of marking prominence. For clinical and therapeutic practice, we assume beneficial effects of cognitive training to preserve communicative function.

References

[1] Hornykiewicz, O. (1998). Biochemical aspects of Parkinson's disease. *Neurology, 51*(2 Suppl 2), 2-9.

[2] Kalbe, E., Rehberg, S. P., Heber, I., Kronenbuerger, M., Schulz, J. B., Storch, A., et al. (2016). Subtypes of mild cognitive impairment in patients with Parkinson's disease: evidence from the LANDSCAPE study. *J Neurol Neurosurg Psychiatry, 87*, 1099-1105.

[3] Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural neurology*, *11*(3), 131-137.

[4] Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research*, *12*(2), 246-269.

[5] Duffy, J. R. (2013). *Motor Speech Disorders - Substrates, Differential Diagnosis, and Management*. St. Louis, Elsevier, Mosby.

[6] Mücke, D. & Grice, M. (2014). The effect of focus marking on supra-laryngeal articulation – is it mediated by accentuation? *Journal of Phonetics, 44,* 47-61.

[7] Cho, T. (2006). Manifestation of Prosodic Structure in Articulation: Evidence from Lip Kinematics in English. *Laboratory Phonology, 8,* 519-548.

[8] Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., ... & Dubois, B. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement disorders*, *23*(15), 2129-2170.

[9] Dobinson, C. (2007). *An investigation into the use of computers for conducting home practice of speech exercises for people with dysarthria.* Unpublished PhD Thesis. Newcastle University, Newcastle upon Tyne.

[10] Tischler, L., & Petermann, F. (2010). Trail making test (TMT). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 58*(1), 79-81.

[11] Schretlen, D., Bobholz, J. H., & Brandt, J. (1996). Development and psychometric properties of the Brief Test of Attention. *The Clinical Neuropsychologist*, *10*(1), 80-89.

[12] Wechsler, D. (1997). *Administration and scoring manual.* San Antonio, TX: The Psychological Corporation.

[13] Tykalova, T., Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2014). Acoustic investigation of stress patterns in Parkinson's disease. *Journal of Voice*, *28*(1), 129-136.

Assessing tonal specifications through simulation and classification : The case of post-wh accent in Japanese

Shigeto Kawahara[1], Jason Shaw[2] & Shinichiro Ishihara[3]

[1]*Keio University (Japan),* [2]*Yale University (United States),* [3]*Lund University (Sweden)*
kawahara@icl.keio.ac.jp, jason.shaw@yale.edu, shinichiro.ishihara@ostas.lu.se

Background: Given a phonetic trajectory, be it articulatory or acoustic, deciphering whether a segment has a phonetic target in that trajectory, or is entirely targetless, is not an easy task (e.g. [1, 2]). [3] developed a general computational toolkit to address this issue, which allows us to assess the presence of a phonetic target on a token-by-token basis. [3] analyzed tongue dorsum trajectories, addressing whether devoiced [u] in Japanese retains its articulatory target; they found that some tokens retain a full target, whereas some tokens are entirely targetless. The current paper applies this computational toolkit to intonational patterns.

The empirical focus of the current study is realization of lexical LHL tonal accent following wh-elements in Japanese. [4] argues that lexical accent after wh-elements in Japanese is “eradicated,” and that this domain of eradication continues up to the question particle that licenses that wh-element. This claim led [5] to propose a general theory of wh-movement in natural languages. Japanese allows wh-elements to stay in-situ because wh-elements and their licenser can be prosodically grouped together; since English does not have this prosodic means, it has to overtly move its wh-elements. Other studies however cast doubt on the claim that lexical accent after wh-elements is entirely deleted (e.g. [6, 7]). [6] shows that degrees of rise of lexical accent are in fact *reduced* after wh-elements, but the averaged F0 contours seem to show that there are some visible rises on the lexical items at issue.

Method: The current study thus reanalyzes a subset of the data obtained by [6]. We compare two sentence structures in (1) and (2).

(1) Control sentences: Word_1 $\text{Word}_{2[-\text{wh}]}$ Word_3 Word_4 Verb
(2) Test sentences: Word_1 $\text{Word}_{2[+\text{wh}]}$ Word_3 Word_4 Verb

The sentences in (1) serve as the control sentences in which the lexical accent of both Word_3 and Word_4 is realized (i.e. full target). What we are interested in is whether Word_3 and Word_4 in (2) maintain some traces of lexical accent or whether the accent is completely eradicated. There were six types of sentences for both (1) and (2); nine native speakers of Tokyo Japanese repeated those sentences, together with other sentences, twice. The intonational contours of Word_3 and Word_4, delimited by their L tones, were extracted using YAAPT. These contours were decomposed into a set of DCT coefficients. In addition to the control sentences, we simulated linear trajectories between the two L tones, and injected the same degree of variability as those observed in the sentences in (2)—these contours simulate how targetless intonational contours (i.e., tone eradication) would be realized given naturalistic variability. Finally, a Bayesian classifier was trained, and for each tonal contour for the sentences in (2), it assigned a posterior probability of belonging to the linear interpolation category.

Results: Figure 1 shows the posterior probability of linear interpolation for Word_3 and Word_4 For Word_3, many speakers (Speakers 1, 2, 4, 5, 7, 8, 9) show at least some tokens that have a high posterior probability of linear interpolation. These tokens support the view expressed by [4]. However, a Speakers 6, 7, 8, and 9, show a large number of tokens that are better classified as belonging to the full target category; these tokens show no trace of reduction. We finally observe those tokens whose posterior probabilities are in the middle range (Speakers 2, 4, 5, 7); these tokens are phonetically reduced. For most speakers (all except 3), within-speaker variability is evident. For Word_4, most speakers (all but Speakers 1 and 6) show tokens of high linear interpolation probability. Speakers 1, 3, 5, 6, and 8, also produced tokens with full tonal targets and there are many tokens as well that are phonetically reduced. Again, just like Word_3, we observe both inter- and intra- speaker variability.

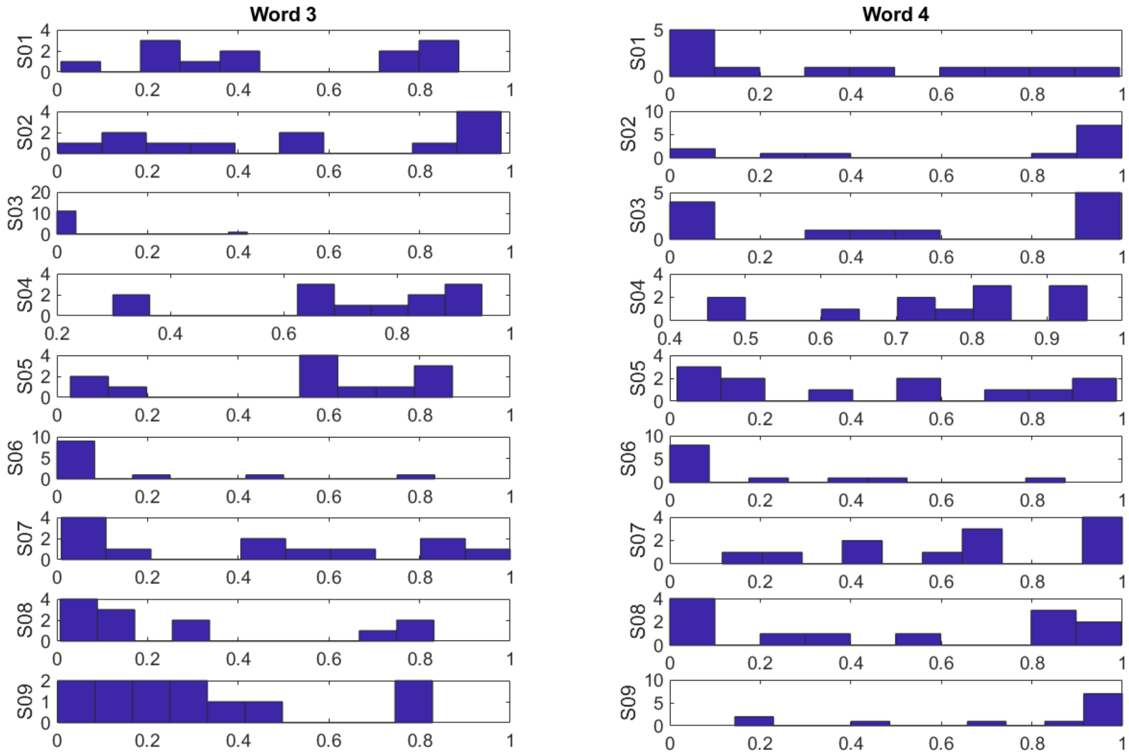


Figure 1: Posterior probabilities of linear interpolation. Left = Word$_3$; Right = Word$_4$.

Discussion: Our analysis shows that some speakers show some tokens that are best characterized as eradicated; at the same time, however, no speakers consistently show eradication. These results imply that [4]'s observation was correct at some level of analyses, but the current results pose an interesting challenge to [5]. If eradication is what allows Japanese wh-elements to stay in-situ, how come those tokens without eradication show no wh-movement? How come Speaker 6, who almost always showed high probability of full target, does not move wh-elements?

Next, we discuss some advantages of our analytical toolkit. First the toolkit allows us to assess the presence of intonational targets on a token-by-token basis. Most previous studies analyze averaged contours, but analyzing only averaged contours can be misleading. Take the case of Word$_4$ for Speakers 3 and 4, for example. In our analysis, Speaker 4 shows reduction for all tokens; Speaker 3 on the other hand shows a bimodal distribution of full targets and eradication. If we were only looking at averages, we would have erroneously concluded that both speakers show reduction. This highlights the importance of analyzing each token separately. Second, [7] addressed the question of whether "deaccented" phrases and unaccented phrases are different or not, by fitting linear regression lines (see also [2]), and found that the regression lines are different between deaccented phrases and unaccented phrases. One distinct advantage of our approach is that it does not have to assume linearity, as the first step of our analysis decomposes trajectories into a sum of cosine waves. Token-by-token analysis offers great promise for the study of intonational variation.

References

[1] Cohn, A. (1993). Nasalisation in English: phonology or phonetics. *Phonology* 10. 43–81.
[2] Pierrehumbert, J. & M. Beckman. (1988). *Japanese tone structure*. MIT Press.
[3] Shaw, J. & S. Kawahara. (2018). Assessing surface phonological specification through simulation and classification of phonetic trajectories. *Phonology* 35. 481–522.
[4] Deguchi, M. & Kitagawa, Y. (2002). Prosody and wh-question. *Proceedings of NELS* 32: 73-92.
[5] Richard, R. (2010). *Uttering Trees*. MIT Press.
[6] Ishihara, S. (2010). Focus prosody in Tokyo Japanese wh-questions with lexically unaccented wh-phrases. *Proceedings of ICPhS* 946-949.
[7] Maekawa, K. (1994). Is there 'dephrasing' of the accentual phrase in Japanese? *Ohio State University Working Papers in Linguistics* 44. 146-156.

Do Individual Differences in Working Memory Capacity Predict Cross-Speaker Variation in Planning Scope? Some Prosodic Tests

Jason Bishop[1,2]

[1]*CUNY-College of Staten Island (USA),* [2]*CUNY-Graduate Center (USA)*
jbishop@gc.cuny.edu

While several phonetic and phonological patterns indicate that speech production planning unfolds in prosodic phrase-sized chunks [1] rather than one- or two-word sequences [2], research in both phonetic science and psycholinguistics has begun to ask how flexible this planning is, and to what extent it may reflect demands both external and internal to speakers [3]. The present study explored the extent to which speakers' planning may vary in relation to the latter, exploring individual differences in working memory capacity (WMC) as a source of speaker-internal—and thus speaker-specific—constraints on planning. In particular, we tested the hypothesis that speakers with higher WMC engage in more extensive/longer-range planning, exploring two prosodic variables.

The first was speakers' average phrase length, counted in syllables, over a sample of connected speech. We predicted that speakers with higher WMC, if they engage in longer-range planning, would tend to produce longer prosodic phrases (intonationall-defined). Notably, this question has a parallel in the literature on implicit prosody (i.e., prosody generated internally during silent reading), where sentence processing tasks have suggested that the size of readers' implicit prosodic phrases varies along with their WMC [4]. To our knowledge, however, the question of whether WMC predicts variation in the length of speakers' overtly-produced phrases has not been explored.

The second variable was silent pause duration. Pause durations are known to increase as the length of an upcoming phrase increases [5,6] and so are widely assumed to, in part, reflect speakers' planning. Interestingly, however, [7] demonstrated that pause duration is not only predicted by the length of an upcoming phrase, but the phrase's internal structure. In particular, and somewhat counterintuitively, speakers tend to produce shorter pauses when an upcoming phrase is structurally complex. For example, on average, speakers of English and German will produce shorter pauses before an Intonational phrase (IP) that contains two or more smaller intermediate phrases (ip) than before an IP that contains just one ip. [7,8] suggested this reflects the options afforded by complex phrases; complex phrases can be approached more incrementally, while more monolithic structures cannot (see **Fig.1**). The question we asked was whether WMC predicts variation in the extent to which speakers avail themselves to this more incremental planning option. It was predicted that, because speakers with higher WMC should be more likely to plan larger chunks (and thus less likely to engage in the more incremental planning option), their pause durations should be less affected by the complexity of an upcoming IP. Relative to speakers with lower WMC, then, speakers with higher WMC should have longer, not shorter, pause durations before complex IPs.

The hypotheses relating WMC to these two prosodic variables were tested on a corpus of read speech collected in the context of another study [9]. The corpus was based on a production study with a large group of native English speakers from the US (N=100) who completed a standard measure of WMC and who read the same 160-word passage aloud (in a sound-attenuated booth, using a Shure SM10A headworn mic, digitally recorded at 41kHz). The phrase structure of the produced passages was identified using ToBI conventions for Mainstream American English [10]; the length (in syllables) of all speakers' fluent ips and IPs were then counted, and the durations of all pauses preceding fluent IPs were measured. Mixed-effects linear regression was then used to model phrase length and pause duration, with the goal of identifying any effect of WMC that could be related to speakers' planning in the ways described above.

Results indicated the following. First, WMC was to some extent predictive of speakers' phrase lengths; speakers with higher WMC tended to produce longer IPs and (IP-medial) ips (see **Fig.2**), although the effect was only significant when phrase levels were collapsed (*est=-.0417, SE=.0136, t=-2.99 p<.01*). One interpretation of this finding (anticipated in [8]) is that speakers can manage their WMC-related limitations in different ways, with some speakers adjusting the lengths of their ips, others their IPs, and yet others a combination of both. Second, WMC was also found to predict the sensitivity of speakers' pause durations to the complexity of an upcoming IP; when the length of an upcoming IP was held at its mean, there was no main effect of its complexity on pause durations (*est=.006, SE=.001, t=.355,p>.1*). Instead, the effect of complexity depended on speakers' WMC; longer pause durations preceded complex phrases, but only for speakers with higher WMC (*est=.040, SE=.027, t=2.03, p<.05*), consistent with more advanced—and less incremental—planning by these speakers. We discuss the implications of these findings for speech production planning; for models of how prosodic structure interacts with other components of grammar and language production; and for the use of individual differences approaches in the investigation of these issues.

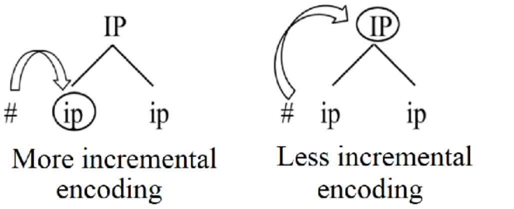


Fig.1 Planning options in complex prosodic structures (based on [8]); speakers can plan at lower nodes or (less incrementally) at higher ones. Planning strategies will be reflected in pause durations (indicated by '#' in the figure).

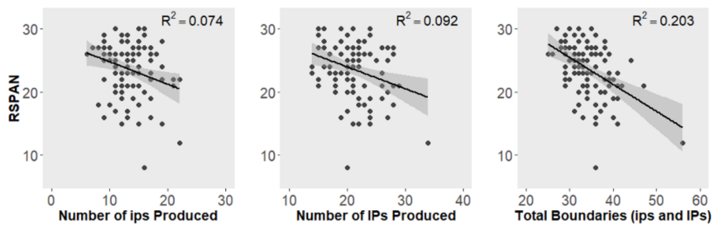


Fig.2 Relationship between speakers' WMC (as measured by reading spans) and the number of phrase breaks produced in the recorded passage. The strongest relation is between speakers' WMC and the total number of boundaries, collapsing phrase levels (ip and IP).

References

[1] Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics,*101, 112–156.

[2] Levelt, W.J. (1989). *Speaking: From intention to articulation.* Cambridge: MIT Press.

[3] Wagner, V., Jescheniak, J., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 423−440.

[4] Swets, B., Desmet, T., Hambrick, D., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: a psychometric approach. *J. Experiment Psychology: General*, 136, 64-81.

[5] Ferreira, F. (1993). Creation of prosody during sentence production. *Psycholinguistic Review*, *100,* 233–253.

[6] Whalen, D., & J. Kinsella-Shaw (1997). Exploring the relation of inspiration duration to utterance duration. *Phonetica* 54, 138–152.

[7] Krivokapić, J. (2007). The planning, production, and perception of prosodic structure. Ph.D. dissertation, University of Southern California.

[8] Krivokapić, J. (2012). Prosodic planning in speech production. In S. Fuchs, M. Weihrich, D. Pape, & P. Perrier (eds.): *Speech planning and dynamics*. Peter Lang. pp. 157–190.

[9] Bishop, J. (resubmitted). Exploring the similarity between implicit and explicit prosody: prosodic phrasing and individual differences. Ms, City University of New York.

[10] Beckmann, M., & Ayers-Elam, G. (1997). Guidelines for ToBI labeling (Version 3). Unpublished ms. Ohio State University.

Distal rhythmic structure and speech rate in perception: exploring prosodic and rate-dependent influences

Jeremy Steffman

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

Background: The role of prosody in modulating listeners' perception of phonetic cues has been a recent topic of interest in the literature. Work investigating these questions has focused on adjacent, or proximal prosodic effects (e.g. boundary marking) [1,2]. The present study explores how *distal*, non-adjacent rhythmic patterns influence listeners' processing of phonetic (durational) cues in speech, as informed by the documented importance of rhythmic/metrical structure in word segmentation/lexical processing [3,4].

Listeners are argued to project a prosodic structure in segmenting speech based on the rhythmic properties of a distal preceding context, grouping ambiguous strings of sounds into words based on patterns of alternating duration/pitch [4]. These findings are couched in the *perceptual grouping hypothesis* (PGH), which predicts that perceptual grouping of alternating patterns will affect listeners' expectations about incoming information in the speech signal. The present study extends the PGH to test if grouping influences perception of vowel duration as a cue to coda voicing [5]. Predictions based on rhythmic grouping are compared with expected durational contrast effects [6] (outlined below). Three experiments tested rhythmic effects in perception by manipulating the rhythmic structure and rate of a series of CV syllables preceding a target sound from a "coat" ~ "code" continuum which varied only in vowel duration (90-150ms, 15ms steps). The target was categorized in a 2AFC task. Participants were native speakers of American English.

Experiments 1 & 2: In Experiment 1 (n = 32), the crucial manipulation was whether the precursor formed a sequence of durational trochees (long-short), or iambs (short-long). Preceding syllables were CV ([t^hɑ]), and were identical except for duration of the vowel. Three trochaic or iambic feet preceded a final foot in which the target was grouped with either a long syllable, forming a potential trochee (the TROCHEE condition), or a short syllable, forming a potential iamb (the IAMB condition; Fig. 1). Following the PGH, if listeners group the target as the second syllable in a foot based on preceding durational alternations, expectations about the duration of the target vowel might change based on whether it was the implied second syllable in a trochee (where it would be expected to be shorter), or an iamb (where it would be expected to be longer). An expectation of relative lengthening in the IAMB condition predicts that "code" responses *decrease* in the *IAMB condition*, reflecting the influence of distal rhythmic patterns. Proximal durational contrast effects [6] predict an opposite shift in categorization. Given that a relatively *longer* syllable precedes the target in the TROCHEE condition, categorization would be expected to shift to *longer* required vowel durations for a "code" response (*decreasing* "code" responses in the *TROCHEE condition*). As these two effects make opposite predictions, this design directly tests whether distal rhythmic structure or proximal duration will influence perception. Results are assessed by mixed-effects logistic regression (by-participant random intercepts, maximal random slopes). **Results:** A main effect of rhythm was found whereby the IAMB condition significantly *decreased* "code" responses ($p < 0.01$; Fig. 2). This result supports predictions based on the PGH, suggesting that listener *expectations* about duration based on rhythmic grouping influence the perception of vowel duration as a cue to voicing, in defiance of proximal contrast effects. These results thus provide novel insight into how rhythmic/prosodic context affects listeners' processing of durational cues. To test if proximal contrast effects are indeed present, Experiment 2 (n = 32) used *only the immediately preceding syllable* from the stimuli in Experiment 1 (boxed region in Fig. 1) and blocked stimuli by condition; it was otherwise identical. Results show that a preceding longer precursor (the TROCHEE condition in Experiment 1) significantly *decreased* "code" responses ($p < 0.05$; Fig. 3), confirming proximal contrast effects obtain in the absence of distal context.

Experiment 3: Experiment 3 (n = 32) explored how changing the *rate of repetition* of rhythmic structure influenced perception. Stimuli from Experiment 1 were manipulated to create a FAST condition (full precursor linearly compressed by 20%) and a SLOW condition (precursor linearly expanded by 20%); resulting in four conditions in a 2x2 design (FAST TROCHEE, SLOW TROCHEE, etc.). The experiment was otherwise identical to Experiment 1. **Results:** A main effect of rate was found whereby FAST rate increased “code” responses ($p < 0.05$; not shown). This is expected based on previous findings in the speech rate literature (e.g. [7,8]). A main effect of rhythm was also found, whereby the TROCHEE condition significantly *decreased* “code” responses ($p < 0.01$; see Fig. 4). Perhaps surprisingly, this result matches the proximal contrast effects observed in Experiment 2. Because the *entire* precursor was temporally expanded, these rate manipulations include expansion and compression of the syllable precedingly the target and therefore also exaggerate *proximal* differences in duration. This result therefore highlights that rhythmic effects compete with proximal contrast effects such that they disappear when proximal differences are larger in magnitude, and rate varies. An interaction between rhythm and rate was also found ($p < 0.05$). Post hoc testing of the interaction finds a significant effect of rate only in the TROCHEE condition ($p < 0.01$), and *no* effect in the IAMB condition (cf. Figs. 5 & 6). Given that all participants were speakers of American English, this interaction raises an intriguing possibility: that listeners more readily entrain to changes in rate when the rhythmic structure of speech more closely matches language patterns, given that English is largely trochaic, and listeners have been shown to be sensitive to this sort of metrical structure in other tasks [3]. This prediction is currently being tested in a follow-up experiment. Crucially, Exp. 3 shows that in the presence of increased variability in rate, the rhythmic grouping effects seen in Exp. 1 are not observed.

Summary: These experiments find an intricate interplay between durational contrast effects and rhythmic context in perception. Extending these results will therefore better our understanding of distal prosodic cues in perception and their interface with other perceptual processes.

Fig.1 Waveforms are bracketed according to listeners’ hypothesized grouping. The longer precursor syllable is bolded. The gray box highlights proximal context.

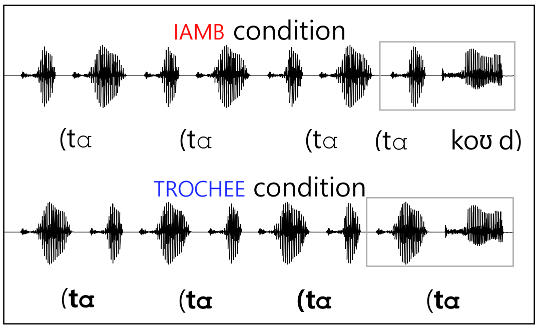


Fig.2-6 Categorization is plotted split by condition. Target vowel duration is on the *x* axis, the *y* axis shows the proportion of “code” responses. Parenthetical (short)/(long) refer to the syllable immediately

Fig.2 Exp. 1

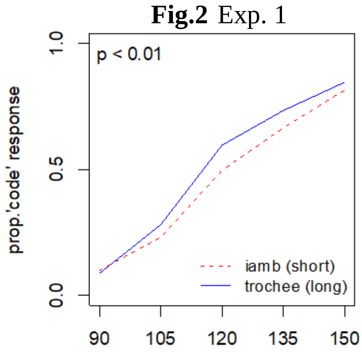


Fig.3 Exp. 2

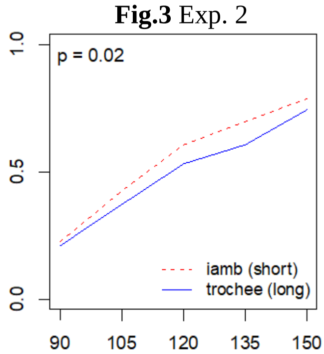


Fig.4 Exp. 3

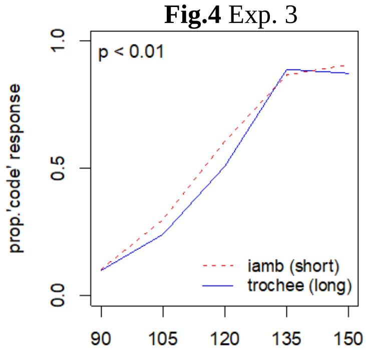


Fig.5 Exp. 3

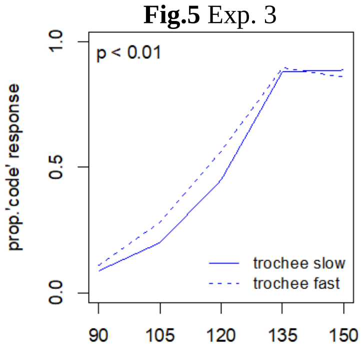
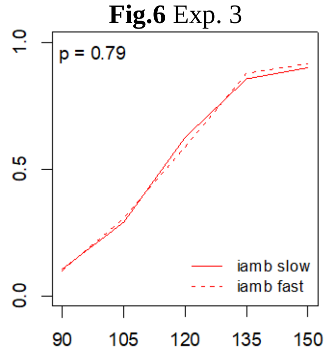


Fig.6 Exp. 3



References

[1] Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *JPhon*, *54*, 68–79.
[2] Steffman, J. (2019). Intonational structure mediates speech rate normalization in the perception of segmental categories. *Jphon*, *74*, 114–129.
[3] Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psych: Human percept. and performance*, *14*(1), 113.
[4] Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *JML.*, *59*(3), 294–311.
[5] 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.
[6] Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *JASA*, *85*(5), 2154–2164.
[7] Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attn., Percept. & Psychophysics*, *79*(1), 333–343.
[8] Reinisch, E. (2016). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attn., Percept. & Psychophysics*, *78*(4), 1203–1217.

Oral Presentations

(Day 2, Special Session)

Prosodic processing in sentences with "only" in L1 and L2 English: An ERP study

Rachida Ganga[1], Haoyan Ge[2], Marijn Struiksma[1], Virginia Yip[3] & Aoju Chen[1]

[1]*Utrecht University (the Netherlands),* [2]*Open University of Hong Kong,* [3]*Chinese University of Hong Kong (China)*
r.ganga@uu.nl, hge@ouhk.edu.hk, m.struiksma@uu.nl, vcymatthews@cuhk.edu.hk, aoju.chen@uu.nl

Speakers of West Germanic languages use focus particles such as "only" to signal upcoming contrast but can differ in the exact manner. For example, English speakers prefer placing "only" adjacent to the verb and indicate the locus of contrast via accentuation (e.g. I have only carried the bag vs. I have only carried the bag) [1]. In contrast, Dutch speakers prefer placing "only" ("*alleen*") adjacent to the contrastive, accented word [2]. Such cross-linguistic differences imply that Dutch and English listeners may use different processing strategies in their brain in sentences with "only" in their native language (L1). It has been shown that Dutch listeners expect accentuation immediately following "*alleen*" in online processing (early positivity at 100-200ms, followed by an "accent positivity") [3]. When they hear accentuation further away from "*alleen*", expectation (Expectancy Negativity) and sentence reanalysis (P600) occur. Furthermore, linguistic context speeds up the recognition of accentuation (shorter latency of "accent positivity") [4]. However, nothing is known on the brain's response to sentences with "only" in L1 English listeners and Dutch learners of English (L2 English). The present study aims to find out whether L1 and L2 English listeners exhibit different expectancy patterns of accentuation when processing English sentences with "only". Dutch listeners may experience L1 influence when processing L2 English prosody [5] or use learner-specific approaches that do not resemble L1 Dutch or English processing [6]. We hypothesize that L1 and L2 English listeners process accentuation immediately following "only" similarly but accentuation not adjacent to "only" differently, and that presence of context will enhance these differences.

Methods: Advanced Dutch learners of English (n=33, 14m) and native English listeners (n=8, 2m, more being tested) listened to four types of English stories (60 per type), differing in the presence/absence of context and accentuation on verbs/objects (Table 1). ERPs were recorded from the onset of verbs and objects (t=0) to 1000ms afterwards. We conducted mixed effect analyses using R (lme4) for verbs and objects separately in three time windows, depending on the respective word length (Figure 1), with ACCENT (verb, object), CONTEXT (absent, present), ANTERIORITY (front, central, back), LATERALIZATION (right, middle, left), and GROUP (Dutch, English) as fixed factors and PARTICIPANT as the random factor. Pairwise comparisons were done with Bonferroni correction.

Results – Verbs: In 100-200ms after verb onset, we found significant interactions of ACCENT X ANTERIORITY and ACCENT X CONTEXT. Pairwise comparisons revealed that accentuation was recognized early (more positivity) than deaccentuation in frontocentral brain regions and context elicited an expectancy response (more negativity) only when the verbs were deaccented. In 200-390ms, we found a significant interaction of ACCENT X ANTERIORITY X GROUP. Pairwise comparisons revealed that the acoustic properties of accentuation were processed with more positivity across the whole scalp by L2 listeners but not by L1 listeners.

Results – Objects: In contrast to verbs, we found neither a significant main effect of ACCENT nor significant interactions involving ACCENT in 100-200ms and in 200-390ms after object onset, suggesting that early processing was not guided by contrast or prosody. In 500-900ms, we found significant interactions of ACCENT X ANTERIORITY and CONTEXT X ACCENT X GROUP. Pairwise comparisons revealed that accentuation made sentence reanalysis more effortful (more negativity) in frontocentral regions than deaccentuation for L2 listeners, independent of context, but for L1 listeners only when the context was absent.

Discussion & Conclusion: In spite of showing a stronger response to the acoustic properties of accentuation, L2 English listeners processed accentuation in the verb similarly to L1 English listeners. Both groups were similar to L1 Dutch listeners in [7]. This can be explained by the fact that it is common to accent the verb immediately following "only" in both Dutch and English. Interestingly, accentuation in the object made sentence reanalysis more effortful for L2 listener with or without context, but for L1 listeners only in the absence of context. These findings can be explained by influence from Dutch in processing L2 English sentences, in particular, the expectation of an accented word right after “only” and the markedness of sentences with “only” and with the object accented without any context. To conclude, our hypotheses are largely borne out by the initial results. L2 English processing is influenced by processing strategies in L1 Dutch even in advanced Dutch learners of English. Additional data from L1 English listeners will help to verify our current conclusions.

Table 1. Examples of experimental stimuli. Pitch accents are represented in bold.

Context sentences	Target sentences	
The dinosaur has a pumpkin and a bucket. He was going to throw them and kick them. Then he changed his mind.	(A) The dinosaur is only **throwing** the bucket.	(B) The dinosaur is only throwing the **bucket**.
-	(C) The dinosaur is only **throwing** the bucket.	(D) The dinosaur is only throwing the **bucket**.

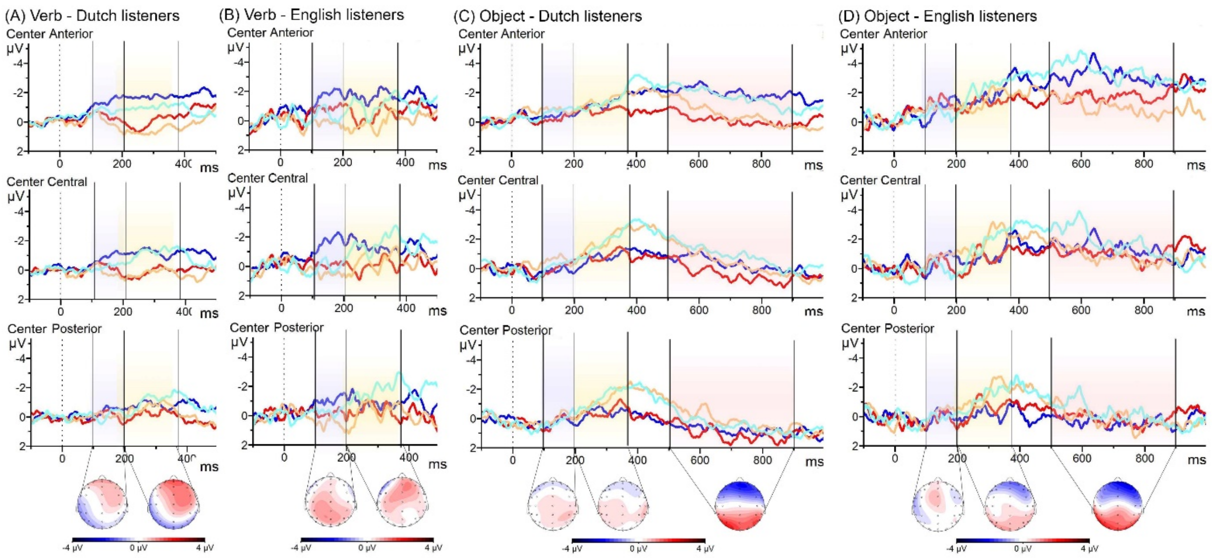


Figure 1. Grand-average ERPs for verbs in L2 (A) and L1 English listeners (B), and objects in L2 (C) and L1 English listeners (D) from word onset (t=0). Baseline correction was performed in the 100ms prior to word onset. Conditions: accented verbs with context (red) and without context (orange), and accented objects with context (dark blue) and without context (light blue). Vertical lines indicate time windows. Topographies reflect accent effects (accented–deaccented). The time window of 500ms till 900ms in verbs was not analysed as this overlapped with the onset of words after the verbs.

References

[1] Bouma, G., Hendriks, P., & Hoeksema, J. (2007). Focus Particles Inside Prepositional Phrases: A Comparison of Dutch, English, and German. *J Comp Ger Linguist,10*, 1–24.

[2] Foolen, A., van Gerrevink, R., Hogeweg, L., & Prawiro-Atmodjo, P. (2009). The placement of focus particles in Dutch. *Linguistics in Netherlands, 26*(1), 51–63.

[3] Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. J. (2010). Focus Particles and Prosody Processing in Dutch: Evidence from ERPs. In: *Speech Prosody*. Chicago;. p. 1–4.

[4] Dimitrova, D. V., Stowe, L. A., & Hoeks, J. C. J. (2015). When Correction Turns Positive: Processing Corrective Prosody in Dutch. *PLoS One, 10*(5): e0126299.

[5] Ge, H., Mulders, I., Yip, V., & Chen, A. (In prep). Processing of sentences with “only” in Dutch-speaking second language learners of English: An eye tracking study.

[6] Chen, A., & Lai, V. (2012). Comb or Coat: The Role of Intonation in Online Reference Resolution in a Second Language. *Sound Sounds Utr,* 57–68.

[7] Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. J. (2012). Less Is Not More: Neural Responses to Missing and Superfluous Accents in Context. *J Cogn Neurosci, 24*, 2400–18.

Musicality influences the processing of accentuation in sentences with *only* by L2 English learners

Nelleke Jansen, Rachida Ganga, Marijn Struiksma & Aoju Chen

Utrecht University, The Netherlands
{n.g.jansen} {r.ganga} {m.struiksma} {aoju.chen}@uu.nl

Research on the relationship between music and language has revealed advantages of musicality in the perception of acoustic variation in prosodic parameters in a foreign or second language (L2), e.g. pitch [1] and duration [2]. In this study, we investigate whether these benefits in speech perception extend to higher order processing in L2 prosody such as the processing of the prosodic cues to information structure. Specifically, we address the following question: **Does musicality facilitate the processing of accentuation in English sentences with *only* by Dutch learners of English?** Sentences containing the focus particle *only* need accentuation to determine the locus of contrast (e.g. I have only CARRIED the bag vs. I have only carried the BAG), and thus require the processing of the accentuation-to-contrast mapping. These sentences may pose a challenge for Dutch learners of English, because the Dutch equivalent of *only* ('alleen') typically directly precedes the word in contrastive focus [3], whereas in English, *only* usually precedes the verb, regardless of which word is focal [4]. Dutch listeners have been shown to expect an accent on the word directly following *alleen* [5]. They may also expect the accent to fall on the verb following *only* in English. Given that musicality is related to heightened sensitivity to acoustic variation in L2 prosody, we hypothesised that learners with higher musicality are more efficient in higher order prosodic processing in L2 and demonstrate less influence of L1 processing patterns.

Method: Thirty-three advanced Dutch learners of English participated in this study. In a 64-channel event-related potentials (ERP) experiment, the participants were instructed to listen to short trials recorded by a native speaker of British English. Each trial contained a sentence with *only* in which either the verb or the object was accented; the sentence with *only* was preceded by a context in half of the trials, creating four conditions (Table 1). ERPs were recorded from the word onset to 1000ms after the word onset during the word *only,* the verb, and the object noun. Following the ERP experiment, the participants completed a short version of the Profile of Music Perception Skills (PROMS) test [6]. The PROMS scores were used as a measure of musicality, which refers to one's musical perceptual ability.

Analysis: ERPs in three time windows were analysed using Mixed Effect Modelling (lme4 in R): 100-200ms (analysed for *only*, verbs, and objects), 200-390ms (analysed for verbs and objects), and 500-900ms (analysed for objects). The average ERPs per condition per area were used. CONTEXT (present vs. absent), ACCENT (verb vs. object), LATERALISATION (right, middle, or left) and ANTERIORITY (front, central, or back) and MUSICALITY were included as fixed factors and STIMULUS-LIST and PARTICIPANT as random intercepts.

Results: We found a significant three-way interaction of CONTEXT x ACCENT x MUSICALITY for *only* in the time window 100-200ms ($p < .001$). Subsequent analyses showed that in the accented-object condition with context, lower PROMS scores led to more negativity than higher PROMS scores (Figure 1A), suggesting more processing effort in the participants with lower musicality. Acoustic analysis of our stimuli revealed that the words before *only* had a significantly higher pitch and *only* had a longer duration in the accented-object condition than in the accented-verb condition. Our participants might thus have anticipated the locus of accentuation upon hearing *only* [7]. The presence of context appears to enhance the L1-based expectancy of accentuation in the verb, making the anticipated lack of accentuation in the verb more effortful to process for the participants with lower musicality. The higher processing effort in participants with lower musicality thus suggests that they are less flexible in adjusting their expectations when listening to L2 prosody than listeners with higher musicality.

The modelling on the ERPs in the verb did not yield a main effect of MUSICALITY or a significant interaction involving MUSICALITY, despite the fact that the average ERPs suggested a difference between the participants with lower musicality and those with higher musicality. Our sample size was probably too small to observe a relationship between variation in musicality scores and ERPs in the verb region. In subsequent modelling, we re-coded MUSICALITY as a categorical variable by dividing the participants into two groups on the basis of the median value of the PROMS scores (i.e. 40), low musicality (lower than 40, N=16) and high musicality (40 or higher, N=17). We found a significant interaction of CONTEXT x ACCENT for the verb in the time-window 200-390ms ($p < .001$) for the high-musicality group, but not for the low-musicality group. Further analyses showed that in the accented-object condition without context, the high-musicality group exhibited less negativity than the low-musicality group (Figure 1B). The negativity found in the low-musicality group suggests a continued expectancy of accentuation on the verb. Here, the effect of CONTEXT is different than in *only*; the absence of context makes the lack of accentuation in the verb difficult to process for the low-musicality group, possibly because it is strange to use *only* in an isolated sentence and even stranger to then deaccent the verb. These differences in expectancies provide further evidence that listeners with high musicality are more flexible when processing L2 prosody.

Conclusion: Our study has provided the first evidence for an influence of musical abilities on high order prosodic processing in L2, i.e. the processing of the prosodic cues to information structure. Higher musicality is related to better online adjustment of expectations for accent placement while listening to sentences with *only*. The heightened sensitivity to acoustic variation in pitch and duration that has been related to higher musicality may play a role in the more efficient higher order prosodic processing in L2.

Table 1. Examples of experimental stimuli. Pitch accents are represented by capitals.

Context sentences	Target sentences	
The dinosaur has a pumpkin and a bucket. He was going to throw them and carry them. Then he changed his mind.	(A) The dinosaur is only THROWING the bucket.	(B) The dinosaur is only throwing the BUCKET.
-	(C) The dinosaur is only THROWING the bucket.	(D) The dinosaur is only throwing the BUCKET.

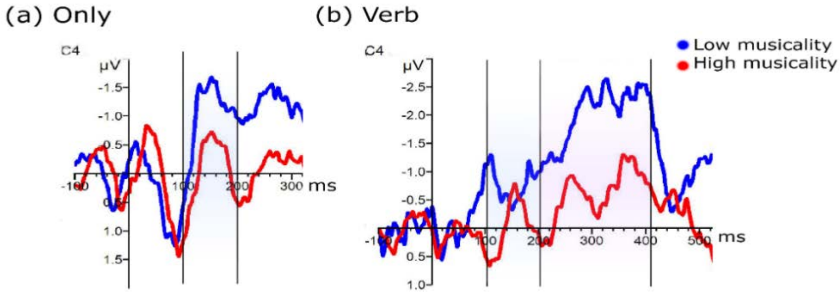


Fig. 1 Grand-average ERPs for the low- and high-musicality group from the word onset (t = 0) for *only* (A) in the condition accented objects with context, and for verbs (B) in the condition accented objects without context.

References

[1] Marques, C., Moreno, S., Luís Castro, S., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453-1463.
[2] Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: a cross-linguistic study. *Acta psychologica*, *138*(1), 1-10.
[3] Foolen, A., Gerrevink, R. V., Hogeweg, L., & Prawiro-Atmodjo, P. (2009). The placement of focus particles in Dutch. *Linguistics in the Netherlands*, *26*(1), 51-63.
[4] Bouma, G., Hendriks, P., & Hoeksema, J. (2007). Focus particles inside prepositional phrases: A comparison of Dutch, English, and German. *The Journal of Comparative Germanic Linguistics*, *10*(1), 1-24.
[5] Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. (2010). Focus particles and prosody processing in Dutch: Evidence from ERPs. In *Speech Prosody 2010-Fifth International Conference*, paper 979.
[6] Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PloS one*, *7*(12), e52508.
[7] Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, *6*(2), 81-96.

Neural responses to tonal and segmental gaps in Mandarin Chinese

Chiung-Yu Chang[1] & Feng-fan Hsieh[1]
[1]*National Tsing Hua University (Taiwan)*
s101080012@m101.nthu.edu.tw, ffhsieh@mx.nthu.edu.tw

Introduction: In this study, we conducted a passive listening ERP experiment to see if the results of previous behavioral experiments could be replicated (e.g., [1,2,3], among many others). In the literature, it has been repeatedly shown that speakers are able to access fine-tuned, gradient judgments about the acceptability of nonce words in their native languages. Regarding Mandarin Chinese, however, there is a complication that hypothetical words may come in two types, namely, (i) Segmental gaps (SG) refer to impossible or non-existing CV or VC combinations, e.g., [ki1] and (ii) Tonal gaps (TG) mean (more than) one of the four lexical tones may be absent in a meaningful monosyllable, e.g., [su3], while [su1], [su2] and [su4] are real words. It is not surprising to see that Mandarin speakers treat TGs as more real-ish words, whereas SGs are much more often rejected as a possible word. As a matter of fact, the prediction is borne out in results of previous behavioral studies [1,2,3]. Thus, our main research goal is to explore if there is evidence for a gradient neural response based on the different types of nonce words in Mandarin Chinese. This is not a trivial issue because it has been reported that results of the behavioral experiments may be completely absent in a neurophysiological experiment (e.g., [4]).

Method: Eleven right-handed native speakers of Taiwanese Mandarin with normal hearing participated in the experiment. A go/no-go task was integrated with the passive listening paradigm to keep the subjects attentive. Participants were asked to press a button only if they hear a pure tone. The critical stimuli consist of 80 TGs, 80 SGs, and 80 attested syllables chosen from all possible combinations of open syllables with the onsets [p, p^h, m, f, t, t^h, n, l, k, k^h, x, ʐ]. The electroencephalography was recorded with 32 Ag/AgCl-sintered electrodes connected to a Neuroscan SynAmps2 amplifier. The data were time-locked to the onset of the auditory stimulus. Epochs were extracted from -150 ms to 750 ms, low-pass filtered (30 Hz) and baseline corrected to the pre-stimulus period. Those with artifacts were rejected by visual inspection. Mean amplitude, 15% and 50% fractional area latencies (hereinafter FAL) were measured in the time window of 300-500 ms at 9 electrodes (FC3, FCZ, FC4, C3, CZ, C4, CP3, CPZ, CP4) using the ERP Measurement Tool in the ERPLAB package [5].

Results: The grand average ERP waveforms from 4 selected electrodes were plotted in Figure 1. A negativity which has the typical centroparietal distribution of the N400 (Figure 2) developed around 250 ms and dominated the 300-500 ms window in all conditions. Friedman tests performed on the data from each electrode revealed significant differences between the 15% FALs for the three conditions at FC3 and CZ. Post-hoc comparisons using Wilcoxon-Nemenyi-McDonald-Thompson tests showed that 15% FAL was shorter for TGs than that for real words at FC3 and that it was shorter than that for SGs at CZ. Similarly, significant differences between the 50% FALs for the three conditions were found at FC3, FCZ, C3, and CZ. Post-hoc comparisons showed that 50% FAL for TGs at these electrodes were shorter than that for real words. In addition, 50% FAL was shorter for TGs than that for SGs at CZ and FC3. As for the mean amplitude, no significant difference between conditions at the group level was observed.

Discussion: One of our principal findings is that there is *no* statistically significant difference of the amplitude of N400 across SG, TG and real words, even though the amplitude of the N400 component has been shown to be affected by the lexical status [6] and phonological well-formedness [4]. In other words, our participants do not exhibit a gradient neural responses to the well-formedness and attestedness of the stimuli (cf. [4]). Nevertheless, a closer examination reveals that the main differences between TGs and the rest of the stimuli (i.e., SG and real words) lie in the *latency* of N400, suggesting that upon hearing a potential word, our participants first analyzed an attested combination of tone and segmental makeup or an unattested combination. Taken together, these findings seem to lend *limited* support to a pre-lexical phonological processing, according to which stimuli are first checked against phonotactic restrictions before being subjected to a full lexical search. One possible explanation is that the co-occurrence restrictions on CV syllables (e.g., *[ki]) are not encoded in UG, unlike the Sonority Sequencing Principle (SSP), which induces a significant N400 effect in English (e.g., [4]). Another possibility is that the phonotactic, but not the tonotactic, restrictions are processed at a later point, but this is not confirmed in the present study.

Conclusion: Our results, albeit inconclusive, suggest that tone plays a role in phonotactics as well, in addition to other types of violations at the segmental level.

Finally, in the conference, we shall report results from a similar ERP experiment using the lexical decision paradigm.

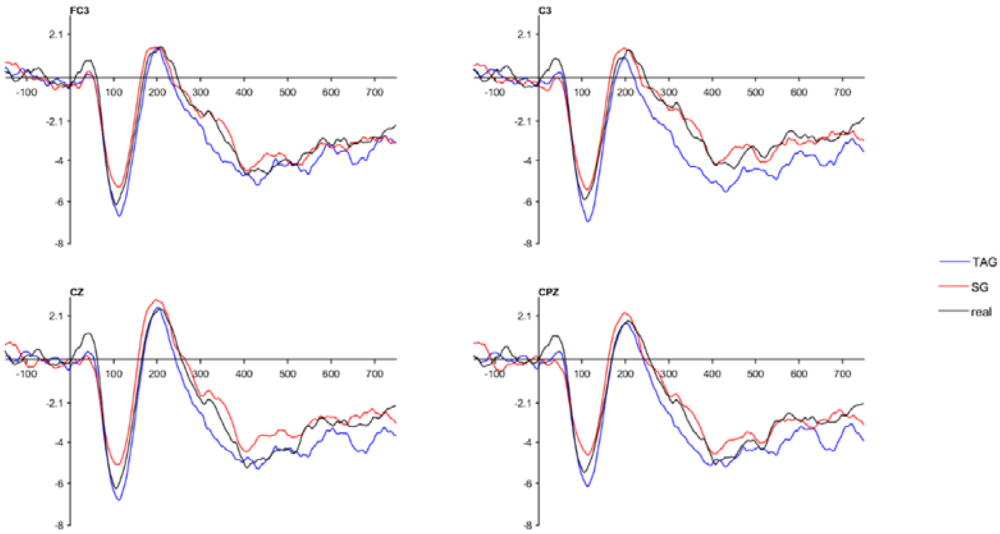


Fig.1 The grand average ERP waveforms of the three conditions.

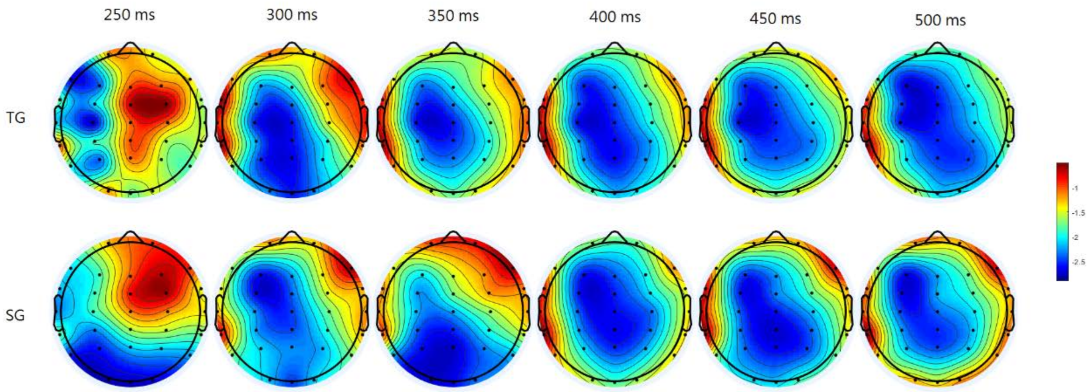


Fig.2 Topographic maps for the TG (top) and SG (bottom) conditions.

References

[1] Lai, Y.-C. (2006). *A perceptual investigation on Mandarin tonotactic gaps.* (Master thesis, National Tsing Hua University, Unpublished).

[2] Myers, J. (2002). An analogical approach to the Mandarin syllabary. *Chinese Phonology*, *11*, 163-190.

[3] Wang, H. S. (1998). An experimental study on the phonotactic constraints of Mandarin Chinese. *Studia Linguistica Serica*, 259-268.

[4] White, J., & Chiu, F. (2017). Disentangling phonological well-formedness and attestedness: An ERP study of onset clusters in English. *Acta Linguistica Academica*, *64*(4), 513-537.

[5] Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, *8*, 213.

[6] Domahs, U., Kehrein, W., Knaus, J., Wiese, R., & Schlesewsky, M. (2009). Event-related potentials reflecting the processing of phonological constraint violations. *Language and Speech*, *52*(4), 415-435.

No word-specific mismatch negativity effect in Mandarin speakers

Stephen Politzer-Ahles[1], Suyeon Im[1]

[1]*The Hong Kong Polytechnic University (Hong Kong)*
sjpolit@polyu.edu.hk, suyeon.im@polyu.edu.hk

A key feature of the architecture of language processing is presumably that words of a language are represented and processed differently than nonwords. One piece of evidence for this comes from the mismatch negativity, a component of brain activity which is larger when hearing real words than when hearing nonwords, and which is not due to low-level physical differences between them. This lexicality effect on the mismatch negativity has been observed in English [1,2], Finnish [3-5], and Thai [6]. These effects suggest that the mismatch negativity can serve as a bio-marker for determining whether a stimulus is processed like a word or not like a word (e.g., if it activates long-term memory traces or initiates computations associated with linguistic processing of morphologically meaningful units). The present study aimed to take advantage of this fact to examine the brain-level processing of Mandarin words, nonwords, and allomorph stimuli: sounds that do not exist as citation forms of real words but that are frequently produced (and heard) as outputs of a phonological alternation (like rising-tone [kʰa˧˥], which native speakers will generally say is not a word, but which is an allomorph of the low-tone real word [kʰa˨˩] 卡 when it appears in certain contexts).

The mismatch negativity is measured by presenting sounds in an oddball paradigm, where participants hear one category of sounds frequently and another category of sounds rarely. The mismatch negativity is the brain response to the rarely presented sounds, minus the response to the frequently presented sounds. Low-level physical differences between words and nonwords can thus be controlled by including those differences in both the frequently-presented and rarely-presented sounds, so they will be "subtracted out" of the mismatch negativity response. The mismatch negativity response itself reflects the detection of the change between the frequent and rare stimuli, rather than the exogeneous response to purely physical aspects of the sound; since the mismatch negativity for words is often bigger than that for nonwords, it may also reflect the activation of long-term memory representations associated with the rare stimulus.

In the present study, 48 native speakers of Mandarin (pre-tested to ensure that they did not believe they know any characters with pronunciations corresponding to the sounds that we are treating as nonwords) participated in a passive oddball paradigm in which they heard the following types of blocks (the stimuli before the ellipse represent the frequently-presented sounds and the stimuli after represent the rarely-presented sounds):

- **Word stimulus as rarely-presented sound:** [kʰai˨˩] [kʰai˨˩] [kʰai˨˩] … **[kʰa˨˩]**
- **Nonword stimulus as rarely-presented sound:** [pʰai˨˩] [pʰai˨˩] [pʰai˨˩] … **[pʰa˨˩]**
- **Word stimulus as rarely-presented sound:** [pʰai˧˥] [pʰai˧˥] [pʰai˧˥] … **[pʰa˧˥]**
- **Allomorph stimulus as rarely-presented sound:** [kʰai˧˥] [kʰai˧˥] [kʰai˧˥] … **[kʰa˧˥]**

In each block, they heard 20 of the frequently-presented stimuli in a row, and then heard 330 stimuli pseudorandomly mixed such that 85% of the sounds (290 trials) were the frequently-presented stimulus and 15% (50 trials) were the rarely-presented stimulus; each block occurred three times, so there were 150 trials of each type of rare stimulus. The acoustic difference between the frequent and rare stimuli was always the same (the frequently-presented stimuli had a diphthong rime [ai] and the rarely-presented stimuli a monophthong rime [a]) and the sounds were cross-spliced to ensure the same rimes were used across rarely-presented stimuli that would be compared. This design ensures that differences observed between the mismatch negativity waves cannot be due to physical differences in the stimuli or to easier-to-detect vs. harder-to-detect changes between the frequent and rare stimuli. (It also would not be due to differences in the lexicality of the frequently-presented stimulus, as previous studies have shown this to not have an impact on the mismatch negativity [1].) We predicted that real-word [kʰa˨˩] would elicit a larger (more negative) mismatch negativity than nonword [pʰa˨˩], whereas the difference between real-word [pʰa˧˥] and allomorph token [kʰa˧˥] (not a real word, but a possible allomorph of one) would depend on whether or not the latter is processed like a real word.

Results (from electrode Fz) from our 48 speakers are shown in the figure below, with solid blue lines for real words and dashed or dotted red lines for nonwords and allomorph stimuli. Contrary to expectation, real words did not elicit larger mismatch negativity than nonwords. These results challenge previous understanding of the nature of apparent lexicality effects in the mismatch negativity. While there are many other paradigms showing that words and nonwords are processed differently (such as the prime lexicality effect [7]), the present results suggest that either mismatch negativity is not a reliable biomarker of this difference, or that lexicality in Mandarin has different properties than it does in other languages where this effect has been shown.

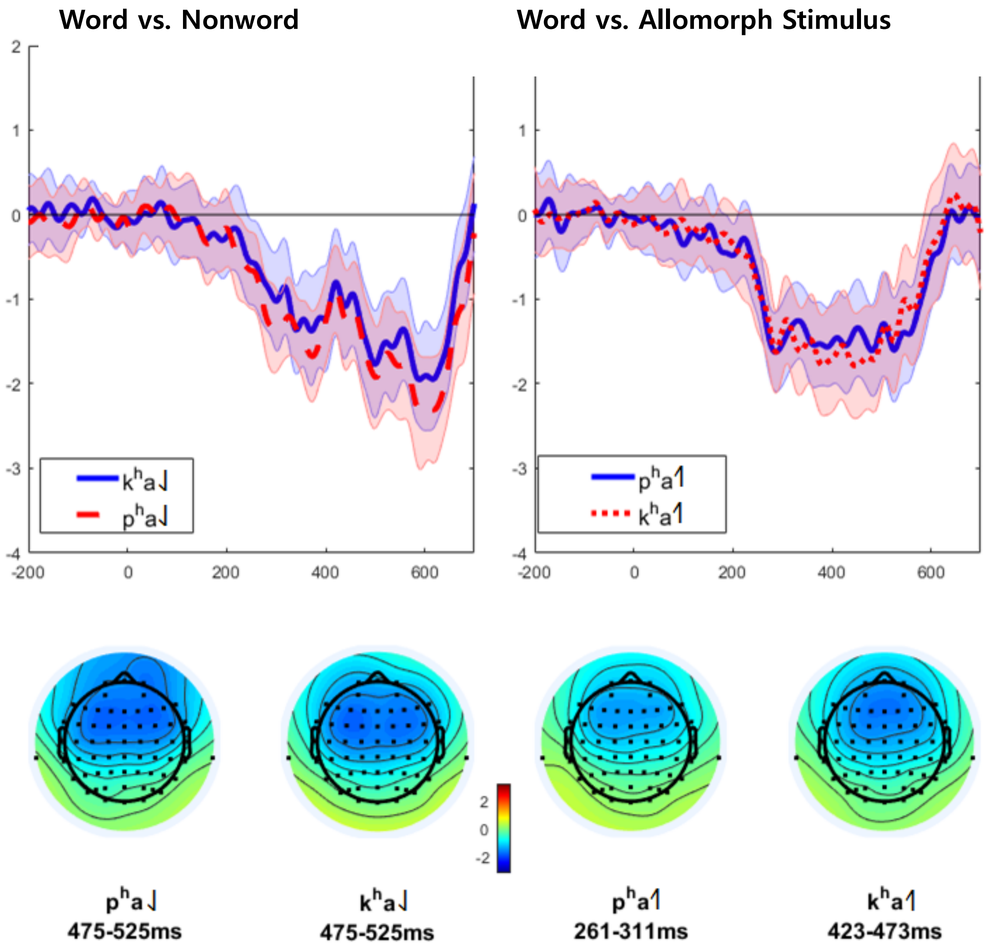


Fig.1 Mismatch negativity waves (deviant minus corresponding standard) for each deviant. Blue lines represent real word stimuli, red dashed lines represent the unambiguous nonword stimulus, and red dotted lines represent the allomorph stimulus.

References

[1] Shtyrov & Pulvermuller (2002). *NeuroReport 13*.
[2] Pettigrew et al. (2004). *Ear & Hearing 25*.
[3] Korpilahti et al. (2001). *Brain & Language 76*.
[4] Pulvermuller et al. (2001). *NeuroImage 14*.
[5] Pulvermuller et al. (2004). *Psychophysiology 41*.
[6] Sittiprapaporn et al. (2003). *Annals of the New York Academy of Sciences 999*.
[7] Qiao et al. (2009). *Cognition 113*.

Poster Presentations
(Day 1)

A cross-linguistic analysis of phonological neighborhood density

Julien Eychenne & Yejin Shon

Hankuk University of Foreign Studies (South Korea),
jeychenne@hufs.ac.kr, shon.yejin@gmail.com

A number of studies have shown that lexical access, in both production and perception, is influenced by phonological neighborhood ([1][2][3] and references therein). The most common metric, neighbor-hood density (ND), is usually operationalized as the number of lexical items that differ from a target word by the addition, deletion or substitution of a single phoneme. Although neighborhood density has featured prominently in phonetic and psycholinguistic research, little is known about how it relates to other properties of the lexicon, and why some words have more neighbors than others. In a landmark study, [4] built and analyzed a database of orthographic and phonological neighborhood densities (CLEARPOND) for five Indo-European languages, namely American English, Dutch, French, German and Spanish. They showed that ND varied across these languages, and that it was correlated with word length and word frequency.

This paper builds upon the work in [4] and offers a more comprehensive analysis of neighborhood density from a cross-linguistic perspective. In order to obtain a more typologically balanced sample, we kept four languages from the CLEARPOND database, namely English, French, Dutch and Spanish, and added six other non-Indo-European languages, including three non-tonal languages (Finnish, Korean and Malay) and three tonal ones (Cantonese, Mandarin and Vietnamese). The inclusion of tonal languages seemed important since it has been shown that ND plays a role at the tonal level as well [5], and conservative estimates suggest that more than 40 % of the World's languages are tonal [6]. In addition to (log) word frequency and (log) word length, we investigated the potential effect of average phoneme probability, normalized pointwise mutual information (NPMI) [7], part of speech (POS), the number of homophones and the tonal/non-tonal nature of the language. In order to make the sample balanced, we only considered the 7,000 most frequent words in each language. (Word frequency was normalized and ND statistics were calculated on this sample.).

The data were analyzed using a negative binomial mixed effects model with the above mentioned predictors as fixed effects, and language as a random effect. In addition to confirming and nuancing the findings in [4], regarding the effect of word frequency and word length, the results of this study show that ND is positively correlated with average phoneme probability and average NPMI (words which have a more phonologically predictable shape tend to have more neighbors), as well as the number of homophones: words with higher ND tend to have more homophones, which suggests that they lie in a dense network of similar words. In addition, we find that verbs have more neighbors than the other categories. Most importantly, we find a clear difference between tonal and non-tonal languages, since we find that tonality significantly interacts with part of speech, NPMI, phoneme probability and homophone density. Our results also reveal that there is important cross-linguistic variation (see Figure 1). We discuss the significance of these results for phonetic and psycholinguistic experiments involving ND.

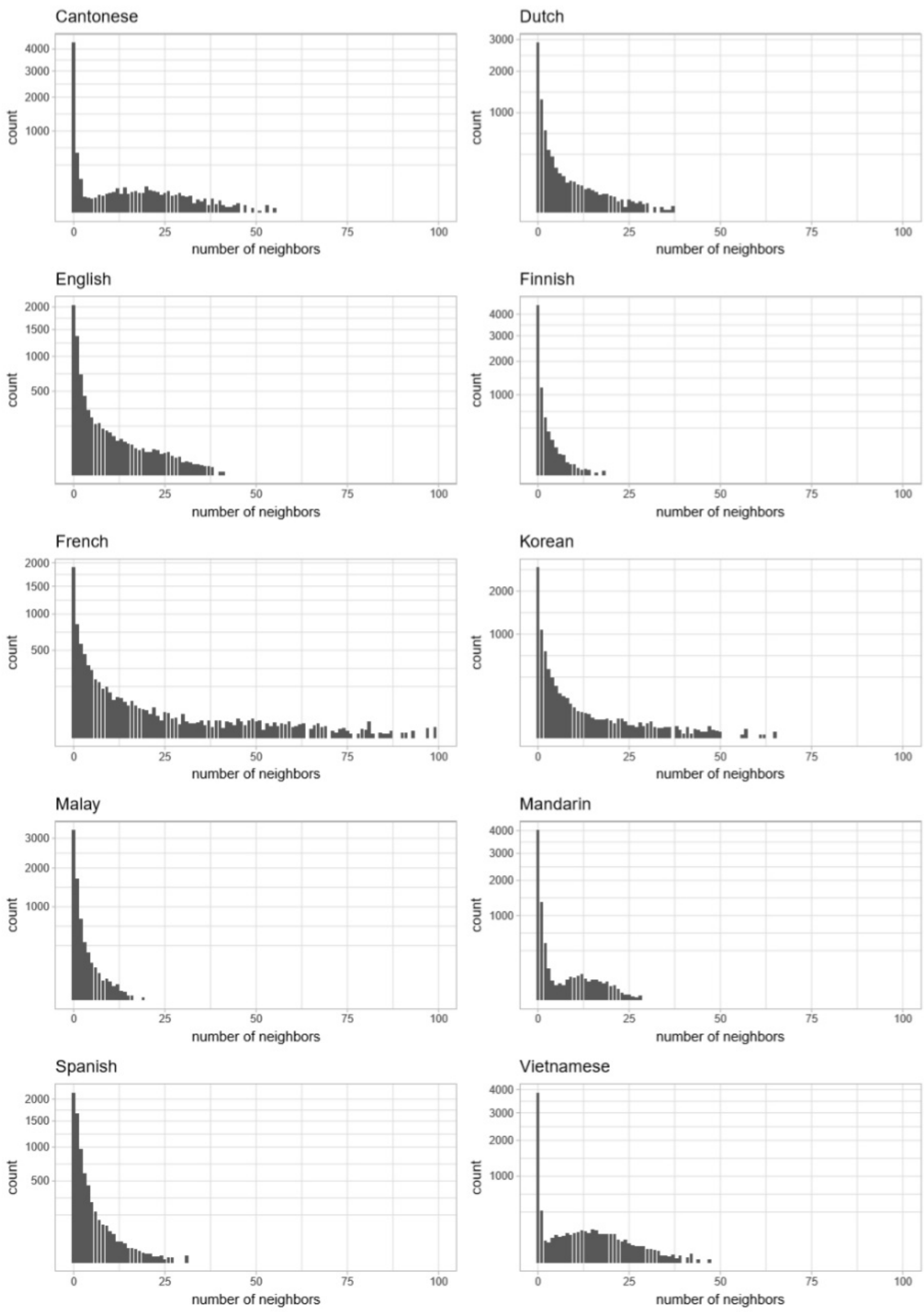


Fig.1 Number of neighbors across 10 languages in frequency spectrum

References

[1] Luce, P. A., & Pisoni, D. B.. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and hearing 19*(1). 1-36.

[2] Munson, B., & Solomon, N. P. (2004). The Effect of Phonological Neighborhood Density on Vowel Articulation. *Journal of Speech Language and Hearing Research 47*(5). 1048-1058.

[3] Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: Re-lating coarticulation and hyperarticulation. *Journal of Phonetics 41*(6). 491-508.

[4] Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLOS ONE 7*(8). e43230.

[5] Yao, Y., & Sharma, B. (2017). What is in the neighborhood of a tonal syllable? Evidence from auditory lexical decision in Mandarin Chinese. *Proceedings of the Linguistic Society of America 2*. 1-14.

[6] Maddieson, I. (2013). Tone. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/13.

[7] Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL* 31-40.

Vowel length of emphasized Cebuano adjectives

Kevin Samejon

Notre Dame of Marbel Univ. (Philippines)
samekevin.ling@gmail.com

Introduction. The Cebuano language does not employ phonetic length distinction to express lexical contrast. However, phonetic length has been generally observed to also indicate emphasis as in the Cebuano utterance, *Lamiiiiii kaayo ang sud-an*, trans. 'The dish is veeeeeery delicious', a pragmatic use of phonetic length. Here, instead of the intensifier adverb *kaayo*, trans. 'very', it is the adjective word *lami*, trans. 'delicious', that undergoes this durational change. Following the observation of an earlier experiment in English [2], the present study investigated this phonetic behavior on Cebuano vowels. An experimental production task was administered. Responses were recorded, annotated, described, and acoustically analyzed.

Methods. Seven (7) native Cebuano speakers, average age of 19 years old, participated in the study. The stimuli for this study included six commonly used adjectives. Target words are disyllabic bearing a CV.CVC structure, with stress on the second syllable and end with a glottal stop. These words are framed first in a non-emphatic sentence, e.g., *Lami kaayo ang sud-an*, and had their vowels replicated to indicate emphasis from level 1 through 5, i.e., *lamiii, lamiiii, lamiiiii, lamiiiiii,* and *lamiiiiiii*. Noticeably, level 1 starts with three orthographic *i*'s because in Cebuano adjectives simply adding another *a* or *i*, e.g., *lamii,* can function as morphemic suffix which transforms an adjective into a verb in imperative mood. It will be read with an intervocalic glottal stop as well. There is a total of 36 stimuli (3 vowels * 2 adjectives * 6 emphasis levels).

What followed a brief practice session is the experimental production task. The production task consisted of seven blocks, with 36 randomized sentences on each block. This totals to 252 tokens. Responses were recorded using Jabra UC Voice 550 MS Duo Lync optimized corded headset, and a computer running Audacity at a 16-bit resolution with sampling rate of 22,050Hz. Stimuli were displayed on the computer screen monitor and participants proceed to the next prompt using the arrow key. Breaks were given in between blocks. Few tokens were skipped and were still mispronounced even after the orthographic precaution takes as described above.

Acoustic analysis was done by marking and extracting durational boundaries on Praat [1] using waveform and spectrogram as indicators. For statistical analysis, Pearson correlation (r) was used to measure correlation between the five emphasis levels and duration, where the "no emphasis" stimuli is excluded. A linear regression analysis on the emphatic conditions followed to determine the increase in vowel duration for each level. Given the multiplicity of comparisons for every condition in this experiment, a pair-wise comparison was not pursued to avoid Type I error. However, as learned from [2], an independent t-test for each speaker via Bonferroni adjustment of significance level to $\alpha = .01$ in each successive pair of comparison was used. Error bars are employed to aid visualization of this variance, and individual speaker data was considered.

Results and discussion. Results revealed positive correlation between emphasis levels and phonetic duration but no other significant comparisons across levels of emphasis for each speaker. This suggests that, based on the articulatory facility of speakers, they can produce gradable phonetic length beyond the usual binary distinction yet seem to find it inconsequential to maintain clear durational distinction on other levels of emphasis. Speakers then seem to subscribe to the usual two-way durational distinction of no emphasis/emphasis [4]. Statistically insignificant p-values on the conducted t-tests, and error bar reversals and overlaps may help demonstrate this, i.e., reversals in Speaker 2, 3, 5, and 6 (see Fig. 1).

It is also worth noting that using orthographic replication as indicator of emphasis in the stimuli does not interfere with the aims of the study. Simply, if participants count orthographic iterations of vowels, we may then expect linear and strict correlations between levels of emphasis and duration. Moreover, if counting has been employed by the speakers, we cannot expect significant jump from level 0 and level 1 [2].

In the future, perception studies can be pursued. Even with a very low regression coefficient of 30 ms in the present study, compared to other parallel studies [2,3], using this value as basis for a perception task may be enough because listeners were found to be sensitive to durational differences of as short as 12.5 ms [5]. Furthermore, a closer look at the prosody of emphatic utterances might reveal greater insight than singularly investigating phonetic cues such as length.

Table 1. Pearson *r*-value, regression coefficients, and maximum vowel/rhyme duration for each speaker.

Cebuano Speaker	*r*	Coefficient (ms)	Max duration (ms)
1	.47	30	643
2	.29	8	344
3	.27	15	487
4	.26	19	752
5	.25	14	575
6	.22	15	634
7	.18	9	530

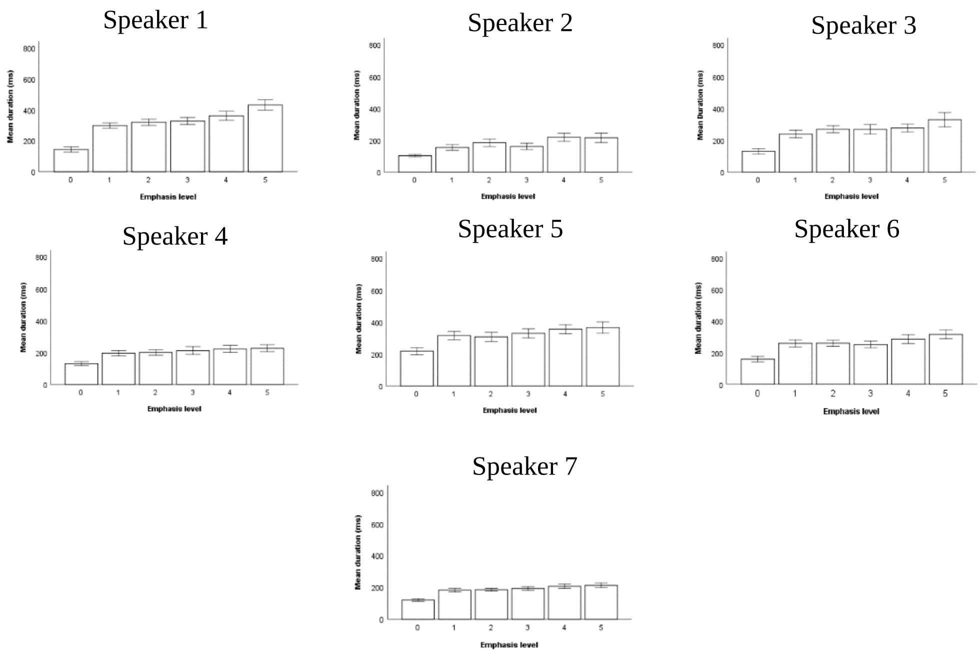


Fig. 1 Error bars for speakers 1 through 7. (Error bars all throughout represent 95% CI.)

References

[1] Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. v.6.0.40 http://www.praat.org/

[2] Braver, A., Dresher, N., & Kawahara, S. (2014). The phonetics of emphatic vowel lengthening in English. In A. Albright, & M. A. Fullwood (eds.), *Proceedings of the 2014 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America.

[3] Kawahara, S., & Braver, A. (2013). The phonetics of multiple vowel lengthening in Japanese. *Open Journal of Modern Linguistics*, *3*, 141–148.

[4] Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839–862.

[5] Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neuroscience Letters*, *107*, 347–352.

English obstruent perception by native Mandarin, Korean, and English speakers

Yen-Chen Hao[1] & Kenneth de Jong[2]

[1]*University of Tennessee (USA),* [2]*Indiana University (USA)*
yhao6@utk.edu, kdejong@indiana.edu

This study investigates the accuracy and confusion patterns of eight English obstruents /p b t d f v θ ð/ in the onset and coda position identified by Mandarin and Korean-speaking L2 learners as well as by a control group of native English speakers. According to theoretical models that underscore the influence of L1 categories on L2 perception (e.g. PAM: [1], [2]; SLM: [3]), L1 Mandarin and Korean speakers are expected to differ in their perception of English onset and coda obstruents due to the different correspondence between their respective L1 consonants and those in English. Specifically, Mandarin has /f/ in the onset position while none of the target fricatives exist in Korean. As for the coda position, Mandarin does not allow any obstruent while Korean neutralizes underlying stops and fricatives into homorganic voiceless lax stops. On the other hand, theories advocating for language universal markedness suggest that some L2 sounds would be more difficult than others regardless of the L1 background. To examine these hypotheses, we compared English obstruent identification by 41 Mandarin, 40 Korean, and 17 English speakers, each tested in their native country. The results showed that all three groups were significantly more accurate in perceiving obstruents in the onset than in the coda position, more accurate in identifying voiceless than voiced targets, stops than fricatives, and labials than coronals. The L1 English speakers generally achieved higher accuracy than the L2 learners except in the identification of /θ/ and /ð/. The two learner groups did not differ in their overall accuracy in the onset position, yet in the coda position the Mandarin group outperformed the Korean group. With regard to the specific obstruents, in the onset position the Mandarin group achieved higher accuracy than the Korean group on /b/, /d/, and /f/, while in the coda position the Mandarin group achieved higher accuracy on /d/. The three groups' mean accuracy rates on the eight target sounds in the two prosodic positions are presented in Figure 1.

In addition to accuracy rates, all three groups' erroneous responses were analysed based on the voicing, manner, and place confusions. It is found that both the Mandarin and Korean groups exhibited a stronger bias toward voiceless consonants in the coda position than the English group. As for the manner confusion, the two learner groups were more biased toward fricative responses in the coda position than the English group, while the Korean group showed additional confusion of misidentifying fricatives as stops. Finally, all three groups showed some bias toward labials in the coda position, which appears to result from an aversion to the /θ/ and /ð/ responses. These findings suggest that L1 experience plays an important role in L2 sound perception. However, the presence of a similar L1 counterpart of the L2 target sound does not necessarily guarantee the ease of its acquisition. While some of the learners' featural confusion patterns can be explained by their L1 influence, the general similarity of the two learner groups is indicative of a robust and pervasive language-independent tendency in L2 speech perception. Furthermore, the comparison of the learners and native speakers shows that the perception of some obstruents may be inherently difficult.

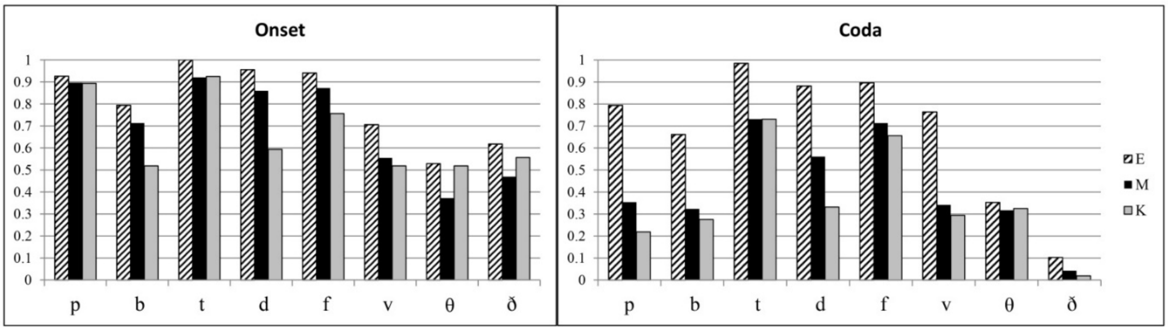


Fig. 1 The three L1 groups' mean accuracy rates for the eight English obstruents in the onset (left) and coda position (right).

References

[1] Best, C. T. (1995). A direct realist view of 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.

[2] Best, C.T., and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn and M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp.13–34). Amsterdam: Benjamin.

[3] 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.

Pitch accent and the three-way laryngeal contrast in North Kyungsang Korean

Young Hwang, Samson Lotven, Kelly Berkson

Indiana University (USA)

hwang24@indiana.edu, slotven@indiana.edu, kberkson@indiana.edu

Korean is typologically unusual in that it has a three-way contrast between phonetically voiceless stops in word-initial position. The three different categories are often called *fortis*, *lenis*, and *aspirated*, and they are differentiated by both voice onset time (VOT) and fundamental frequency (f0) on following vowels [1,2,3,4]. VOT is shortest for fortis, intermediate for lenis, and longest for aspirated (fortis < lenis < aspirated). The fortis and aspirated categories have higher f0, while the lenis category has lower f0 on the following vowels (lenis < fortis, aspirated).

Traditionally, VOT was considered to be the primary acoustic correlate to the stop distinction [5], while f0 was considered to be a redundant cue [3,6,7,8]. However, recent studies have reported that Korean stops are undergoing a diachronic sound change and that VOT values of lenis and aspirated stops have merged due to a lengthened VOT for lenis stops and a shortened VOT for aspirated stops [8]. As a result, the distinction between the lenis and aspirated stops is now coded primarily by an f0 difference on the following vowels [4,7,8,9].

The new pattern is generally observed in Seoul Korean, but not all dialects in Korea are reported to have undergone this change. The South Kyungsang (SK) dialect, one of two Kyungsang dialects which retain the pitch accent system from Middle Korean, has not undergone the change, and even younger speakers keep the traditional pattern [10]. Lee, Politzer-Ahles & Jongman [9] assert that VOT is still used as a main cue of the stop distinction in the SK dialect since f0 is already used for the pitch accent system.

However, the change seems to be happening in North Kyungsang (NK) even though the NK dialect is a pitch accent dialect. Holliday & Kong [11] reported that female speakers of the NK dialect have more VOT overlap between the lenis and aspirated stop categories than male speakers. Very little other work on the NK dialect exists, however, raising questions about how robust this sound change is in NK and how it is progressing. With this in mind, we report on an apparent time study investigating the change-in-progress affecting VOT and f0 in NK stop consonants and its interaction with the pitch accent system. To examine age and gender effects, 4 groups of native NK speakers were included: 6 older males, 5 older females, 6 younger males, and 6 younger females (N = 23 talkers). Stimuli include 36 bisyllabic words with Low-High and High-Low tone patterns beginning with bilabial, alveolar, and velar stops of each of the three types (fortis, lenis, and aspirated) in the onset of the first syllable embedded in a carrier sentence. VOT measured from stop burst to the onset of periodicity and average f0 extracted from the 10% point of the following vowel are reported.

The results indicate that the four groups of talkers are in different stages of the VOT change. The younger groups and the female talkers are more likely to produce the innovative pattern, with longer lenis and shorter aspirated VOT values (meaning lenis and aspirated pattern together). For f0, younger speakers tend to have greater f0 differences for the stop distinction than older speakers. Aspirated stops are not well differentiated from lenis stops by VOT for many younger speakers and for these speakers f0 differences are larger. Younger speakers also have a more minimal f0 distinction between High and Low tones than older speakers.

In addition, we found evidence that there is an interaction between VOT and the tone contrast in the pitch accent system, such that VOT varies by tone category. This study provides evidence that despite the fact that the NK dialect is a pitch accent dialect there can still be the diachronic change in the stop distinction, and that the laryngeal contrasts are becoming merged into the tonal system.

References

[1] Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*, 198-228.

[2] Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics, 1,* 263-272.

[3] Kang, Y., & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua,134,* 62-74.

[4] 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.

[5] Lisker, L., & Abramson A. (1964). A cross-language study of voicing in initial stops. *Word, 20,* 384-422.

[6] Kang, K.-H., & Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of Acoustical Society of America, 119,* 1672-1683.

[7] 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(6)*, 3909-3917.

[8] Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology, 23,* 287-308.

[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(2),* 117-132.

[10] Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association* 42 (2), 145-169.

[11] Holliday, J. J., & Kong, E. J. (2011). Dialectal variation in the acoustic correlates of Korean stops. *International Congress of Phonetic Sciences, XVII,* 17-21.

Choices in abstract phonological analysis have direct consequences for psycholinguistic predictions: the case of phonological neighbourhood networks

Rory Turnbull

University of Hawaiʻi at Mānoa
rory.turnbull@hawaii.edu

Phonological neighbourhood density is a common measure of lexical similarity in psycholinguistics and has been asserted to be relevant to speech perception and production. Neighbourhood is usually defined as a relation which obtains between two words with a phoneme string edit distance of one. However, because phonological descriptions are non-unique in that multiple analyses are possible [1], multiple differing representations of a given language's lexicon are possible. This study investigates how such indeterminacy in phonological representations can influence properties of neighbourhood density and patterns of lexical organization.

A phonological neighborhood network (PNN) is a graph of the lexicon where each word is linked to its phonological neighbors. This graph will show that, for example, the words *bat* and *can* are connected via the intermediate word *cat*. This practical arrangement permits analysis of the lexical structure of languages via the application of tools drawn from graph theory [2, 3]. Several graph-theoretic measures of lexical networks have been shown to be of relevance to language processing: for example, words with a high clustering coefficient – where many of the word's neighbors are also neighbors of each other – have slower speech onset latency in naming tasks [4].

The construction of a PNN requires that lexical entries are already segmented into phonemes, and the presence of phonemes entails a phonological analysis. This point may be obvious but it has far-reaching consequences, as in many cases there are multiple possible phonological analyses with no clear empirical reason to favour one over the other: phonological analysis is indeterminate.

For example, French nasal vowels can be regarded as unitary phonemes (a "surface true" analysis), or as clusters of an oral vowel followed by a nasal stop. These analyses have different implications for our understanding of French (morpho)phonology, but predict the same surface forms. However, they predict different neighbourhood patterns, and therefore different PNN structures. Debates in theoretical phonological analysis therefore have a direct influence on the predictions of psycholinguistic phonetics.

To determine whether this influence was merely a small quirk or a pervasive issue, we constructed multiple lexicons of various phonological analyses of French and Mandarin. Fourteen lexicons of Mandarin were constructed (following [5]) and 180 lexicons of French. PNNs were derived from each of these lexicons. These PNNs were contrasted with PNNs generated from a set of 49 typologically diverse languages.

As expected, the different languages had different network properties, such as graph entropy or average neighbourhood size. Examination of these properties demonstrated two important results:

(1) All networks tended to be highly assortative by degree, meaning that the neighbours of words with many neighbours tended to have many neighbours themselves. This property is common for social networks [6], but PNNs are especially assortative [3]. This finding demonstrates that, regardless of the underlying phonology assumed, the networks have qualitatively similar topology, suggesting organizational constraints to the structure of the lexicon.

(2) The within-language variation among the alternative lexicons of Mandarin and French was as large, if not larger, as the between-language variation among English, Dutch, German, and Korean. This can be seen in Figures 1 and 2 below, which depict distributions of two different network-theoretic measures for each set of languages. This finding demonstrates the sensitivity of PNNs to the phonological representation used as input.

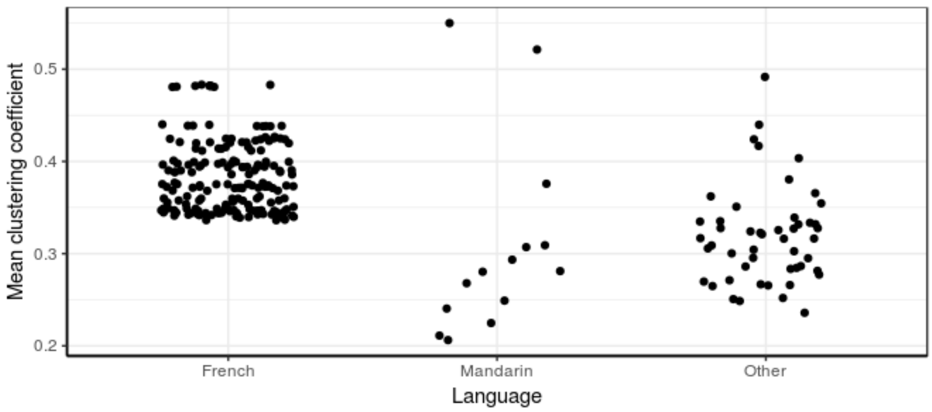


Fig.2 Distributions of mean clustering coefficient on sets of PNNs. The clustering coefficient of a node denotes how many neighbours of that word are neighbours of each other. Note that the variations on French and Mandarin significantly overlap with the "Other" languages, all of which are unique languages.

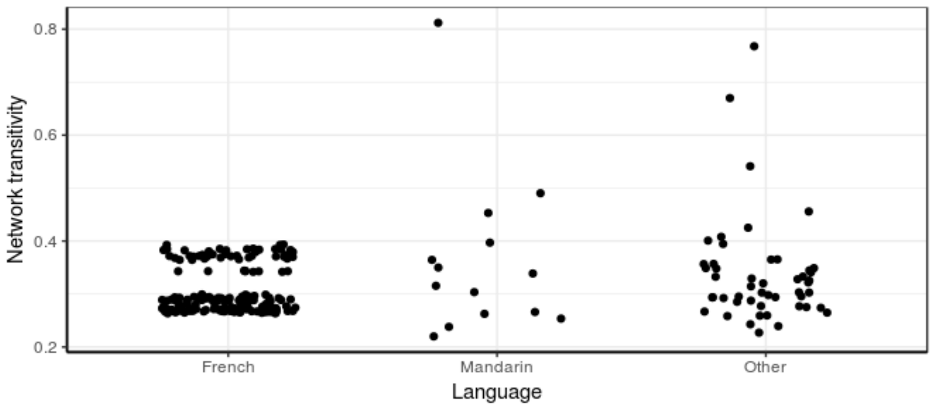


Fig.3 Distribution of network transitivity on sets of PNNs. Transitivity is a measure of ...

While there is a common structural core to all PNNs investigated, the variation between different analyses of the same language was as big if not bigger than variation between different languages. In calculating neighbourhood density, the choice of the phonological analysis is therefore as important as the selection of the language itself. Taken together, this study provides evidence that analytical choices in dealing with phonological indeterminacy can have consequences for the calculation of neighbourhood density.

References

[1] Chao, Y.-R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica*, 4(4):363–397.

[2] Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research*, 51(2), 408–422.

[3] Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, 20(03), 679–685.

[4] Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, 34(4), 685–697.

[5] Neergaard, K., & Huang, C. R. (2016). Graph theoretic approach to Mandarin syllable segmentation. *Proceedings of the Fifteenth International Symposium on Chinese Language and Linguistics (IsCLL).*

[6] Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E*, 67(2), 026126.

Second Language Perception of English Stops by Korean-Speaking Child Learners: Effects of Position and Lexical Knowledge

Haerim Hwang

University of Hawaiʻi (USA)
haerim@hawaii.edu

The Speech Learning Model (SLM) [1] posits that "sounds in the [first language (L1)] and [second language (L2)] are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level" (p. 239). According to this proposal, the position of a category within a word affects the difficulty of perceiving it [2] and it is positional allophones rather than abstract phonemes that play a role in L2 category perception. Although position effects have been observed in a few L2 studies on category perception [3, 4, 5], most studies have tested only advanced adult L2 learners (L2ers). Furthermore, it remains unknown whether L2 categories are mapped onto L1 positional allophones or L1 phonemes. The current study addresses these gaps in the research by testing beginning-level L1-Korean child L2ers to determine (a) whether their perception of the voicing contrast in English stops depends on word position, (b) whether they rely on L1 allophones or L1 phonemes, and (c) whether lexical knowledge effects can be detected.

Stops in Seoul Korean show a three-way contrast between lenis (e.g. /k/), aspirated (e.g. /k^h/), and fortis (e.g. /k'/) variants at the phonemic level. Importantly, although there is no voicing contrast, voiced allophones (e.g. [g]) are used between sonorants. In the coda position, all stops are neutralized to lenis consonants. [6] Depending on which L1 category Korean L2ers of English associate with L2 sounds, we can imagine two possible scenarios for the word-medial voicing contrast (e.g. [pɪki] vs. [pɪgi]): either L2 sounds (e.g. [k], [g]) are respectively mapped onto L1 positional allophones (e.g. [k], [g]) as stated by the SLM, in which case L2ers should have no difficulty with the contrast, or L2 sounds (e.g. [k], [g]) are mapped onto L1 phonemes (e.g. /k/), in which case L2ers should have trouble perceiving the contrast. However, both scenarios make the same predictions for word-initial stops (e.g. [k^hoʊt] vs. [goʊt]), where the aspiration cue is expected to facilitate perception of the voicing contrast (e.g. English [k^h] and [g] perceived as Korean [k^h] and [k] or as /k^h/ and /k/), and for word-final stops (e.g. [bæk] vs. [bæg]), where perception of the contrast should be difficult due to the neuralization in Korean.

Method: Forty 9-year-old L1-Korean L2ers of English with beginning-level proficiency in Seoul, Korea completed a two-talker AX discrimination task in English containing either real words (RW group, n=21) or nonsense words (NW group, n=19). Participants were asked to judge whether the sounds in each word pair were the same or different. Nine real word pairs and nine nonsense word pairs were selected for use in these tasks. Each pair consisted of one word containing a voiceless stop (/p, t, k/) and another containing a voiced stop (/b, d, g/). The contrasting categories appeared in word-initial (k=3), word-medial (k=3), or word-final position (k=3), depending on the condition. The word order in each pair was manipulated to produce four stimuli. Each task consisted of 36 critical trials and 36 fillers in total.

Results (Figure 1): The participants' responses were converted to d' scores and analyzed using a mixed ANOVA with *Group* (RW; NW) as a between-subjects factor and *Position* (word-initial; word-medial; word-final) as a within-subjects factor. The analysis revealed a main effect of *Group* (p<.001) and a significant interaction between *Position* and *Group* (p<.001). In post hoc analyses, both groups showed a significant effect of *Position* (p<.05): Whereas the NW group's d' scores were significantly higher for the word-initial contrast than for the word-medial contrast (p<.05) and the word-final contrast (p<.01), the RW group's d' scores were significantly lower for the word-initial contrast than for the other two contrasts (p<.05). This finding indicates that the perception of an L2 category is affected by its position in the word. In particular, the NW group's performance suggests that, contra the SLM, beginning-level child L2ers map L2 sounds onto L1 phonemes. Furthermore, the difference observed between the RW and NW groups provides evidence of lexical knowledge interference effects in L2 category perception. The RW participants may have experienced increased processing load due to activating L2 lexical representations along with L2 phonological representations, which could have inhibited their ability to distinguish the target sounds.

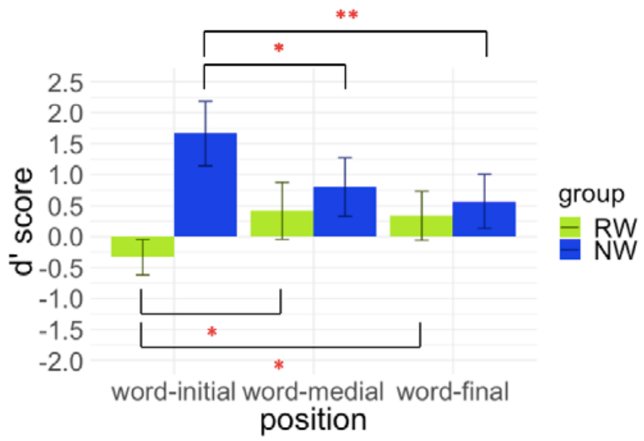


Fig. 1 Mean *d'* score per position and group. Error bars show standard errors. Significance level: *p<.05; **p<.01.

References

[1] 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.

[2] Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech*. Cambridge: Cambridge University Press.

[3] Park, H., & de Jong, K. J. (2008). Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics*, *36*, 704–723.

[4] Park, H., & de Jong, K. J. (2017). Perceptual category mapping between English and Korean obstruents in non-CV positions: Prosodic location effects in second language identification skills. *Journal of Phonetics*, *62*, 12–33.

[5] Takagi, N. (1993). *Perception of American English /r/ and /1/ by adult Japanese learners of English* (Unpublished doctoral dissertation). University of California-Irvine, Irvine, CA.

[6] Kim, H., & Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics*, *24*, 295–312.

Effects of orthographic input on L2 production: the case of Korean-speaking learners of Mandarin Chinese

Sang-Im Lee-Kim

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

Of the many factors affecting the acquisition of second language (L2) phonology, the role of orthographic input has only recently been considered. Despite some empirical findings demonstrating a positive role of orthographic input in L2 phonological acquisition [1, 8], other experimental results have suggested a negative influence [2, 5]. When orthographic input is not entirely consistent with corresponding phonetic forms, in particular, phonological developments may be significantly hindered. The present study is primarily concerned with L2 production to shed light on the role of orthographic input in the development of the fine-grained sub-phonemic features of non-native sounds. Specifically, we explored the development of stop categories and lexical tones in Mandarin by native speakers of Seoul Korean. The learners' development was closely tracked in a longitudinal study.

Fourteen Seoul Korean speakers with no prior experience with tone languages participated in the production study. They were randomly divided into two groups based on the type of stimuli presented in the experiment: orthographic ("Pinyin group", 5F, 3M) or audio ("Audio group", 2F, 4M). For the Pinyin group, stimuli were presented orthographically without auditory stimuli whereas participants in the audio group were presented with auditory stimuli produced by a female native speaker of Mandarin. Participants were college students (*M*(age) = 25.8) in Korea who had registered for Elementary Chinese courses. The courses took place in a formal class setting four hours a week, including three hours of lectures taught in Korean by a Korean instructor and a one-hour practicum taught in Mandarin by a Chinese assistant. The stimuli were monosyllabic words consisting of Mandarin voiceless stops contrasting in aspiration (e.g., unaspirated: <bā bá bǎ bà> vs. aspirated: <pā pá pǎ pà>). A total of 24 stimuli (3(unasp, asp, nasal) x 2(lab, cor) x 4(tones)) were repeated three times in each task, and the same materials were used throughout the longitudinal study testing the same participants three times over a semester: Phase 1 (after one week of Mandarin exposure), Phase 2 (after 8 weeks), and Phase 3 (after 15 weeks).

The results of stop VOT analyses showed that while all participants correctly produced long VOTs for the aspirated stops, four out of the eight speakers in the Pinyin group produced inordinately long VOTs for the unaspirated stops (Figure 1), especially in the earlier stages of learning. The four other learners in the Pinyin group produced correct VOT targets for the unaspirated stops, so did all six learners in the audio group. Importantly, a clear correlation was observed between the erroneous patterns in VOT and in *f0* of Tone 1 (X^{55}, high-level) production: those learners with overestimated VOTs for the unaspirated stops produced the stops with lower *f0* (Figure 2). In contrast, those who produced accurate stop VOTs showed no such separation in T1 *f0* trajectories between aspirated and unaspirated stops, nor did any of the participants in the audio group. This result highlights the negative influence of orthographic input on the production of L2 sounds. Korean learners of English have been reported to map English voiced stops <b, d> onto Korean lenis stops more so than onto fortis stops [6, 7]. As all Korean participants in the present study were familiar with the English alphabet, Pinyin orthography (e.g., <b d>) may trigger a mapping onto Korean lenis stops [Long VOT–Low *f0*], overriding actual acoustic-phonetic input of Mandarin unaspirated stops [Short VOT–High *f0*(T1)].

As for the tone production, T2 (X^{35}, <bá>) – T3 ($X^{21(4)}$, <bǎ>) contrasts presented a particular challenge, consistent with previous studies showing non-native listeners' difficulty to perceive the difference between these two tones (Huang & Johnson 2010). In particular, the learners' T2 and T3 *f0* trajectories tended to be more similar, unlike native Mandarin speakers' production wherein T2 and T3 *f0* trajectories diverge to a greater extent toward the end of a syllable [4]. More interestingly, many learners' ability to accurately produce T2 and T3 diminished over time, rather than improving. Figure 3 illustrates some representative cases. This pattern is likely driven by non-phonetic factors. The Pinyin tone diacritic denoting T3 (<bǎ>) implies that the ending of the tone should be high rather than low. Moreover, the Tone 3 sandhi rule (/T3/→[T2]/_T3) may also contribute to the perceptual assimilation of the two tones. Beginner learners are introduced to this rule quite early due to common expressions such as /**ni**$^{\mathbf{214}}$hao^{214}/ ([**ni**$^{\mathbf{35}}$hao$^{21(4)}$]) *you-good* 'hello'. Along with the Pinyin T3 diacritic, the T2-T3 phonological alternation may mistakenly lead learners to assume that the *f0* rise at the end of T3 is indeed an indispensable part of this tone, which may make T2 and T3 even more perceptually similar than they actually are.

Taken together, the results of the study show that novice learners' difficulty with L2 contrasts arises from various L2 experiences including orthographic input, not solely from the acoustic properties of sounds.

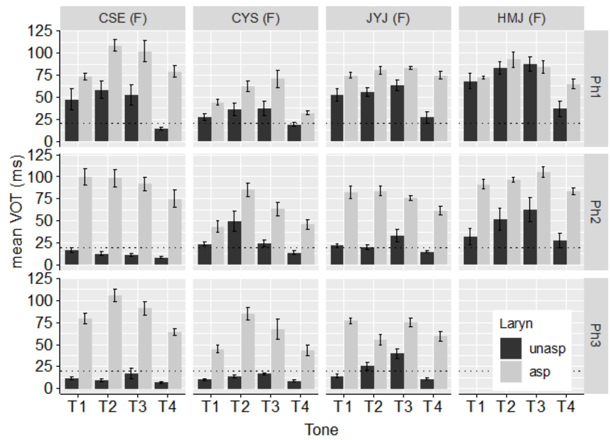


Fig.1 Pinyin-based learners showing inordinately long VOTs for the Mandarin unaspirated stops.

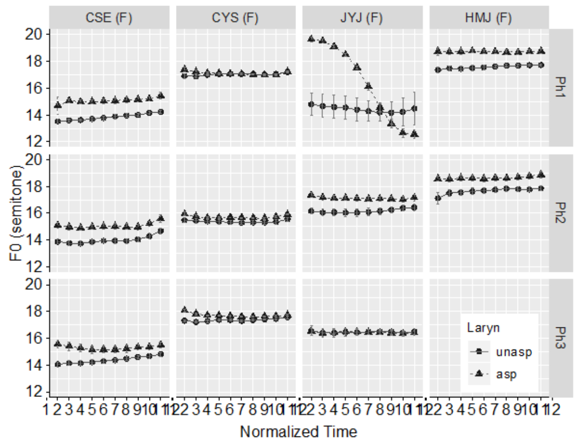


Fig.2 Mean f0 trajectories of T1 for Pinyin-based individual speakers with VOT overestimation for the Mandarin unaspirated stops.

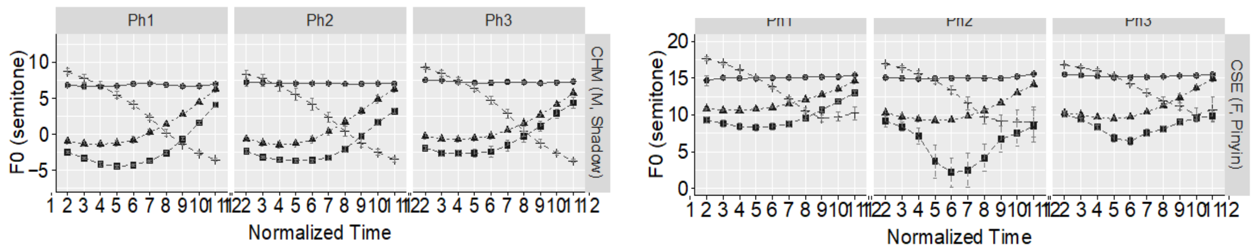


Fig. 3 Mean *f0* trajectories of four lexical tones following the aspirated stops. Representative cases illustrating the (partially) diminishing T2-T3 contrasts.

References

[1] Escudero, P., Hayes-Harb, R., Mitterer, H. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36, 345–360.

[2] Hayes-Harb, R., Cheng, H.-W. 2016. The influence of the Pinyin and Zhuyin writing systems on the acquisition of Mandarin word Forms by native English speakers. *Frontiers in Psychology*, 7, 1–13.

[3] Huang, T., Johnson, K. 2010. Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, 67. 243–267.

[4] Kuang, J. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *Journal of the Acoustical Society of America*, 142(3), 1693–1706.

[5] Mathieu, L. 2016. The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research*, 32(2), 145–170.

[6] Park, H., de Jong, K. J. 2008. Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics*, 36, 704–723.

[7] Schmidt, A. M. 1996. Cross-language identification of consonants. Part 1. Korean perception of English. *Journal of the Acoustical Society of America*, 99(5), 3201–3211.

[8] Showalter, C., Hayes-Harb, R. 2013. Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, 29(2), 185–210.

Enhancement of sibilant contrasts in near merger during word processing by Min-Mandarin bi-dialectal speakers

Waan-rur Lu[1] & Sang-Im Lee-Kim[2]

National Chiao Tung University (Taiwan)[1, 2]
lynnwlulu@gmail.com, sangimleekim@nctu.edu.tw

Many Chinese dialects including Southern Min lack retroflex sibilants in the system. Accordingly, bi-dialectal Min-Mandarin speakers in Taiwan tend to show dental-retroflex sibilant merger conditioned by multiple factors. For example, Shih [3] demonstrated young adults convey a more distinctive Mandarin three-way sibilant contrast (dental, retroflex, alveopalatal) compared to older generations. Chuang and Fon [2] showed merger is further conditioned by gender, region, and the frequency of the use of Min. The lack of sibilant contrasts by those bi-dialectal speakers may have significant consequences in word processing. Wu and Ma [4], for example, have shown that early Hakka-Mandarin bilinguals demonstrated a sign of automatic phonological activation of both lexical items containing dental and retroflex sibilants, leading to a significant delay in word processing. The present study explored the nature of sibilant merger implemented by young bi-dialectal speakers in Taiwan using varying experimental paradigms.

First, a reading task was conducted to examine participants' regular sibilant production. This task included a set of disyllabic stimuli with all six sibilants (dentals /s ts tsh/ vs. retroflexes /ʂ tʂ tʂh/) in word-initial position. Forty college students in total were recruited and divided into three groups based on whether they conveyed clear retroflexes in contrast to dentals in their production: Merger (11F/10M), Intermediate (3F/4M), and Contrast (6F/6M) groups. Sibilant merger among young speakers was more gradient than previously assumed. The intermediate group, in particular, singled out the speakers who variably conveyed the contrast: approximately half of the stimuli carried the contrast, while the other half showed merger.

In a priming study, experimental stimuli included disyllabic Mandarin compounds containing word-initial sibilants as well as many filler items. The target words were paired with either congruent (/**tʂ**an^{51}li^{51}/ 'stand still' – /**tʂ**an^{51}li^{51}/ 'combat capability'; /**s**i^{55} yi^{35}/ 'master of ceremonies' – /**s**i^{55} yi^{35}/ 'personal relationship') or incongruent primes (/**ts**i^{51}li^{51}/ 'self-reliance' – /**tʂ**i^{51}li^{51}/ 'Chile'; /**ʂ**an^{55} jiao^{214}/ 'hillside' – /**s**an^{55} jiao^{214}/'triangle') (Table 1). The word frequencies were matched across conditions using Academia Sinica Balanced Corpus of Modern Chinese. A word-naming study was carried out in e-prime designed in a backward masked prime paradigm. After a fixation cross, a brief prime (50 ms) was presented followed by a 50 ms mask (#######) on a computer screen. A target item then appeared on the screen for three seconds. Upon seeing the target, participants were asked to read aloud the target as quickly and accurately as possible. All the stimuli in both studies were presented visually in Chinese characters which are logographic and do not carry phonological information of the words.

Table 1. Examples of experimental stimuli.

Congruent prime	Target	Incongruent prime	Target
站立 'stand still' /**tʂ**an^{51} li^{51}/	戰力 'combat capability' /**tʂ**an^{51} li^{51}/	自立 'self-reliance' /**ts**i^{51} li^{51}/	智利 'Chile' /**tʂ**i^{51} li^{51}/
retroflex	retroflex	**dental**	**retroflex**
司儀 'show host' /**s**i^{55} yi^{35}/	私誼 'personal relationship' /**s**i^{55} yi^{35}/	山腳 'hillside' /**ʂ**an^{55} jiao^{214}/	三角 'triangle' /**s**an^{55} jiao^{214}/
dental	dental	**retroflex**	**dental**

The frication noise during the primed sibilant production was annotated and submitted to a time-averaging spectral analysis. The distribution of CoG (Center of Gravity) values was summarized in Figure 1. Overall, the CoG distance between dental and retroflex sibilants (ΔCoG) was the smallest for the Merger group (1,465 Hz), followed by Intermediate (2,300 Hz), and the largest for the Contrast group (2,957 Hz). Within the Merger group, male speakers were more likely to merge sibilants than females (ΔCoG(male)= 1,116 Hz, ΔCoG(female)= 1,783 Hz), similar to the findings in Chuang and Fon (2010). Nevertheless, the Merger group showed small but consistent spectral differences between dentals and retroflexes, suggesting a contrast in near merger, not complete neutralization [5]. Further, the results generally showed dissimilation of the sibilant targets from the incongruent primes: dentals become more *dentalized* (higher CoG) when primed with *retroflexes*, whereas retroflexes become more *retroflexed* (lower CoG) when primed with *dentals*, resulting in the enhancement of the dental-retroflex contrast. This pattern was most evident for the male speakers in the Merger group (ΔCoG(Congruent)= 513 Hz vs. ΔCoG(Incongruent)= 1,938 Hz), as verified by a significant interaction between TargetPlace and PrimeType (p =.016). Despite some trend, the interactions were not significant for other groups.

Taken together, the results suggest that the speakers showing dental-retroflex near merger, in fact, have fully distinct representations of the contrasting sibilants, which may be implemented explicitly when primed with directly contrasting sounds. In a formal read speech, however, the contrast may remain covert – being realized as near merger – presumably due to the social prestige associated with non-full-retroflexion for the retroflex sibilants in Taiwan [1].

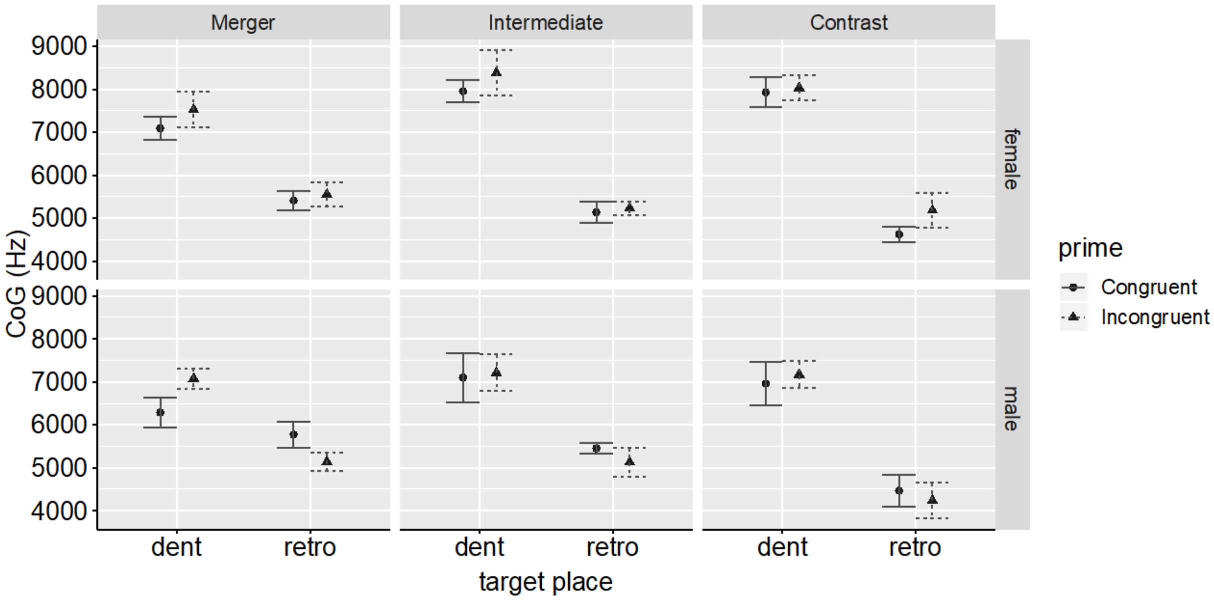


Fig. 1 Mean CoG values (Hz) of the target sibilants presented with congruent vs. incongruent primes. Participants were divided into three groups based on whether they conveyed sibilant contrasts in the regular reading task.

References

[1] Chung, K. S. (2006). Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication*, 16(2), 197–214.

[2] Chuang, Y. Y. & Fon, J. (2010). The effect of prosodic prominence on the realizations of voiceless dental and retroflex sibilants in Taiwan Mandarin spontaneous speech. *Proceedings of the 5th International Conference on Speech Prosody*.

[3] Shih, Y. T. (2012). Taiwanese-Guoyu bilingual children and adults' sibilant fricative production patterns. Doctoral dissertation, The Ohio State University.

[4] Wu, S. & Ma, Z. (2017). Native-language phonological interference in early Hakka–Mandarin bilinguals' visual recognition of Chinese two-character compounds: Evidence from the semantic-relatedness decision task. *Journal of Psycholinguistic Research*, 46 (1), 57–75.

[5] Yu, A. (2005). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology*, 24, 187–214.

Experimental evidence for perceptual hypercorrection in American r-dissimilation

Nancy Hall, Bianca Godinez & Megan Walsh

California State University Long Beach (USA)
nancy.hall@csulb.edu

The cause of phonological dissimilation is much debated; it is often described as an “unnatural” process, as opposed to the articulatorily natural tendency towards assimilation. Ohala 1993 [1] proposes that dissimilation originates from perceptual hypercorrection for assimilation. Certain features, such as rhoticity, affect acoustics across several syllables. This drawn-out realization can cause perceptual masking of similar nearby sounds. For example, in American *surprise* /sɚpɹaɪz/, listeners may misinterpret the rhoticity of the first vowel as anticipatory assimilation to the later rhotic, and posit the representation /səpɹaɪz/. This has in fact become a common alternate pronunciation.

Although widely cited, Ohala’s proposal has rarely been empirically tested. As Garrett & Johnson 2011:21 [2] note, “there are almost no controlled observations suggesting that listeners hypercorrect in speech perception.” It has proven difficult to produce perceptual dissimilation in laboratory settings [3,4].

Experiment 1: Perception of 1 vs 2 /r/s.

We tested perception of nonce words that mimic the typical environments for /r/-dissimilation in American English. We created 26 pairs of stimuli by splicing naturally produced syllables containing unstressed /r/ to two different continuations: one that contained /r/ and one that did not (as below). We predicted that listeners would be more likely to miss the presence of the first /r/ when spliced to a continuation with another /r/ (Figure 1).

The nonce words were presented to listeners in naturalistic sentences accompanied by a picture. This method of presentation was designed to mimic the experience of acquiring a new word from conversation. Listeners typed the unfamiliar word, spelling it however they heard it. 20 listeners were divided into two groups, where each group heard half of the stimuli with two /r/s and half with one /r/.

As predicted, the target /r/ was omitted from the written response significantly more often when the continuation also contained /r/: 13/260 times in words with 2 /r/s, versus 4/260 times in words with 1 /r/ ($\chi 2 = 4.9$, $p = .027$). Moreover, the patterns of perceptual error largely followed the tendencies of actual American spoken /r/-dissimilation. For example, /r/-deletion was more common if the two /r/s were in adjacent syllables. We interpret this as evidence that Ohala’s perceptual hypercorrection theory is viable, although we acknowledge other possible explanations of the results as well, in particular, the possibility that listeners apply a grammatical dissimilation rule.

Experiment 2: Effect of intervening r-coarticulation.

In a second experiment currently underway, we attempt to distinguish between grammatical and perceptual explanations by testing whether the presence or absence of rhotic coarticulation on syllables intervening between two /r/s affects the rate of dissimilation. In a nonce word like [maɹˈnɪkjəlɚ], we predict that if we splice in a token of [nɪkjəl] that was extracted from a word without rhotics (e.g. from [məˈnɪkjələ]), the absence of rhotic coarticulation on those syllables should make listeners less likely to interpret the first [ɹ] as anticipatory coarticulation with the [ɚ]. The grammatical approach to dissimilation would predict no effect.

This experiment uses 35 sets of words, each set including 4 conditions. A target /r/ is spliced to continuations with or without another /r/ (trigger and control conditions), with intervening portion with or without /r/-coarticulation, as shown in Figure 2.

We predict that the presence of a second /r/ and the presence of intervening /r/-coarticulation will each decrease recognition of the first /r/. We will present results of this experiment and discuss their implications for or against the perceptual hypercorrection hypothesis.

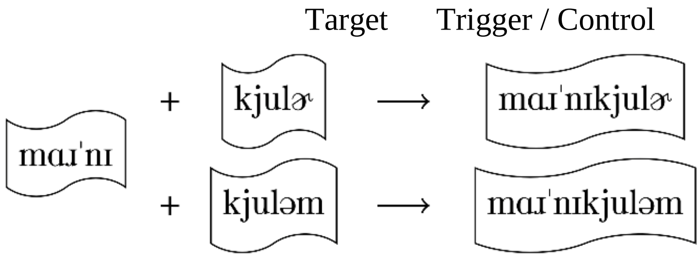


Fig.1

<table>
<tr><th>Target</th><th>Middle</th><th>Trigger / control</th></tr>
<tr><td rowspan="4">maɹ</td><td rowspan="2">ˈnɪkjəl
(with r-coarticulation)</td><td>ɚ</td></tr>
<tr><td>əm</td></tr>
<tr><td rowspan="2">ˈnɪkjəl
(without r-coarticulation)</td><td>əm</td></tr>
<tr><td>ɚ</td></tr>
</table>

Fig. 2

References

[1] Ohala, J. J. (1993). The phonetics of sound change. *Historical linguistics: Problems and perspectives*, 237-278.

[2] Garrett, A., & Johnson, K. (2011). Phonetic bias in sound change. In *UC Berkeley Phonology Lab Annual Reports*

[3] Abrego-Collier, C. (2011). Liquid dissimilation as listener hypocorrection. In Annual Meeting of the Berkeley Linguistics Society (Vol. 37, No. 1, pp. 3-17).

[4] Harrington, J., Kleber, F., & Stevens, M. (2016). The relationship between the (mis)-parsing of coarticulation in perception and sound change: evidence from dissimilation and language acquisition. In *Recent advances in nonlinear speech processing* (pp. 15-34). Springer, Cham.

The Influence of Native Intonational and Tonal Categories on Nonnative Tone Learning

Jessica Chin & Mark Antoniou

The MARCS Institute for Brain, Behaviour and Development, Western Sydney University (Australia)
j.chin@westernsydney.edu.au, m.antoniou@westernsydney.edu.au

As we age, it becomes increasingly difficult to acquire nonnative linguistic features. The native language influences how nonnative speech sounds and prosodic patterns are perceived. With training, however, perception of nonnative linguistic features can improve.

Lexical tone is a linguistic feature that distinguishes the meanings of words using pitch. Tone languages vary in the number of tones within their inventories, and individual tones may differ in height, direction and pitch trajectory. Tone language experience modulates perception of nonnative tone contrasts [1], [2], sometimes facilitating it, other times interfering with it [3]–[5]. Ease of tone learning may depend on how nonnative tones map onto L1 tonal or intonational categories [6]. But, it is not well understood how the complexity of the native tone system influences nonnative tone perception. Complexity has been defined as the number of tones within a tone system [7], but has also referred to similarity of pitch slopes and the presence of level, contour and/or checked tones.

Here, we examined if tone language experience facilitates nonnative tone learning, and whether experience with a more complex tone system confers additional benefit. We compared native speakers of nontonal (Australian English) and tonal languages (Mandarin Chinese and Vietnamese) in their ability to learn the tones of Meixian Hakka, a language with four regular and two checked tones [8]: Tone 1 (33) is a mid-level tone, tone 2 (11) is low-level, tone 3 (41) is mid-falling, and tone 4 (51) is high-falling [9]. The first checked tone (55) is high-level, and the second (41) is mid-falling. Permissible stop codas include /p/, /t/, /k/, and can appear in the VC or CVC syllable contexts [9].

Hakka differs from the tone languages of our participants. Mandarin has four tones: high-level (55), mid-rising (35), low-dipping (214) and high-falling (51) [10], and no checked tones. Southern Vietnamese has five tones: mid-level, low-falling, mid-rising, low falling-rising and falling-rising. Checked (mid-rising and low falling-rising) tones appear when a syllable ends in /p/, /t/ or /k/ [11].

Participants completed five sessions of tone word training, held on separate days. In session 1, participants completed a demographic questionnaire, a tone identification pre-test, and their first training session. Participants repeated the training task in sessions 2 to 4. In session 5, they completed their final training session, a generalisation task and a tone identification post-test.

The Mandarin (MDiff = 21.8, 95% CI [12.1, 30.5], $p < .001$) and Vietnamese (MDiff = 10.4, 95% CI [1.5, 19.3], $p = .017$) speakers identified Hakka tones more accurately than English speakers, and the Mandarin speakers also outperformed the Vietnamese speakers (MDiff = 11.4, 95% CI [2.3, 20.5], $p = .009$). The interaction between test and language was not significant ($p = .632$).

In tone training, all groups improved across sessions. The Mandarin speakers outperformed the English speakers (MDiff = 22.3, 95% CI [9.9, 34.7], $p < .001$), while the English and Vietnamese groups did not differ (MDiff = 10.6, 95% CI [-2.1, 23.3], $p = .130$). The difference between the Mandarin and Vietnamese groups was only marginally significant (MDiff = 11.7, 95% CI [-1.3, 24.7], $p = .09$). A similar pattern was observed for generalisation to a novel talker.

In identification, patterns of tone confusions revealed that all groups consistently identified the correct tone, but tonal language speakers were most accurate. For instance, English speakers identified level tones 2, 5 and 6 quite successfully, but struggled with other tones (63-69%). Mandarin speakers identified falling tones 3 and 4 slightly less accurately (79-84%), while Vietnamese speakers misidentified tones 2, 3, 4 more than other tones (70-79%). Accuracy decreased further for certain contrasts, and these contrasts differed across groups. For instance, English speakers identified tones 1 and 3 only 66% and 52% of the time when distinguishing between the two, and tones 1 and 4 63% and 56% of the time. Mandarin speakers correctly identified tones 3 and 4 71% and 69% of the time, while Vietnamese speakers correctly identified tones 2 and 3 72% and 66% of the time.

The findings show that although native speakers of nontonal and tonal languages benefit from tone training, tonal language speakers showed an advantage from the commencement of training. Further, the learning of nonnative tones does not seem to depend on the complexity of the native tonal language, but rather how well the tones map onto native tonal or intonational categories.

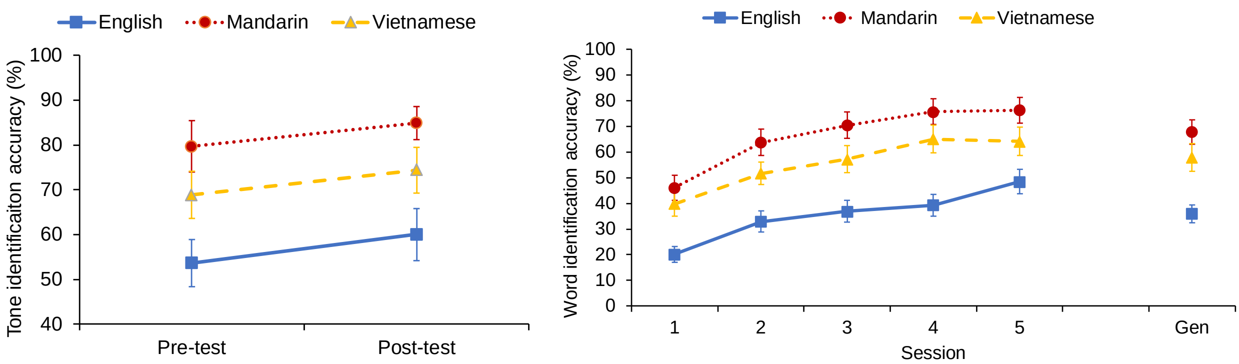


Fig.1 Left panel: Tone identification accuracy (%) at pre- and post-test. Right panel: Tonal (Mandarin, Vietnamese) and nontonal (English) learners' word identification accuracy (%) across five training sessions and generalisation test (Gen). Error bars depict SEM.

References

[1] Wayland, R., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *J. Phon.*, *36*(2), 250–267.

[2] Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Appl. Psycholinguist.*, *24*(1), 113–129.

[3] Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J. Exp. Psychol. Hum. Percept. Perform.*, *32*(1), 97–103.

[4] So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Lang. Speech*, *53*(2), 273–293.

[5] Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *Mod. Lang. J.*, *97*(1), 144–160.

[6] Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *J. Phon.*, *36*(2) 268–294.

[7] Tong, X., & Tang, Y. C. (2016). Modulation of musical experience and prosodic complexity on lexical pitch learning. Presented at *Speech Prosody*, Boston, USA, pp. 217–221.

[8] Hashimoto, M. J. (2010). *The Hakka dialect: A linguistic study of its phonology, syntax and lexicon*. Cambridge: Cambridge University Press.

[9] Cheung, Y. M. (2011). Vowels and tones in Meixian Hakka: An acoustic and perceptual study. PhD dissertation, City University of Hong Kong, Hong Kong.

[10] Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

[11] Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *J. Phon.*, *37*(1), 79–96.

A Sociophonetic study of Stop Consonants in Nepali

Neelam Singh

The English and Foreign Languages University (Hyderabad, India)
neelam@efluniversity.ac.in

The present study accounts for the fundamental frequency (f0) perturbations of stop types in Nepali. Nepali is an Indic language spoken in Nepal and in the north-eastern states of India. The study examines the stop consonants of Nepali spoken in the Maram region of Manipur, India. Nepali exhibits a 4-way laryngeal contrast; voiceless stops (VLS), voiceless aspirated stops (VLAS), voiced stops (VS) and voiced aspirated stops (VAS). The study examines the social factors: age, gender, level of proficiency, intra-inter-lingual contact and change due to extensive language contact with Meitei, a tonal language. Cross linguistically, voiced obstruents have been found to lower f0 in the following vowel (House and Fairbanks 1953, Hombert 1979, Clements 2002 and Christovich 1969). This lowering has been attributed to physiological and phonetic factors by few (Stevens2000, Atkinson 1978, and Honda 2004), while some argue that f0 lowering following voiced obstruents serves to maintain a phonological contrast between voiced and voiceless obstruents (Ohde 1984, Kingston & Diehl 1994, and Svantesson & House 2006).

The results indicate that Nepali speakers maintain the f0 perturbation patterns expected in 4-way laryngeal contrast, despite being in contact with Meitei for a very long time. Our findings lend support to the claim that while physiological and phonetic factors explain the expected f0 perturbation, however, long standing contact with tonal languages fail to provide any significant interaction on the speakers' f0. Following an extensive language background questionnaire, Nepali speakers were divided into four levels of proficiency in Meitei; Very High, High, Medium and Low. Three repetitions of each word in a frame sentence were recorded under focal and non- focal conditions and analyzed using Praat. Time-normalized f0 contours were measured for 10 intervals into the vowel and these measures were subjected to z-score normalization to reduce subject effects. A linear mixed effects model was used to analyze the data with R (R Development Core Team 2009). In our model the fixed effects included laryngeal setting (VS and VAS) and level of proficiency (LoP), and item and subjects acted as random effects. Likelihood ratio tests were performed to compare the fixed effects model with the random effects null models. Our results indicate that f0 perturbation patterns follow universal claims, in that VAS lowers f0 more than VS. However, despite sociophonetic conditions like very high level of contact and proficiency in a tonal language fail to impact the regular f0 pattern following VS and VAS of Nepali speakers. These findings lend support to the claim that physiological and phonetic factors determine f0 perturbation in the following vowel but sociophonetic conditions, such as high level of proficiency fail to provide any significant effect on the diasporic Nepali speakers'f0 with regard to the two stop type (VS and VAS) in the present context.

References

[1] Atkinson, J.E. 1978. Correlation analysis of the physiological factors controlling fundament frequency. *Journal of the Acoustical Society of America*. 63,211-222.

[2] Chelliah, S., 2003.Meithei. In Graham Thurgood and Randy LaPolla eds., *The Sino-Tibetan Languages*, 427-438. Surrey, England: Curzon Press.

[3] Chelliah, S. L. 2005. A study of Manipuri grammar. *Doctoral dissertation,* UMI Ann Arbor.

[4] Chistovich, L.A. 1969. Variations of the fundamental voice pitch as a discriminatory cue for consonants. *Soviet Physics – Acoustics.* 14, 372-378.

[5] Hombert, J.M. 1979. Consonant types, vowel quality and tone. *In Tone: A Linguistic Survey.* V.A. Fromkin ed. Academic, New York, pp. 77- 111.

[6] Honda, K.2004. Physiological factors causing tonal characteristics of speech: from global to local prosody. *Speech Prosody 2004*. Nara, Japan.

[7] House, A. & Fairbanks, G.1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America. 25:1, 105113.*

[8] Kingston, J & Diehl, R. 1994. *Phonetic knowledge. Language.* 70:3, 419- 454.
[9] Ohde, R. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America*. 75, 224- 230.
[10] R Development Core Team. 2009. R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna,Austria.* ISBN 3- 900051-07-0, URL http://www.R-project.org.
[11] Stevens, K. N. 2000. *Acoustic Phonetics*. MIT Press.
[12] Svantesson, J-O. & House, D.2006. Tone production, tone perception and Kammu tonogenesis. *Phonology*. 23, 309- 333.
[13] Thomas, Erik. 2011. *Sociophonetics- An Introduction. New York:* Palgrave Macmillan.

Ganong effects for lexicality but not for frequency

Stephen Politzer-Ahles[1], Leon Ka Keung Lee[1]

[1]*The Hong Kong Polytechnic University (Hong Kong)*
sjpolit@polyu.edu.hk, leon-leechk@hotmail.com

Ambiguous sounds are often judged in a way that would yield a real word: for example, when identifying a token that is ambiguous between /d/ and /t/, participants are often more likely to choose /t/ when the sound is in the context *_arp* (where identifying it as /t/ yields the real word *tarp*, whereas identifying it as /d/ yields the nonword **darp*) than when the sound is in the context *_am* (where identifying it as /t/ yields the nonword **tam*, whereas identifying it as /d/ yields the real word *dam*). This top-down influence on ambiguous sound judgment is the Ganong effect [1]. One might expect frequency to exert a similar top-down influence, i.e., people might identify the sound as /t/ more often in the context *_ime* (where it would yield a relatively high-frequency word *time*, whereas identifying the sound as /d/ would yield a lower-frequency word *dime*), and less often in the context *_or* (where it would yield a word relatively low-frequency word *tore*, whereas identifying the sound as /d/ would yield a higher-frequency word *door*).

Such an effect was observed in an experiment by Connine and colleagues [2]. While that experiment used a large sample of continua and edited the stimuli in a systematic and well-controlled way, one potential source of variance in the findings is that the continua biased towards or against aspirated identifications were from separate recordings (e.g., recordings of *deem* and *team* for a continuum where /t/ responses are more likely, and recordings of *dear* and *tear* for a continuum where /d/ responses are more likely). Thus, in spite of the careful and systematic manipulation of the stimuli, there may be low-level acoustic differences between those continua. To control for such differences, Shen and Politzer-Ahles [3] used bisyllabic Mandarin stimuli so that the exact same recordings could be used in different continua. For example, *duìhuà* ("conversation") is a higher-frequency word than *tuìhuà* ("degeneration"), whereas *duìyì* ("play chess") is a lower-frequency word than *tuìyì* ("retire). Shen and Politzer-Ahles [3] made one aspiration continuum with the syllable {*d*/*t*}*uì*, and spliced different second syllables onto it, in order to create continua with identical acoustics on the critical syllables but with different top-down biases. Their results replicated those of Connine and colleagues [2], but the careful acoustic control came at the expense of generalizability: they only used two continua, whereas Connine and colleagues [2] used 46. To test the generalizability and replicability of a frequency-based Ganong effect using acoustically matched stimuli, our pre-registered study (https://osf.io/6e35g/) adopts the same design as Shen and Politzer-Ahles [3] but with more stimuli; we also include word-nonword continua as a control condition to test whether the lexicality-based Ganong effect [1] is replicated.

The continua we used are shown below. For each three place of articulation, one item (two continua) was used to test the frequency-based Ganong effect and one item (two continua) to test the lexicality-based Ganong effect. Stimuli were recorded by a female speaker of standard Mandarin, and aspiration continua were created by taking aspirated tokens of each first syllable (*pì*, *tàn*, and *kuān*) and systematically removing 5 ms of aspiration at a time. Different second syllables were then spliced to these to yield the different continua, and the categorical boundaries were identified with a discrimination test with four Mandarin speakers. Tokens at the boundary and up to 4 steps before or after the boundary were used, resulting in 9-step continua.

	bilabial	alveolar	velar
Aspirated yields a high-frequency word	{*p*/*b*}*ìhuà*	{*t*/*d*}*ànwàng*	{*k*/*g*}*uānchǎng*
Aspirated yields a low-frequency word	{*p*/*b*}*ìjìng*	{*t*/*d*}*ànshì*	{*k*/*g*}*uānxīn*
Aspirated yields a real word	{*p*/*b*}*ìrú*	{*t*/*d*}*ànsuǒ*	{*k*/*g*}*uānróng*
Aspirated yields a nonword	{*p*/*b*}*ìmiǎn*	{*t*/*d*}*àngāo*	{*k*/*g*}*uāndiǎn*

Results from 53 participants (out of a planned 70) are shown below. It is clear that, while there is a strong Ganong effect for the lexicality manipulation (*b*=0.76, *z*=5.61, *p*<.001 in a binomial mixed-effects model), there is no significant frequency-based Ganong effect (*b*=.01, *z*=0.28, *p*=.780). The size of the frequency-based Ganong effect was not reliably moderated by reaction time (*b*<0.01, *z*=0.09, *p*=.925), which is consistent with Connine and colleagues [2]. There is suggestive evidence that it was moderated by place of articulation ($\chi^2(2)=8.00$, *p*=.018), but only to the extent that the frequency-based Ganong effect was very marginal in alveolar stimuli (*b*=0.13, *z*=1.70, *p*=.090) and substantially lower in velar and, particularly, labial stimuli.

The present results suggest that, contrary to previous reports, word frequency may not have the same kind of top-down effect on ambiguous speech sound categorization as lexicality does. The apparent frequency effect observed by Shen and Politzer-Ahles [3] may have been due to other, unknown confounding factors in the item used in that study. It is less clear how the results from Connine and colleagues [2] and the present experiment can be reconciled. One possibility is that frequency-based Ganong effects occur in English but not Chinese—although we had no *a priori* reason to hypothesize such a cross-linguistic difference—or that they occur in one-syllable simplex words but not in two-syllable compound words. Other possibilities are that the effects observed by Connine and colleagues [2] were due to low-level acoustic differences in the stimuli rather than to frequency (although it is unlikely that all 46 continua used in that study would have confounding acoustic differences in the same direction), or that the failure to observe a frequency Ganong effect in the present study represents a Type 2 or Type M error (although this is unlikely, given that the lexicality-based Ganong effect observed was so robust).

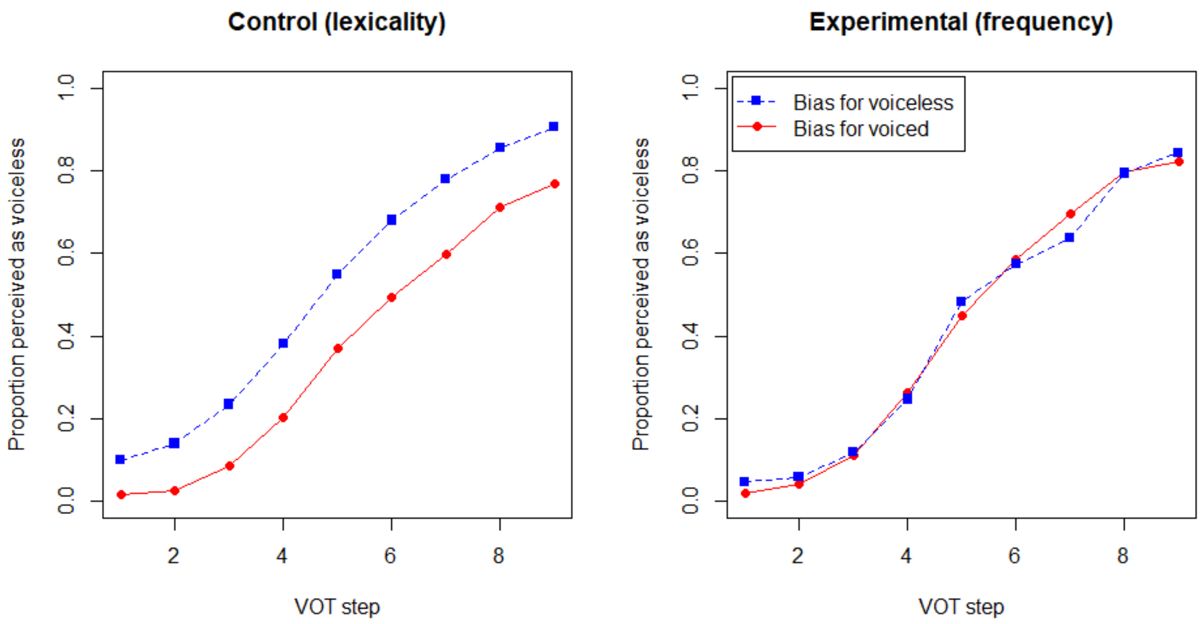


Fig.1 Results of the Ganong experiment for the lexicality contrast (left) and the frequency contrast (right). Where a Ganong effect is present, the blue dashed line with square markers should be higher than the red solid line with round markers.

References

[1] Ganong, W. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110-125.

[2] Connine, C., Titone, D., & Wang, J. (1993). Auditory word recognition: extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 81-94.

[3] Shen, L., & Politzer-Ahles, S. (2018). Analysis of the influence of word frequency in auditory perception. *HISPhonCog 2018*.

The Effect of Morpho-syntax and Prosodic Boundary on Taiwanese Min Juncture Tones

Ho-hsien Pan[1], Hsiao-tung Huang[1] & Shao-ren Lyu[1,2]

[1]*National Chiao Tung University (Taiwan)*, [2]*McGill University (Canada)*

hhpan@faculty.nctu.edu.tw, ferna823@gmail.com, shaorenlyu@gmail.com

Morpho-syntax and prosody interface to determine juncture tone production in Taiwanese Min, a language known for its chain-shift tone sandhi rules, 55, 13 → 33 → 31 → 53 → 55 and 5 → 3 → 5. Within the domain of a tone sandhi group (TSG), the final syllable carries juncture tone, whereas the non-final syllables carry sandhi tones. According to the tone sandhi rules, in /pɛʔ5 tsai31 **lɔ53**/ [pɛʔ3 tsai31 **lɔ53**] "cabbage stew," 白菜滷, the syllable /tsai31/ [tsai53] surfaces with a sandhi tone, whereas the syllable /lɔ53/ [lɔ53] surfaces with a juncture tone. However, in /lɔ53 pɛʔ5 **tsai31**/ [lɔ55 pɛʔ3 **tsai31**] "stew cabbage," 滷白菜, the syllables /tsai31/ and /lɔ53/ surface with juncture tone [tsai31] and sandhi tone [lɔ55], respectively. The surface tone values may change according to the syllable positions in a TSG domain. To date, prosodic nature of TSG has not been investigated fully. It has been proposed that the TSG domain is determined by syntactic domain and that a juncture tone occurs before the domain of an XP that acts as an argument or adjunct [1, 2, 3, 4]. Furthermore, only juncture tones were found to occur before the morpho-syntactic attribute modification marker "的" [ɛ] which carries a neutral tone (0) in a nominal phrase [XP ɛ YP]. For example, by placing the modification marker, [ɛ0], after the XP noun phrase /hai53 bĩ55/ "seashore," the noun phrase /hai53 **bĩ55** ɛ 0/ modifies the following YP noun phrase / ts^hu 31/ "house" as in /hai53 **bĩ55** ɛ0 ts^hu 31/ [hai55 **bĩ55** ɛ0 ts^hu31] "beach house." The syllable /bĩ55/ [bĩ55] before the modification marker, [ɛ0], carries juncture tone.

Aside from morpho-syntax, prosodic boundary was also found to affect juncture tone production [5]. There were more juncture tones than sandhi tones before high level intermediate phrase (ip) and intonation phrase (IP) boundaries. On the contrary, there were fewer juncture tones than sandhi tones before syllable and word boundaries.

This study investigates the effects of prosodic boundaries, including word, ip and IP, on the juncture tone production, along with the effect of morpho identity, including modification marker /ɛ0/, and final particles /a0, la0, hõ, ɛ0, o0, ma0, nɛ0, hẽ0, ba0/ with frequencies over 100, on the occurrence of juncture tones. The juncture to sandhi ratios of penultimate syllables preceding modification marker or final particle before word, ip and IP boundaries (__/a0, la0, hõ, ɛ0, o0, ma0, nɛ0, hẽ0, ba0/+ word, ip and IP) were investigated. Adult spontaneous corpus (TaiMinSS, www.taimin.tw) was used. TaiMinSS contains 30 monologues elicited from 40 speakers, female and male, of six dialect regions. The speakers were either above 40 years of age, or under 30 years of age. Monologues were transcribed using Praat at an utterance tier, an orthography tier with Chinese characters, a word tier with SAMPA symbols parsed for each word, an underlying tone tier, a surface tone tier with actual tones produced, a syllable tier, a segment tier reflecting the actual segmental pronunciations, a linguistic tier and a miscellaneous tier. Results showed that, first, there were more juncture tones than sandhi tones before the modification marker and before final neutral tones as well (Figure 1). Second, results of linear mixed effect regression models (morphemes × prosodic boundaries) with speakers as a random factor showed that juncture to sandhi tone ratios (J/S) were significantly lower before word boundary than before ip and IP boundaries. The J/S ratios before the modification marker /ɛ0/ were significantly higher than those preceding final particles at word, ip or IP domain-final positions. Third, the J/S ratios were significantly lower before word boundary than before ip or IP boundaries. It is proposed that neutral tone, morpheme identity and prosodic boundary have profound impacts on J/S ratios of penultimate syllables.

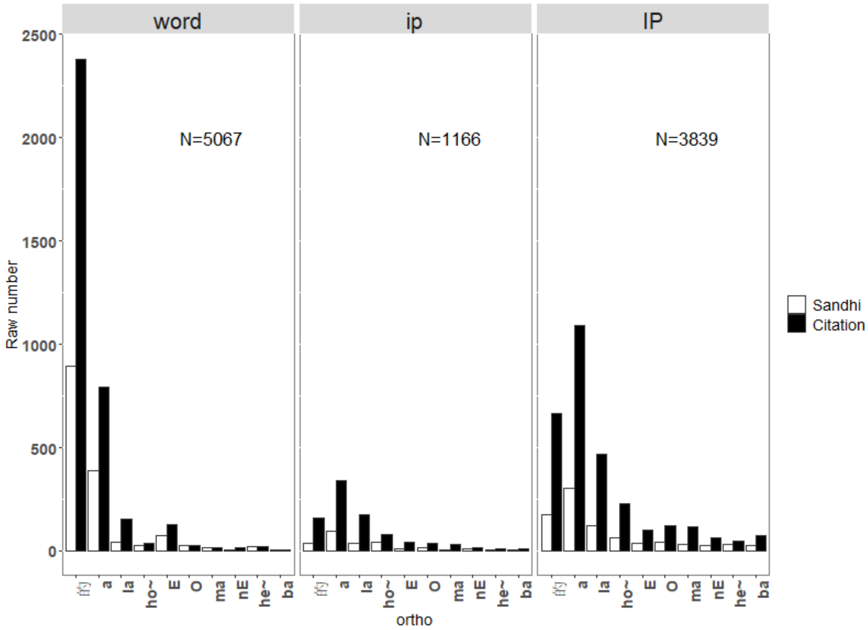


Fig.1 The numbers of juncture (citation) tones and sandhi tones at penultimate positions preceding modification marker ”的” /ɛ0/ and final particles /a0, la0, hõ, ɛ0, o0, ma0, nɛ0, hẽ0, ba0/ at word, ip and IP domain-final positions.

References

[1] Cheng, R. (1968). Tone sandhi in Taiwanese. *Linguistics*, *41*, 19-42.

[2] Chen, M. Y. (1987). The syntax of Xiamen tone sandhi. *Phonology Yearbook, 4,* 109-150.

[3] Chen, M. Y. (2000). *Tone Sandhi Patterns across Chinese Dialects*. Cambridge, UK: Cambridge University.

[4] Lin, J. (1994). Lexical government and tone group formation in Xiamen Chinese, *Phonology*, *11*, 237-275.

[5] Pan, H. H. (2019). Prosodic and syntactic interface on Taiwanese Min sandhi and juncture tones, *Proceedings of International Congress on Phonetic Science(ICPhS).* Melbourne, Australia.

Co-articulation between consonant and vowel in Taiwanese

Wai-Sum Lee[1], Feng-fan Hsieh[2] & Yueh-chin Chang[2]

[1]*Department of Linguistics and Translation, City University of Hong Kong (Hong Kong)*
[2]*Graduate Institute of Linguistics, National Tsing Hua University (Taiwan)*
w.s.lee@cityu.edu.hk, ffhsieh@mx.nthu.edu.tw, ycchang@mx.nthu.edu.tw

1. Introduction. This study investigates the co-articulation between the consonant and vowel in Taiwanese bisyllabic words that contain the sequences of [V**.**CV] and [VC**.**V], where the intervocalic C may be the onset of the following syllable, or the coda of the preceding syllable. It determines (*i*) the degree of consonant-vowel and vowel-consonant co-articulation during the four types of sequences, [**.**CV], [VC**.**], [C**.**V], and [V**.**C], by examining the changes in tongue position with respect to the different consonant and vowel types in the sequences and (*ii*) the difference in the degree of co-articulation between the tautosyllabic sequences, [**.**CV] and [VC**.**], and the heterosyllabic ones, [C**.**V] and [V**.**C].

2. Method. Speech samples of 46 meaningful Taiwanese bisyllabic words were elicited from three male adult speakers in their early 20s who were born into and grew up in Taiwanese-speaking families in Taiwan. The test bisyllabic words contain the [V**.**CV] and [VC**.**V] sequences, where C = [p t k] and V = [i a u]. Ten repetitions of each test bisyllabic word were digitally recorded of the speakers. The NDI Wave was used to collect synchronized audio signals and articulatory actions at three different points on the tongue (henceforth, tongue points), which are tongue blade, anterodorsum, and posterodorsum. To determine the degree of consonant-vowel and vowel-consonant co-articulation, Euclidean distance (henceforth, ED) was computed (as performed in the other co-articulatory studies, e.g., [1]) at the three tongue points (*i*) between the positions at the vowel onset and vowel mid-point during the [**.**CV] and [C**.**V] sequences and (*ii*) between the positions at the vowel mid-point and vowel offset during the [VC**.**] and [V**.**C] sequences. A large or small ED is taken to indicate a corresponding large or small degree of co-articulation between the neighboring segments.

3. Results. The average ED data across all the three tongue points for the four types of test sequences, [**.**CV], [VC**.**], [C**.**V], and [V**.**C], for three male Taiwanese speakers (M1, M2, M3) are presented in Table 1. A summary result is given as follows.

3.1. Tautosyllabic [.CV] sequences. (*i*) For [**.**pV], the ED is small (0.81-3.62 mm), regardless of the vowel type. The small ED suggests a minimal carryover effect of [p] on the articulation of the following vowel. (*ii*) For [**.**tV], there is an increase in the ED relative to the ED for [**.**pV]. The increase in ED is significant when V = [a] (10.42-12.88 mm) or [u] (9.67-13.18 mm), suggesting the antagonism between the articulatory gestures associated with the alveolar [t] and the following low vowel [a] or high back vowel [u]. (*iii*) For [**.**kV], the ED is small when V = [i] (0.77-1.20 mm) and large when V = [a] (6.05-11.26 mm), relative to the ED when V = [u] (2.16-5.97 mm). The data (*a*) indicate that [k] is palatalized before [i] and (*b*) suggest that there is a large carryover co-articulation effect of [k] on [a]. (*iv*) Overall, the orders of decreasing degree of co-articulation, as indicated in the average ED across the three tongue points for the [**.**CV] sequences for each of the three speakers (M1/M2/M3), are (*a*) when C = [t] (9.21/9.27/8.22 mm) > when C = [k] (3.32/6.14/4.41 mm) > when C = [p] (2.11/1.68/2.31 mm), and (*b*) when V = [a] (7.34/8.96/7.55 mm) > V = [u] (5.94/6.34/4.61 mm) > V = [i] (1.36/1.78/2.77 mm).

3.2. Tautosyllabic [VC.] sequences. (*i*) Compared with the ED for [**.**pV] (0.81-3.62 mm), there is an increase in the ED for [Vp**.**] (2.45-6.53 mm), suggesting the anticipatory co-articulation effect of [p] on the preceding vowel is larger than the carryover co-articulation effect of [p] on the following vowel. (*ii*) The ED is much large for [Vt**.**] (8.34-14.48 mm) and [Vk**.**] (6.21-9.57 mm), except for a single case of [it**.**] (2.22-4.11 mm), further suggesting a large degree of vowel-consonant co-articulation during the [VC**.**] sequences. The data appear to support the phonological structuring of the syllable, where the syllable-final consonant (the coda) and the preceding vowel (the syllable nucleus) are the constituent units of the rhyme. (*iii*) Overall, the orders of decreasing degree of coarticulation, as indicated in the average ED across the three tongue points for the [VC.] sequences for each of the three speakers (M1/M2/M3), are (*a*) when C = [t] (10.04/8.98/7.53 mm) or [k] (7.29/9.12/7.67 mm) > when C = [p] (5.25/3.86/2.74 mm), and (*b*) when V = [u] (13.43/11.21/10.15 mm) or [a] (9.53/8.08/6.64 mm) > when V = [i] (4.40/5.82/4.45 mm). Note that in Taiwanese the vowel [u] in the [VC.] sequences can only be followed by [t], but not [p] or [k].

3.3. Heterosyllabic [C.V] sequences. (*i*) Regardless of the consonant or vowel type the ED for the [C.V] sequences is small (not exceeding 3.60 mm) for all the three speakers. (*ii*) Compared with the ED for the [.CV] sequences, the ED is much smaller for [t.V] (0.50-2.89 mm) and [k.V] (0.60-2.72 mm) than for [.tV] (2.10-13.18 mm) and [.kV] (0.77-11.26 mm), due to the absence of alveolar or velar articulation at the onset of the following vowel in [t.V] and [k.V]. (*iii*) Note that in the [C.V] sequences, the burst release of the oral closure of the preceding coda plosive is followed by a short pause before V in the [C.V] sequences. The data suggest re-syllabification does not take place across the syllable boundary in the [C.V] sequences, resulting in a small degree of between-segment co-articulation in the [C.V] sequences.

3.4. Heterosyllabic [V.C] sequences. (*i*) Differing from the [C.V] sequences, the cross syllable boundary co-articulation effect occurs in the [V.C] sequences, as evidenced by an increase in the ED for the [V.C] sequences (2.05-23.55 mm) relative to the ED for the [C.V] sequences (0.50-3.59 mm). (*ii*) The ED is particularly large for [a.C] (5.94-23.40 mm), due likely to the articulatory antagonism between the oral closure for the following plosive consonant and the mouth opening during the preceding vowel [a]. (*iii*) The ED is also large for [u.C] when C = [t] (16.00-23.55 mm) or [k] (5.99-9.08 mm). (*iv*) The ED is significantly larger for the [V.C] sequences, [a.p], [a.t], [a.k], [u.t] and [i.t], than for the [VC.] sequences, [ap.], [at.], [ak.], [ut.] and [it.]. The data suggest that (*a*) C in the [V.C] sequences is ambisyllabic, i.e., functioning simultaneously as the coda of the preceding syllable and the onset of the following syllable, and (*b*) the syllable boundary does not block the vowel-consonant co-articulation, in contrast to the consonant-vowel co-articulation in the [C.V] sequences. (*v*) Overall, the orders of decreasing degree of co-articulation, as indicated in the average ED across the three tongue points for the [V.C] sequences for each of the three speakers (M1/M2/M3), are (*a*) when C = [t] (17.14/16.70/13.71 mm) > when C = [k] (9.67/10.52/10.23 mm) > when C = [p] (6.70/3.72/3.28 mm), and (*b*) when V = [a] (18.04/16.23/13.00 mm) > V = [u] (11.79/11.10/7.94 mm) > V = [i] (3.68/3.62/6.28 mm).

[.CV]	M1	M2	M3	**[VC.]**	M1	M2	M3	**[C.V]**	M1	M2	M3	**[V.C]**	M1	M2	M3
[.pi]	1.21	1.47	2.60	[ip.]	3.97	4.29	3.03	[p.i]	0.69	1.07	1.93	[i.p]	3.49	2.85	2.05
[.pa]	3.62	2.74	2.31	[ap.]	6.53	3.42	2.45	[p.a]	2.79	3.59	1.12	[a.p]	12.23	6.61	5.94
[.pu]	1.51	0.81	2.01	*	*	*	*	[p.u]	0.68	1.88	0.71	[u.p]	4.37	1.70	1.85
[.ti]	2.10	2.67	4.58	[it.]	2.22	3.59	4.11	[t.i]	1.11	0.71	1.55	[i.t]	4.82	4.18	5.93
[.ta]	12.34	12.88	10.42	[at.]	14.48	12.13	8.34	[t.a]	0.87	2.68	0.69	[a.t]	23.04	23.40	19.20
[.tu]	13.18	12.25	9.67	[ut.]	13.43	11.21	10.15	[t.u]	2.89	2.88	0.50	[u.t]	23.55	22.53	16.00
[.ki]	0.77	1.20	1.14	[ik.]	7.00	9.57	6.21	[k.i]	0.60	0.66	1.21	[i.k]	2.72	3.82	10.85
[.ka]	6.05	11.26	9.93	[ak.]	7.59	8.68	9.13	[k.a]	2.72	2.35	1.05	[a.k]	18.85	18.66	13.86
[.ku]	3.14	5.97	2.16	*	*	*	*	[k.u]	1.70	1.32	1.15	[u.k]	7.44	9.08	5.99

Table 1. Average ED (in mm) across the three tongue points (*i*) between the vowel onset and vowel mid-point for the [.CV] and [C.V] sequences and (*ii*) between the vowel mid-point and vowel offset for the [VC.] and [V.C] sequences for three male Taiwanese speakers (M1, M2, M3); * = non-occurring.

4. Conclusion. The articulatory data on the ED at three tongue points during the sequences of [.CV], [.VC], [C.V], and [V.C] in Taiwanese bisyllabic words reveal that the degree of co-articulation varies with the consonant type, vowel type, segment position in the syllable, and presence/absence of the syllable boundary in the consonant-vowel or vowel-consonant sequences.

Reference

[1] Recasens, D., Espinosa A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *JASA* 125(4), 2288-2298.

[*Acknowledgement*: This research is supported by an International Scholarly Exchange Research Grant (#9500023/RG003-P-14) from the Chiang Ching-kuo Foundation, Taiwan.]

Prosodic marking of focus in Korean learners of Mandarin

Jun Liu[1,2], Jing Wu[1] & Yong-Cheol Lee[1]

[1]*Cheongju University (Korea)*, [2]*Liaocheng University (China)*
1748046227@qq.com, wuqionganqi@163.com, soongdora@gmail.com

We conducted a production experiment using phone number strings to examine whether Korean learners of Chinese produce a native-like prosodic marking of focus. We found that both advanced and intermediate groups did not produce clear prosodic effects of focus nor noticeable post-focus compression, when directly compared to Mandarin native speakers. This study demonstrates that both advanced and intermediate groups had a strong negative prosodic transfer, derived from their first language (L1), although this negative transfer is known to decrease as their language proficiency increases [1]. This study also suggests that due to the interaction between tone and intonation [2], focus prosody in a tone language is not quite easy to acquire for non-tonal language speakers. Finally, the current study underlines that, as established in previous studies (e.g., [3]), post-focus compression is hard to transfer between languages.

Although both Mandarin and Korean are characterized by on-focus expansion and post-focus compression [4], the two languages differ greatly in at least two fundamental ways in focus marking. First, Mandarin is a tone language, and its prosodic focus is realized through expanding a pitch range conforming to its tonal structure [5]. Second, the degree of prosodic modulation of narrow focus, relative to broad focus, is shown to be greater in Mandarin than in Korean [6]. According to the Contrastive Analysis Hypothesis [7] for language learning, characteristics similar to L1 are easy to acquire, but those different from L1 are harder to acquire. Therefore, given that Mandarin and Korean differ considerably in terms of how prosodic focus is encoded, we posit that Korean learners of Mandarin will face a challenge in producing a native-like prosodic marking of focus. Furthermore, considering the interaction effect between tone and intonation, a tone language is acquired more slowly than a non-tone language. Accordingly, we also posit that even advanced Korean learners of Mandarin will differ quite from native speakers, showing negative prosodic features, particularly for post-focus compression.

100 ten-digit phone number strings were used in the form of (NNN)-(NNN)-(NNNN). The strings were randomly generated so that (1) each digit (0-9) appears equally often in each position of each digit string and a combination of every two digits appears equally often in each digit string. Two groups (advanced and intermediate) of Korean learners of Mandarin (2 females and 3 males in each group) produced the target stimuli embedded in two focus conditions: neutral and corrective focus. The neutral-focus stimuli were recorded in isolation as a background reading. In the corrective-focus condition, the same sequences were designed in a Q&A form, in which speakers made a response by correcting only one incorrect digit in the question (“No, Mary’s number is **7**87-412-4699”). Five native Mandarin speakers were also recruited for control data.

Among the four tone types in Mandarin, we only selected tone 1 and tone 4 digits (tone 1: 1, 3, 7, 8; tone 4: 2, 4, 6) for further analysis because the other tone digits are not sufficient among the ten digits (0-9) in a digit string. Our basic analysis was to make a direct comparison between broad focus and corrective focus by observing pitch patterns for each focus condition. Figure 1 displays time-normalized pitch contours of tone 1 and tone 4 in the two focus conditions. The pitch contours were averaged over the digit strings (**3**89-343-4492 for tone 1, **6**37-686-7664 for tone 4), produced by five speakers in each group. In the figure, “target” refers to a focus position, the area shaded in gray represents post-focus positions, and the dotted line indicates a phrase boundary, demarcated by hyphens in the digit string (NNN-NNN-NNNN). Figure 1 illustrates that corrective focus shows a higher pitch peak in the focus position and a lower pitch valley in the post-focus positions in the native group. However, both advanced and intermediate groups exhibit no such clear indication of prosodic changes in the focus and post-focus positions. The results demonstrate that even the advanced group produce neither clear on-focus expansion nor post-focus compression although their performance seems slightly better than the intermediate group, suggesting that the interference of L1 was evident in their L2 prosody for both groups. Finally, we view that, due to the interaction of tone and intonation in Mandarin, Korean learners of Mandarin seem to have particular difficulty in the acquisition of on-focus expansion and post-focus compression.

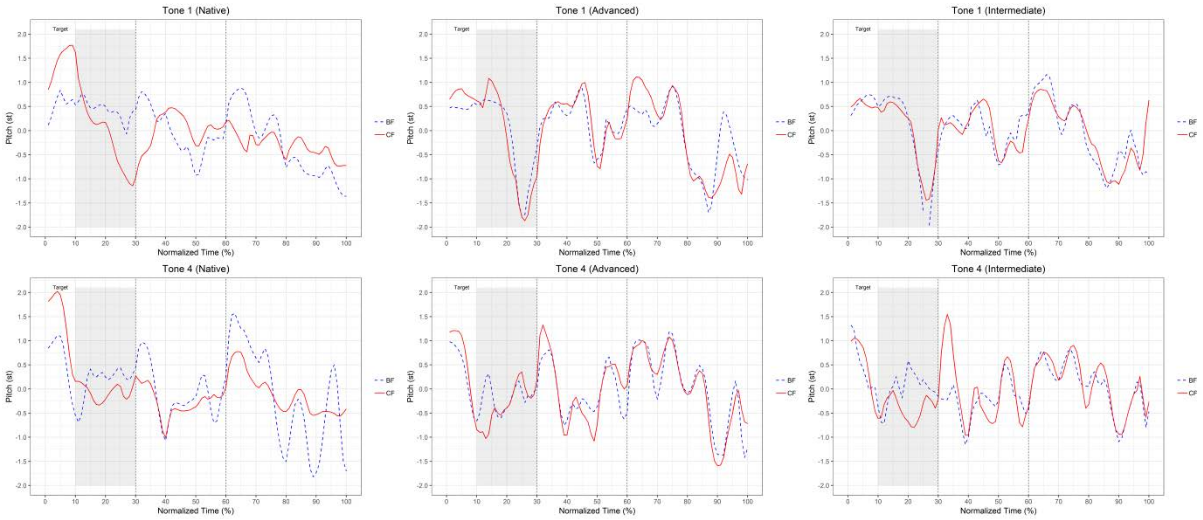


Fig. 1 Time-normalized pitch contours of the digit strings (389-343-4492 for tone 1, 637-686-7664 for tone 4) produced by five speakers of each group in the two focus conditions, separated by tone and language group.

References

[1] Swerts, M., & Zerbian, S. (2010). Prosodic transfer in Black South African English. In *Proceedings of Speech Prosody 2010* (pp. 100198: 1-4).

[2] Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

[3] Chen, Y., Guion-anderson, S., & Xu, Y. (2012). Post-focus compression in second language Mandarin. *Proceedings of Speech Prosody 2012*, 410–413.

[4] Xu, Y., Chen, S., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review*, *29*(1), 131–147.

[5] Lee, Y., Wang, T., & Liberman, M. (2016). Production and perception of tone 3 focus in Mandarin Chinese. *Frontiers in Psychology*, *7*(1058), 1–13.

[6] Lee, Y., Wang, B., Chen, S., Adda-Decker, M., Amelot, A., Nambu, S., & Liberman, M. (2015). A crosslinguistic study of prosodic focus. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015)* (pp. 4754–4758).

[7] Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press.

Phonemic and Featural Modeling of New L2 Sound Acquisition

Kakeru Yazawa[1] & Paola Escudero[2]

[1]*Waseda University (Japan),* [2]*Western Sydney University (Australia)*
k-yazawa@aoni.waseda.jp, paola.escudero@westernsydney.edu.au

It is widely accepted that nonnative speech perception patterns are predictable from phonetic and phonological similarities between the first (L1) and second (L2) language sound categories [1]. Previous L2 studies and models therefore often assume that phonemes are the most basic unit, but emerging evidence suggests that phonological features constituting the phonemes play an important role [2, 3]. This study conducts two kinds of computational modeling, phonemic and featural, which provide different kinds of explanations regarding how a new L2 sound is acquired.

Escudero and Boersma [4] proposed that perception of sound categories can be modeled by Optimality-Theoretic (OT) constraints such as "a value of *x* on the auditory continuum *y* should not be perceived as the phonological category *z*" (e.g., "[F1 = 800 mel] is not /a/"). Tableau 1 illustrates how such constraints can express the perception of a typical token of /a/ with the acoustic values of [F1 = 800 mel] and [F2 = 1100 mel]. In this example, the constraint "[F2 = 1100 mel] is not /e/" is ranked the highest, perhaps because an F2 of 1100 mel is too back for /e/. Thus, the alternative candidate /a/ is chosen as the winner ('perceived'). Escudero [1] argues that this kind of acoustic-to-phoneme modeling adequately explains perception.

In contrast, Boersma and Chládková [5] argue that acoustic-to-feature modeling better explains speech perception. They proposed that a sound category (e.g. /e/) should be expressed as a bundle of phonological features (e.g. /mid, front/). Feature co-occurrence is prohibited by constraints such as "*/mid, central/" whose strictness depends on the organization of features in the particular language [6]. Tableau 2 shows the perception of [F1 = 800 mel, F2 = 1100 mel] (same as Tableau 1) as /low, central/ (i.e. /a/). Here, the constraints "*/low, front/" and "*/mid, central/" are ranked high, perhaps because these combinations of features do not occur in this language. The third constraint "[F2 = 1100 mel] is not /front/" then prohibits /mid, front/, and thus /low, central/ is chosen as the winner.

I will now apply the two approaches to a real L2 learning scenario, namely L1 Japanese listeners' perceptual acquisition of L2 American English (AmE) vowels /ɛ, æ, ʌ, ɑ/. According to Strange et al. [7], these vowels are spectrally assimilated to either /e/ or /a/ in Japanese. The learner's task is to acquire new phonemes, or new sets of features, out of the existing ones to deal with the L2 listening environment. Phonemic and featural simulations were performed based on Stochastic Optimality Theory and the Gradual Learning Algorithm [8], using real acoustic values [9].

Figs. 1 and 2 show the results of L1 Japanese and L2 AmE simulations, respectively. Phonemic and featural modeling yielded very similar outcomes, but there was a crucial difference in how these models were implemented. First, the virtual phonemic listener learned to perceive two categories (/e, a/) in their L1 because these are the only possible categories, which they transferred to L2 perception. In order for new L2 category formation to take place, the model needed to be explicitly told that there were four categories (/ɛ, æ, ʌ, ɑ/) instead of two in the L2. This equals introducing new candidates (e.g. /æ/) and constraints related to the candidates (e.g. "[F1 = 800 Hz] is not /æ/") in terms of OT modeling.

On the other hand, the featural listener learned to perceive /mid, front/ (i.e. /e/) and /low, central/ (i.e. /a/) in their L1, even though perception of the other feature combinations (/mid, central/ and /low, front/) was also possible in principle. This is because constraints against infrequent feature co-occurrences are ranked high, while those against frequent co-occurrences are ranked low. As the virtual listener got exposed to L2 AmE in which /mid, central/ (i.e. /ʌ/) and /low, front/ (i.e. /æ/) do occur unlike the L1, they gradually lowered the ranking of "*/mid, central/" and "*/low, front/." The resultant model could perceive all four possible combinations of features: /mid, front/ (/ɛ/), /mid, central/ (/ʌ/), /low, front/ (/æ/) and /low, central/ (/ɑ/).

Although both phonemic and featural modeling can express the acquisition of new L2 sounds, featural modeling may provide an ecologically more valid account. The strength of featural modeling is that a new ‘category’ can emerge based on re-used L1 features, unlike phonemic modeling where new categories have to be intentionally added. Featural modeling may also explain why certain L2 sounds are perceived as poor exemplars of an L1 sound. For example, the featural L1 Japanese listener is aware that AmE /æ/ is /low/ and /front/, but nonetheless perceives it as /mid, front/ (/e/) or /low, central/ (/a/) because the /low/ and /front/ features do not occur simultaneously in Japanese.

Tableau 1 Perception of [F1 = 800 mel, F2 = 1100 mel] as /a/.

[F1=800 mel, F2=1100 mel]	[F2=1100] not /e/	[F1=800] not /e/	[F1=800] not /a/	[F2=1100] not /a/
/e/	*!	*		
☞/a/			*	*

Tableau 2 Perception of [F1 = 800 mel, F2 = 1100 mel] as /low, central/.

[F1=800 mel, F2=1100 mel]	*/low, front/	*/mid, central/	[F2=1100] not /front/	[F1=800] not /mid/	*/low, central/
/mid, front/			*!	*	
/mid, central/		*!		*	
/low, front/	*!		*		
☞/low, central/					*

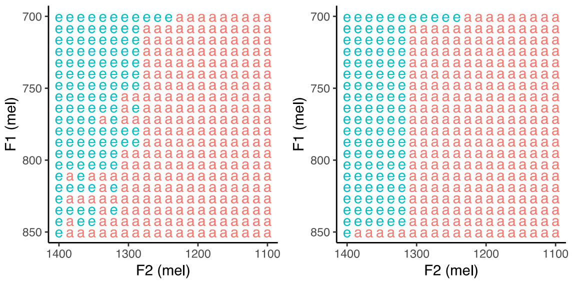


Fig. 1 Phonemic (left) and featural (right) simulations of L1 Japanese perception.

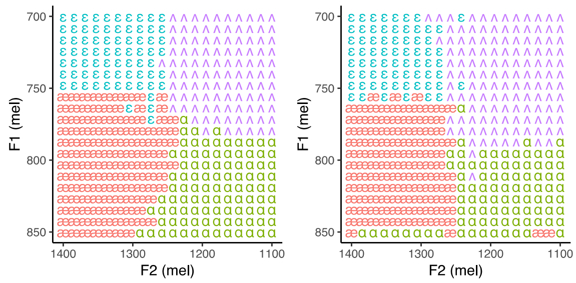


Fig. 2 Phonemic (left) and featural (right) simulations L2 AmE perception.

References

[1] Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Doctoral dissertation). Utrecht University.

[2] Balas, A. (2017). More evidence for selective attention to features: Perception of the Turkish high back unrounded vowel. In *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2017* (pp. 65–70).

[3] Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010.

[4] Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585.

[5] Boersma, P., & Chládková, K. (2011). Asymmetries Between Speech Perception and Production Reveal Phonological Structure. In *Proceedings of the 17th International Congress of Phonetic Science* (pp. 328–331).

[6] Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1013–1016).

[7] Strange, W., Akahane-Yamada, R., Kubo, R., Trent-Brown, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, *26*(4), 311–344.

[8] Boersma, P. (1998). *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives* (Doctoral dissertation.) University of Amsterdam.

[9] Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The Journal of the Acoustical Society of America*, *124*(1), 576–588.

Comparative Modelling of Speech Prosody: AM Theory vs. PENTA Model

Albert Lee[1] & Faith Chiu[2]

[1]*The Education University of Hong Kong (Hong Kong),* [2]*University of York (United Kingdom)*
albertlee@eduhk.hk, faith.chiu@york.ac.uk

Several rival models of speech prosody have coexisted for several decades. In the case where these competing models contribute to mutually exclusive proposals, it becomes necessary for the prosody researcher to directly compare and assess the models and the theories behind each of them. Having originated by accounting for phenomena from different language families, each of these models was thus proposed to serve different purposes (e.g. with the main focus of representing theoretical grammatical constructs, or to generate f_o curves identical to what is observed acoustically). As a result, the models also tend to vary on their fundamental assumptions, their f_o generating mechanisms, and the number and levels of input specification required. These differences have made it difficult for researchers to fully understand rival models well enough to assess them fairly, thus further contributing to their continuous coexistence.

Comparative modelling is one objective way of directly comparing models. By resynthesizing a given dataset based on the workings of several theories, their respective synthesis accuracy can serve as a gauge for fair assessment. For example, the Common Prosody Platform [1] is a recent endeavor intended to offer a user-friendly platform for such comparison.

In this paper, we propose four factors that users should take into consideration when comparing theoretical models based on synthesis accuracy. They are (i) the underlying targets that generate surface f_o contours, (ii) the level of target specification, (iii) the degree of freedom permitted in the number of tiers which contributes to information encoding of each target, and (iv) how they implement underlying units into generation of f_o contours.

To illustrate these points, we present the synthesis accuracy performance of a 6,400-sentence corpus of Japanese utterances [2] that contrasts in narrow focus condition (initial / medial / final / neutral), sentence type (declarative / interrogative), and lexical accent condition (initial accent / unaccented). We implemented speaker-dependent resynthesis and speaker-independent predictive synthesis based on two models: the Autosegmental-Metrical Theory (AM henceforth, e.g. [3]) and Parallel Encoding and Target Approximation Model [4].

We generated AM-styled and PENTA-styled annotation files (see examples overleaf) using a PRAAT script. These files were manually checked and rectified before being submitted to speaker-dependent resynthesis. Subsequently, we used the Jackknife procedure [5] to carry out speaker-independent predictive synthesis. On the whole, it was found that PENTA yielded better synthesis accuracy than AM. We take this to argue that PENTA's (iv) f_o-generating mechanism can achieve a better curve fit.

Nevertheless, AM and PENTA also differ in terms of (i) the nature of underlying targets, (ii) level of target specification, and (iii) degree of freedom in the number of tiers permitted to encode information. As it stands, there is no fair way to quantitatively compare models in these three aspects. It is concluded that the Common Prosody Platform has taken a crucial step forward for prosody research, with much that remains to be achieved, and that there are aspects of theories that awaits further developments before they could be assessed by computational or mathematical means.

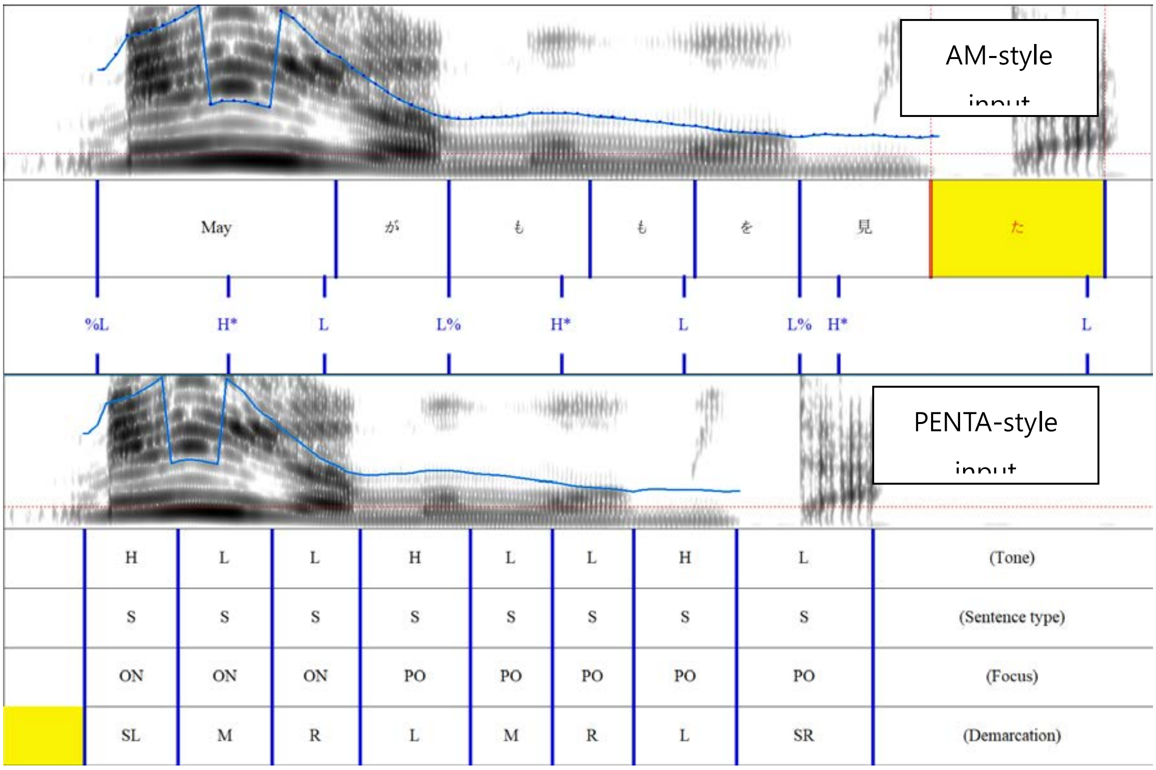


Fig.1 AM-style and PENTA-style annotation of speech data in preparation for model training.

References

[1] Prom-on, S., Xu, Y., Gu, W., Arvaniti, A., Nam, H., &Whalen, D. H. (2016). The Common Prosody Platform (CPP)—where theories of prosody can be directly compared. In *Proceedings of the 8th International Conference on Speech Prosody (SP2016)* (pp. 1–5). Boston, MA.

[2] Lee, A., &Xu, Y. (2018). Conditional realisation of post-focus compression in Japanese. In *Proceedings of the 9th International Conference on Speech Prosody (SP2018)* (pp. 216–219). Poznań, Poland.

[3] Pierrehumbert, J. B., &Beckman, M. E. (1988). *Japanese Tone Structure*. Cambridge, MA: Massachusetts Institute of Technology.

[4] Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, *46*, 220–251.

[5] Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, *43*, 353–360.

The Impact of Accidental Gaps on Tonal Categorization

Tzu-Hsuan Yang, Shao-Jie Jin & Yu-An Lu

National Chiao Tung University (Taiwan)
thy2105@tc.columbia.edu, courtney0419003.fl04@nctu.edu.tw, yuanlu@nctu.edu.tw

Previous studies have demonstrated that speech perception may be shaped by a speaker's linguistic knowledge. For example, Ganong [1] employed a continuum of stops varying in VOT, where one endpoint was a word and the other a non-word (e.g., *task*-**dask*, **tash*-*dash*), and asked English native speakers to identify the sound they heard as a [t] or [d]. The results showed that the boundary of categorical perception was biased towards the endpoint which was a word, reflecting an effect of lexical knowledge. Using a similar method, Massaro and Cohen [2] showed that English speakers gave more [r] responses on a [tr]-*[tl] continuum but more [l] responses on a *[sr]-[sl] continuum, reflecting a bias from phonotactic restriction. Other kinds of knowledge that are relevant to this phenomenon include phonological relationship (e.g., [3]), syllable structure (e.g., [4]) and morphological alternation (e.g., [5]). While previous studies have demonstrated a sensitivity to segmental features, there is little evidence of speech sound categorization being shaped by suprasegmental information. The present study aims to investigate the role of suprasegmental information in speech perception by examining the effect of accidental gaps on Mandarin speakers' tonal categorization.

Mandarin is a tone language with four phonemic tones: high-level Tone 1 [X^{55}], rising Tone 2 [X^{35}], falling-rising Tone 3 [X^{214}], and falling Tone 4 [X^{51}]. However, not every allowable syllable carries all four tones. For instance, the syllable [tsʰu] can be combined with T1 ([tsʰu]^{55} "coarse"), T2 ([tsʰu]^{35} "die" in Old Chinese), and T4 ([tsʰu]^{51} "vinegar"), but not with T3 (*[tsʰu]^{214}). These syllable-tone combinations that could but do not exist are termed "accidental gaps" (e.g., [6]). The present study investigates whether tonal categorization may be biased by the knowledge of accidental gaps.

To examine Mandarin speakers' tonal categorization, we conducted a forced-choice identification experiment in which 20 Taiwanese Mandarin speakers (16F, 4M; ages 20-37, M = 22) participated. Mandarin CV syllables carrying a tone along one of two continua were selected: T1-T2 (level to rising) or T1-T4 (level to falling), among which, tone pairs were further selected so that each pair contained a word on one end and an accidental gap on the other (*T1-T2, T1-*T2, *T1-T4, T1-*T4). The eight syllables were naturally produced and recorded by a phonetically trained female Taiwan Mandarin speaker. 10-step continua were resynthesized from these endpoints using TANDEM-STRAIGHT [7]. This method allows for the resynthesis to be done from the entire CV endpoints so that *f0*, as well as other secondary cues to tone perception (e.g., duration and creakiness) were manipulated proportionally [8, 9]. Figure 1 shows the *f0* trajectories and duration of the resynthesized stimuli on the continua (left: T1-T2; right: T1-T4). Participants were instructed to listen to each of the stimuli randomly presented to them in E-Prime [10] and judge whether they heard T1 or T2/T4 by pressing the corresponding key on the keyboard as soon as they were sure. Participants' responses and response times (RTs) were recorded.

Mixed-effects logistic regression models were fitted using the lme4 package in R [11] to interpret the results. The dependent variable was the participants' tonal categorization, with T1 responses coded as 0 and T2/T4 responses coded as 1. The models included Endpoint (e.g., T1-*T4, *T1-T4) and Step (1 to 10, normalized), and an interaction term for Endpoint and Step. The models also included random intercepts and slopes for Participant and Token. We fitted separate models for each tone continuum (T1-T2 and T1-T4). The results showed an effect of Endpoint along the T1-T4 continua (p=.002), indicated by the earlier boundary shift on the *T1-T4 continuum (solid line) than on the T1-*T4 continuum (dash line). The results along the T1-T2 continua also showed a similar trend (p=.06). The RT results provide further support to this perceptual bias, in that, first, the RTs peaked at different steps along the continua, with earlier peaks for the *T-T continua and later for the T-*T continua; second, although listeners successfully identified the gap syllables, they generally spent more time in making the decisions. To summarize, the findings in the present study showed that Mandarin listeners' responses were biased by accidental gaps. The results suggest that listeners' perceptual categorization is sensitive to segmental as well as suprasegmental (tonal) information.

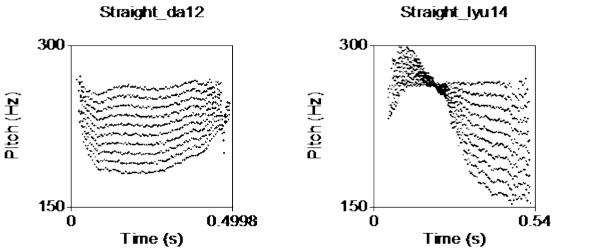


Fig.1 f0 trajectories of T1T2 continua for the token [ta] (left) and T1T4 continua for the token [ly] (right).

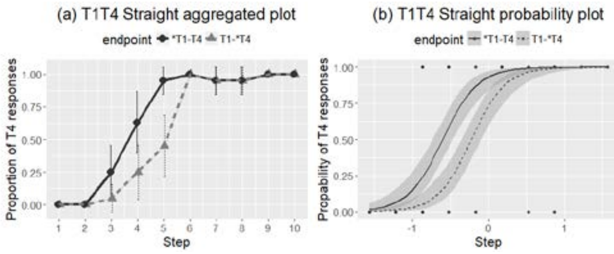


Fig.2 Aggregated and probability plot on the T1-T4 continua showing an earlier boundary on the *T1-T4 continuum than on the T1-*T4 continuum.

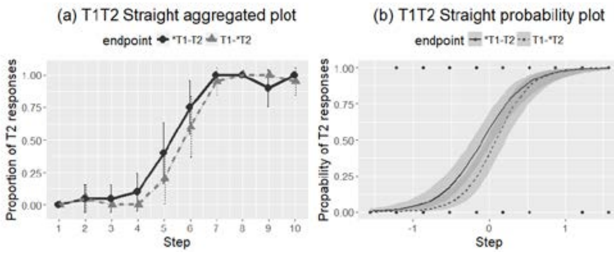


Fig.3 Aggregated and probability plot on the T1-T2 continua showing an earlier boundary on the *T1-T2 continuum than on the T1-*T2 continuum.

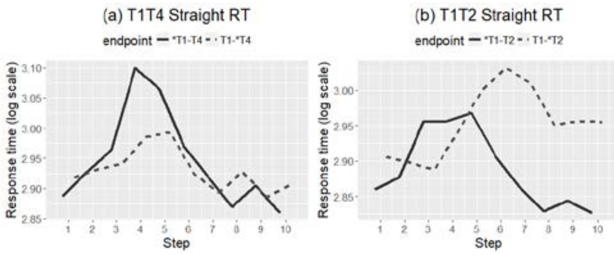


Fig.4 RTs of the identification showing earlier peaks for the *T-T continua and later for the T-*T continua.

References

[1] Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance 6*: 110-125.

[2] Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception and Psychophysics 34*(4), 338–348.

[3] Boomershine, A., Hall, K. C., Hume, E., et al. (2008). The impact of allophony versus contrast on speech perception. In: Avery P, Dresher E and Rice K (eds) *Contrast in Phonology*. Berlin: Mouton de Gruyter, 143-172.

[4] Dupoux, E., Kakehi, K., Hirose, Y., et al. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance 25*: 1568-1578.

[5] Ahn, M. (2008). Morphologically conditioned perceptual bias. *Proc. Chicago Linguistic Society, 44*:1-15.

[6] Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford: Oxford University Press.

[7] Kawahara, H., Morise, M., Takahashi, T., et al. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. Acoustics, Speech and Signal Processing Las Vegas, NV, 3933-3936.

[8] Wu, F., & Kenstowicz, M. (2015). Duration reflexes of syllable structure in Mandarin. *Lingua, 164*, 87–99.

[9] Yu, K. M. (2010). Laryngealization and features for Chinese tonal recognition. Proc. INTERSPEECH-2010 Makuhari, 1529-1532.

[10] Schneider, W., Eschman, A., & Zuccolotto, A. (2002) E-Prime User's Guide. Pittsburgh: Psychology Software Tools Inc.

[11] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Soft. 67*(1), 1–48.

A Cross-dialectal Comparison of *Er*-suffixation in Beijing Mandarin and Northeastern Mandarin: An Electromagnetic Articulography Study

Song Jiang, Yueh-chin Chang, Feng-fan Hsieh

Institute of Linguistics, National Tsing Hua University (Taiwan)
ajiangsong34@gmail.com, ycchang@mx.nthu.edu.tw, ffhsieh@mx.nthu.edu.tw

This study explores articulatory differences of *Er*-suffixation (*Erhua*; diminutive suffixation) between Beijing Mandarin (BJM) (see [1], among many others) and Northeastern Mandarin (NEM). With the help of an NDI Wave, we compared temporal changes of the tongue configuration at different duration deciles of *Er*-suffixed monophthongs (represented with ten solid lines for each stem: t1-t10; the tongue postures of corresponding unsuffixed monophthongs are represented with a green dashed line throughout). Six speakers (BJM=3; NEM=3) participated in the experiments and 10 tokens (5 *Er*-suffixed and 5 unsuffixed) were analyzed for each stem. Only representative results from one BJM speaker (BJ-01) and one NEM speaker (NE-01) are provided here since the results are highly consistent within each dialect group.

High front vowels /i, y, ɿ, ʅ/: We can see from Figs. 1-2 that although tongue root retraction is attested in both BJM and NEM speakers when a high front vowel is being *Er*-suffixed, one significant cross-dialectal difference emerges, namely that BJM speakers employ a "curling-up" posture (negative Tongue Tip Angle (TTA); see [2]) to produce the *Er*-suffix, whereas the curling-up posture is completely absent in NEM. Instead, NEM speakers opt for a downward tongue tip (positive TTA) in *Er*-suffixation. Taken together, it is evident that *Er*-suffixation involves tongue root retraction in both BJM and NEM, while BJM features an "additional" gesture of tongue tip curling. This cross-dialectal difference also leads to that the temporal lingual changes (solid lines) are of much greater magnitude in BJM *Er*-suffixation. Interestingly enough, regarding the "retroflex" apical vowel /ʅ/, no obvious change is found, especially in NEM *Er*-suffixation ([3]).

Back vowels /ɑ, u/: We can see in Figs. 3-4 that BJM also shows a significant tongue shape change in *Er*-suffixation, whereas NEM's tongue postures are kept "stable" throughout.

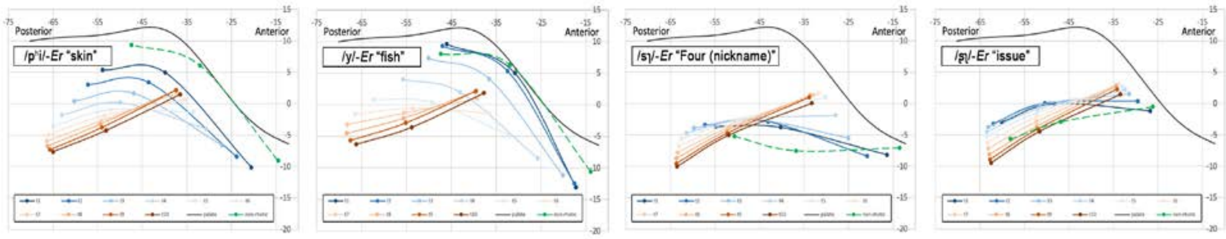


Fig.1 Temporal lingual changes of BJ-01 (Front vowels)

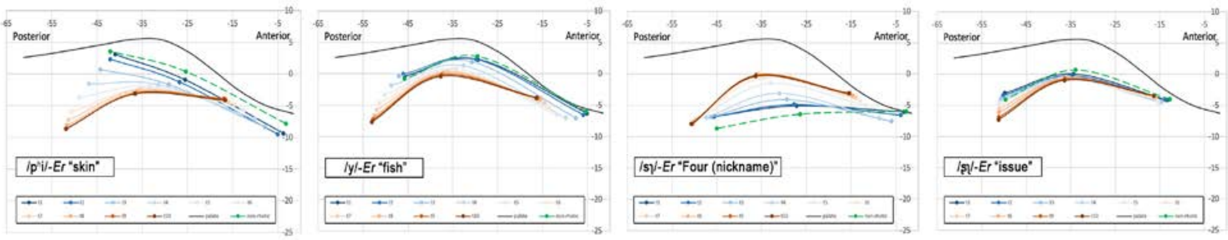


Fig.2 Temporal lingual changes of NE-01 (Front vowels)

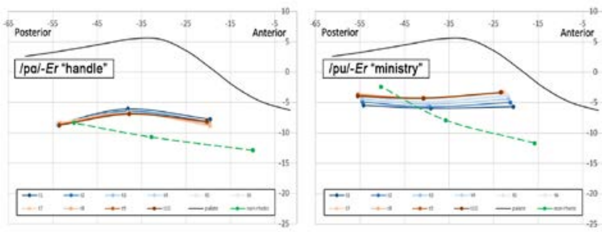


Fig.3 Temporal lingual changes of BJ-01 (Back vowels)

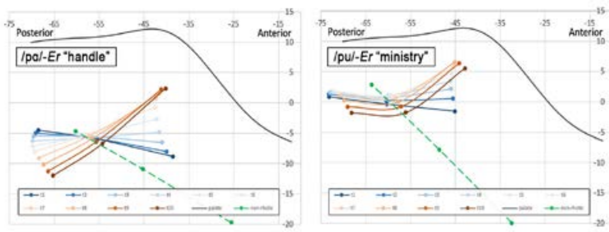


Fig.4 Temporal lingual changes of NE-01 (Back vowels)

It is also important to note that, throughout the entire syllable, an *Er*-suffixed back vowel (solid lines) is substantially different from its corresponding unsuffixed back vowel (green dashed line) in terms of tongue configuration in both BJM and NEM (Figs. 3-4). However, the onset of an *Er*-suffixed front vowel (t1) is similar to its corresponding unsuffixed front vowel in this regard (Figs. 1-2), again in both BJM and NEM. The result also confirms previous impressionistic observations according to which some instances of *Er*-suffixation are "absorbed" into the stem when the stem contains a back vowel, especially in NEM, whereby no tongue tip raising is found.

For the sake of completeness, we further compared the tongue postures of the last decile (offset) of the *Er*-suffix across all tokens (see Figs. 5-6). Boxplot of TTA (Fig. 7) and results of an ANOVA test show a significant difference on Tongue Tip Angle (TTA) between BJM and NEM (F=400.32, p<0.001), confirming that BJM speakers use a curling-up tongue tip (negative TTA), while NEM speakers opt for a downward tongue tip (positive TTA) in *Er*-suffixation.

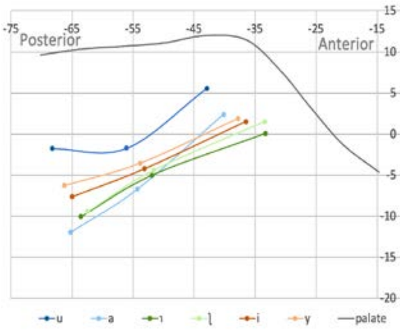


Fig.5 Offsets of Er-suffix (BJ-01)

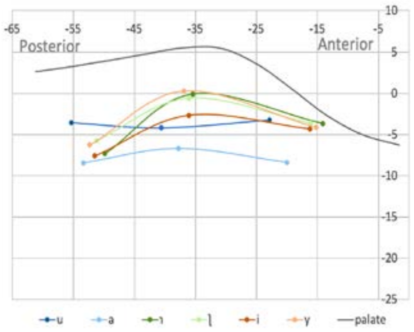


Fig.6 Offsets of Er-suffix (NE-01)

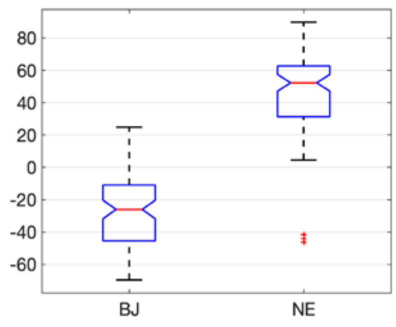


Fig.7 Boxplot of TTA of BJM and NEM speakers

Discussion: The EMA results confirm that *Er*-suffixation may involve both tongue tip raising and tongue root retraction in Mandarin Chinese, too, just like the well-established "retroflex" and "bunched" configurations found in the production of /ɹ/ in American English ([4], [5], among many others). More importantly, our results further reveal that there is an unprecedented cross-dialectal difference between BJM and NEM in *Er*-suffixation, at least in terms of tongue configuration. In addition, we shall report results of the acoustic data in the conference to see if there is any cross-dialectal difference as well.

Conclusion: We have shown in this study that *Er*-suffixation may not be homogenous as previously thought, since BJM and NEM are closely related dialects of Mandarin Chinese. It is worth exploring the sociophonetic or diachronic source of the cross-dialectal difference. Last but not least, for a more comprehensive understanding of the phenomenon in question, our next step is to investigate other cases of which stems are a closed syllable and/or a diphthong/triphthong, whereby incomplete neutralization has been reported to occur (e.g., [3]).

References

[1] Lee, W.-S. (2005). A phonetic study of the "er-hua" rimes in Beijing Mandarin. *In 9th European Conference on Speech Communication & Technology*, pp. 1093-96.

[2] Tiede, M. et al. (2007). Variability of North American English /r/ Production in response to Palatal Perturbation. *Haskins Internal Workshop on Speech Production & Motor Control.*

[3] Li, S.J. 1986. *Hanyu 'Er' /ɚ/ Yinshi Yanjiu [A historical analysis of Chinese /ɚ/]*. Beijing: The Commercial Press.

[4] Boyce, S., Tiede, M., Espy-Wilson, C.Y., & Groves-Wright, K. (2015). Diversity of tongue shapes for the American English rhotic liquid. *Proceedings of the 18th International Congress of Phonetic Sciences.*

[5] Westbury, J. et al. (1998). Differences among speakers in lingual articulation for American English /r/. *Speech Communication, 26*: 203-26.

Prosodic effects of DP-internal word order variation in Xitsonga

Seunghun J. Lee[1,2] & Kristina Riedel [3]

[1]*International Christian University (Japan),* [2]*University of Venda (South Africa),*
[3]*University of Free State (South Africa)*
seunghun@icu.ac.jp, RiedelK@ufs.ac.za

Many Bantu languages have penultimate lengthening where the penultimate syllable of a sentence is lengthened. The domain of penultimate lengthening is described to be an intonational phrase. Qualitative penultimate lengthening can be observed even at the phonological word level. There is an extensive literature on penultimate lengthening in Bantu languages [8], but none of this addresses DP-internal patterns. This paper seeks to address this gap by introducing and analysing syntactic and phonological patterns, including new data from Xitsonga, a southern Bantu language.

Bantu languages generally have a noun-initial DP word order. However, all (or nearly all) Bantu languages also allow for demonstratives (1) and sometimes a quantifier meaning 'each, every' to precede the noun. Beyond this, Bantu languages generally allow changing the relative order of the post-nominal modifiers which leads to subtle (focus-related) changes in meaning but generally Bantu languages do not allow for adjectives, numerals and possessives to appear before the noun. However, Tsonga (S53), Tswana (S31), Haya (JE22) and Basaá (A43a) allow these kinds of nominal modifiers to appear in the pre-nominal position. While the general properties of Bantu noun phrases are well documented, there are few studies about the specific ordering constraints in the various Bantu languages [9], nor any larger cross-linguistic comparisons. Discussions of Bantu language noun phrases generally focus on nominal morphology or the augment [4, 5, 6, 7].

Xitsonga allows numerals and adjectives to appear prenominally (1-3). This word order is associated with focus and the focused element is marked with a longer duration of penultimate lengthening compared to unfocussed elements in non-sentence final position.

(1a)	va-nhu lá-va		(1b)	lá-vá vá:-nhu	
	2-people this-2			this-2 2-people	
	'these people'			'THESE people'	
(2a)	va-nhu	va-mbi:rhí	(2b)	va-mbirhí	vá:-nhu
	2-people	2-two		2-two	2-people
	'two people'			'TWO people'	
(3a)	va-nhu	va-nkú:lú	(3b)	va-nkúlú	vá:-nhu
	2-people	2-big		2-big	2-people
	'big people'			'BIG people'	

Xitsonga also allows for multiple modifiers to appear pre-nominally with N Adj Num being able to appear in any of the logically possible orders. This pattern has not been reported for other Bantu languages, except for Basaá and violates Greenberg's Universal and Cinque's predictions on possible and impossible ordering [1, 2, 3].

In this paper, we offer an overview of noun phrase word order patterns in Xitsonga and analyse these patterns in terms of the effects on penultimate lengthening. The focused element in (1b, 2b, 3b) show longer penultimate vowel than the unfocussed head noun in (1a, 2a, 3a).

References

[1] Cinque, G. (1996). The antisymmetric programme: theoretical and typological implications. *J. of Linguistics* 32: 447-464.
[2] Cinque, G. (2000). On Greenberg's Universal 20 and the semitic DP. *The University of Venice WPL* 10 (2): 45-61.
[3] Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *LI* 36: 315-332.
[4] de Blois, K. F. (1970). The augment in the Bantu languages. *Africana Linguistica* 4:86–165.
[5] Ferrari-Bridgers, F. (2008). A unified syntactic analysis of Italian and Luganda nouns. In C. De Cat and K. Demuth (eds.). *The Bantu–Romance Connection*. Benjamins. 239–258.

[6] Halpert, C. (2015). Argument licensing and agreement. Oxford University Press.
[7] Halpert, C. (to appear). The Augment. In *The Oxford Handbook of Bantu Linguistics.*
[8] Hyman, LM. (2009). Penultimate Lengthening in Bantu: Analysis and Spread. *UC Berkeley Phonology Lab Annual Report*
[9] Letsholo, R. and K. Matlhaku (2014). The syntax of the Setswana noun phrase. *Marang* 24, 22-41.

Function and Prosodic Form of Backchannels in L1 and L2 German

Simon Wehrle & Martine Grice

University of Cologne, Germany
simon.wehrle@uni-koeln.de; martine.grice@uni-koeln.de

Backchannels (BC) are used in conversation to signal understanding or agreement. They are a ubiquitous and essential feature of spoken communication. It has been claimed that deviances from language-specific conventions of BC usage have negative consequences for understanding, comprehensibility and character attribution [1][2][3][4]. In a previous study involving mouse-tracking to indicate how (un)friendly, (in)attentive or (dis)interested a speaker sounded, German native listeners rated rising and falling BCs predominantly positively, whereas they rated flat intonation negatively [5]. In light of these perception results, we investigate how this flat intonation is distributed in BC productions, both in native speakers of German as well as in learners with L1 Vietnamese. In Vietnamese, flat (or slightly falling) BCs are most frequent and perceived as polite and acceptable according to [4].

Our corpus consists of Map Task data from 12 age-matched female speakers in three groups: 4 L1 speakers of German (G_{L1}), 4 L2 speakers of German with L1 Vietnamese (G_{L2}) and 4 L1 speakers of Vietnamese (V_{L1}). G_{L2} speakers were students in Germany and had a proficiency level of B2 CEF. In total, we analysed 135 minutes of dialogue containing 812 BCs. For German, all BC types except "*ja* (yes)", "mm(hm)" and "*genau* (exactly)" had previously been excluded, as these were the only BC types investigated in [5]. 73.2% of all BC tokens remained for analysis. The pitch contours of all BCs were manually corrected and smoothed. f0 was measured at 10% and 90% of token duration and the difference between the two points was calculated in semitones (ST) (cf. [4]). Results are presented on a continuous scale and tokens with a pitch movement of less than 1 ST are considered as "flat". Duration did not enter into the analysis as all tokens are very short, with very little variance (mean: 0.4s; SD: 0.11s).

Figure 1 shows ST values for all BC tokens. For G_{L1}, relatively few tokens are flat (16.7%), whereas for G_{L2}, almost twice as many are flat (30.9%) and both the mean and median values fall within the area +/- 1 ST. For V_{L1}, the overwhelming majority of BC tokens are falling, with very little variance, but also some level tokens (10%). This overall pattern is indicative of transfer effects from V_{L1} to G_{L2}.

Next, we consider different lexical types of BCs in German. Our working hypothesis is that prosodic form plays a special role for the non-lexical "mm(hm)" type, as speakers cannot rely on literal meaning in this case. For $G_{L1,}$ we found that all 67 "mm(hm)" BC tokens are rising, except one produced with a flat contour. For G_{L2}, on the other hand, "mm(hm)" tokens seem to be evenly distributed across rising, level and falling contours. We hypothesise that the G_{L1} distribution arises from an impetus to clearly distinguish "mm(hm)" BCs from filled pauses, which have a similar segmental form in German (*ähm*) and which are also non-lexical. Functionally, however, FPs and BCs play opposite roles in dialogue management. While BCs are used by listeners to signal to the interlocutor that the message is understood and that they may continue their turn, FPs are used by speakers to signal that their own turn is not finished and that they intend to hold the floor. Therefore, speakers may aim to distinguish non-lexical BCs and FPs as much as possible in their prosodic realisation.

Figure 2 shows a comparison of FP tokens (n = 194) with "mm(hm)" BC tokens (n = 119) in German. FPs in G_{L1} are firmly within the region of flat pitch contours, mostly with a very slight fall (cf. [6,7]). FPs in G_{L2} have a similar, but more variable distribution. Crucially, in G_{L1}, there is almost no overlap between FPs and non-lexical BCs. G_{L2} does not show this complementary distribution: FPs and non-lexical BCs overlap completely in their prosodic realisation. This overlap in G_{L2} BC and FP productions may result from the speakers' L1, as speakers of Vietnamese do not appear to produce FPs in any comparable sense, instead repeating or lengthening words to achieve the same function [8].

In sum, we have shown that G_{L2} speakers produce more BC tokens with a flat intonation contour than G_{L1} speakers. Furthermore, G_{L2} speakers do not distinguish the prosodic realisation of non-lexical BCs from that of FPs in the clear and precise way of G_{L1} speakers. In line with [5], both of these behaviours are likely to lead to negative character attributions and misunderstandings in conversation with G_{L1} speakers.

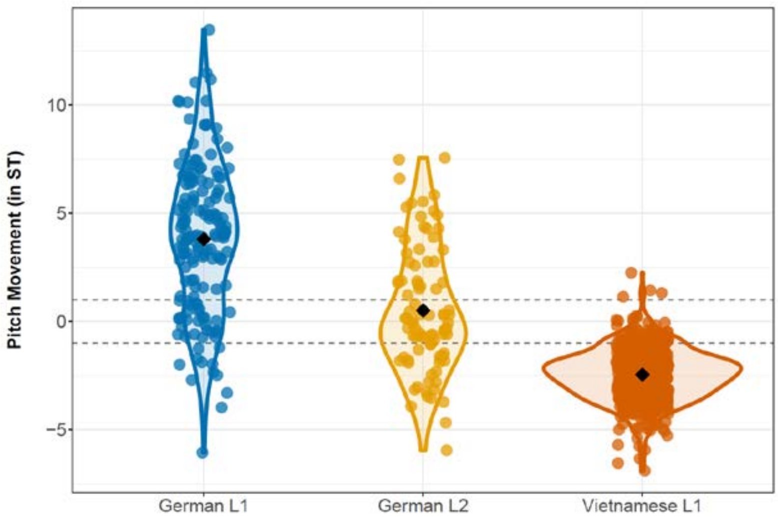


Fig.1 Prosodic realisation of BCs in GL1, GL2 and VL1. Pitch movement in semitones on the y-axis. The dashed lines indicate the "flat" region of +/- 1 semitone. Black diamonds indicate mean values per group. Dots represent individual BC tokens.

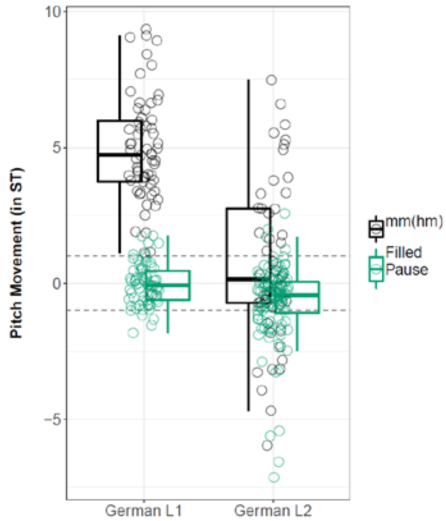


Fig.2 Prosodic realisation of non-lexical BCs and filled pauses in GL1 and GL2. Dashed lines indicate the flat region of +/- 1 semitone. Circles represent individual tokens.

References

[1] Cutrone, P. (2011). Politeness and face theory: Implications for the backchannel style of Japanese L1/L2 speakers. *Language Studies Working Papers, 3*, 51-57.

[2] Li, H.Z. (2006). Backchannel responses as misleading feedback in intercultural discourse. *Journal of Intercultural Communication Research, 35*(2), 99-116.

[3] Shelley, L., & Gonzalez, F. (2013). Back Channeling: Function of Back Channeling and L1 Effects on Back Channeling in L2. *Linguistic Portfolios, 2*(1), 97-108.

[4] Ha, K.P., Ebner, S. & Grice, M. (2016). Speech prosody and possible misunderstandings in intercultural talk - A study of listener behaviour in Vietnamese and German dialogues. *Proc. Speech Prosody 8*, Boston, 801-805.

[5] Wehrle, S., Roettger, T.B. & Grice, M. (2018, November*). Exploring the dynamics of backchannel interpretation: The meandering mouse paradigm*. Paper presented at ProsLang: Workshop on the Processing of Prosody across Languages and Varieties, Wellington, New Zealand.

[6] Shriberg, E. (2001). To 'errrr'is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association, 31*(1), 153-169.

[7] Adell, J., Bonafonte, A., & Escudero-Mancebo, D. (2010). Modelling filled pauses prosody to synthesise disfluent speech. In *Speech Prosody 2010-Fifth International Conference*.

[8] Marc Brunelle, personal communication, March 28, 2019.

A Study on Acoustic Characteristics and Korean EFL Learners' Perception of English Voiceless Fricatives

Zhang Yinjun[1] & Suh Heajin[2]

[1]*Pusan National University (Korea),* [2]*Busan National University of Education (Korea)*
ajadazyj@126.com, wisepearl3@gmail.com

English fricatives are consonants which are produced with a turbulent airstream through a narrow channel [1]. Previous studies on English fricatives mainly focused on the place and manner of articulation of fricative sounds [2,3] and tried to specify distinctive features of English fricative sounds [4]. Recently, research on English fricative sounds focuses on the L2 context and perceiving and producing English fricative sounds became the main research content with regard to learners who learn English as a foreign language (EFL). These studies mainly focused on how differences between English and Korean fricatives interrelate with their perception and production of English and Korean fricatives [5,6,7,8,9,10]. However, few studies examine the role of acoustic characteristics in perception of English fricatives. Therefore, in this study, we will focus on acoustic characteristics of English fricatives and examine the relationship between acoustic characteristics and EFL learners’ perception of English fricatives.

Three female American native speakers (N1, N2, N3) participated in this study. Additionally, twenty Korean EFL learners participated in the perception experiment. The stimuli used in this study were monosyllable consonant-vowel-consonant (CVC) words. The onset consonants were three voiceless fricatives, the interdental fricative /θ/, the alveolar fricative /s/ and the palato-alveolar fricative /ʃ/. In according to [11], due to the lack of corresponding phonemes of English /f, v/ in Korean phonetic system, Korean EFL learners tend to produce Korean sounds /ㅍ,ㅂ/ as substitutions of English /f, v/. English labiodental fricatives are often studied by Korean researchers with Korean /ㅍ,ㅂ/ rather than with English interdental, alveolar and palato-alveolar fricatives. Therefore, we excluded the voiceless labiodental fricative sound in the test. Three vowels were used in this study: the high front vowel /i/, the low back vowel /a/ and the high back vowel /u/. The word-final consonant is the voiceless stop sound /p/. The English stimuli were recorded by native speakers in quiet rooms as mono sound files using a Sony recorder at 22.1 kHz and a bit rate of 16. English stimuli were read and recorded three times in the carrier sentence "Say _______ again". The recorded sound files were segmented, edited and analysed through Praat. To investigate how Korean EFL learners perceive English voiceless fricatives, the MFC listening test (the identification test) was conducted. In the identification test, the subject heard stimuli one by one and there are three answer candidates on the screen. Three candidates are /θ/, /s/ and /ʃ/. Subjects can select only one of the candidates as their response. Four acoustic characteristics were tested in this study, which include spectral peak location, frication duration, center of gravity (COG) and intensity. All these measurements were completed by hand in Praat (version 6.0.49) [12].

The results of the identification test showed that Korean EFL learners can distinguish between English voiceless interdental fricative, alveolar fricative and palato-alveolar fricative sounds. Furthermore, we used the Spearman correlation test to examine whether there is any relationship between acoustic characteristics and Korean EFL learners' perception of English voiceless fricative sounds and found that normalized duration and intensity show statistically significant correlation between learners' perception of English voiceless fricative sounds.

	Group			Chi-square	P value
	/θ/	/s/	/ʃ/		
Correct Answers	39	35	35	14.886	0.001**

* p<0.05 ** p<0.01

Table 1. The Kruskal-Wallis Test Results of the identification test

	1	2	3	4	5
1.Accuracy	--				
2.Spectral peak	0.236	--			
3.COG	-0.100	.452*	--		
4.Frication duration	-0.336	-.563**	0.217	--	
5.Normalized Duration	-.550**	-.428*	0.223	.860**	--
6.Intensity	-.422*	-0.033	0.198	0.367	.539**

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 2. Spearman Correlation analysis

References

[1] Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages.* Oxford: Blackwell.

[2] Behrens, S. J., & Blumstein, S. E. (1988a). Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics,* 16, 295-298.

[3] Lawrence, J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni & R. E. Remez (eds.), *The handbook of speech perception* (pp. 182–206). Malden, MA, Oxford, and Carlton, Victoria: Blackwell publishing.

[4] Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *The Journal of the Acoustical Society of America, 85*(4), 1718-1725.

[5] Cheon, S. Y., & Anderson, V. B. (2008). Acoustic and perceptual similarities between English and Korean sibilants: implications for second language acquisition. *Korean Linguistics, 14*(1), 41-64.

[6] Holliday, J. J. (2012). The acoustic realization of the Korean sibilant fricative contrast in Seoul and Daegu. *Phonetics and Speech Sciences, 4*(1), 67-74.

[7] Lee, S. (2011). Perception and production of English fricative sounds by advanced Korean EFL learners. *Studies in Phonetics, Phonology and Morphology, 17*(2), 259-281.

[8] Seo, M., & Lim, J. (2008). Perceptual discrimination of sibilants by adult Korean learners of English. *Applied Linguistics Association of Korea, 24*(3), 219-238.

[9]Seo, M., & Lim, J. (2009). A study of voiced alveolar and palato-alveolar fricative production by Korean learners of English. *English phonetics, 9*(4), 663-685.

[10] Suh, Y. (2013). The English palatoalveolar fricative in Korean speakers' second language production. *Korea Journal of English Language and Linguistics, 13*(3), 741-758.

[11] Koo, H. S. (2009). A study of production difficulties of English bilabial stops and labiodental fricatives by Korean learners of English. *Phonetics and Speech Sciences, 1*(4), 1115.

[12] Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer. Retrieved from http://www. praat. org/.

Asymmetry in the perceptual assimilation of the Korean laryngeal contrast by Indian listeners

Chiin Ngaihmuan Ngaihte[1] & Jeffrey J. Holliday[2]

Korea University (Korea)
[1]muanmok@gmail.com, [2]holliday@korea.ac.kr

This study explores the naïve perception of word-initial Korean stops and affricates by native (L1) listeners of two Indian languages, Hindi and Paite. Whereas Korean stops and affricates are characterized by a three-way laryngeal contrast among tense, lax, and aspirated categories (e.g. /t*, t, t^h/), all voiceless in word-initial position, Hindi has a four-way contrast of voicing and aspiration (e.g. /d^h, d, t, t^h/), and Paite has a three-way contrast, but different from Korean (e.g. /d, t, t^h/). All of the Hindi and Paite laryngeal contrasts are homorganic with the exception of Paite affricates, which seem to be /ʤ, ʦ, $ʧ^h$/ [1]. Thus, although all three Korean categories are voiceless in word-initial position, both Hindi and Paite contain only two voiceless categories (/t/ and /t^h/), raising the question of whether the existence of two extra voiced categories in Hindi and one extra voiced category in Paite would affect how Korean stops and affricates are perceived by naïve listeners. According to the Perceptual Assimilation Model [2], non-native segments should assimilate to the most gesturally similar L1 categories, which would be the Hindi and Paite voiceless unaspirated and aspirated categories (i.e. /p, t, k, ʦ, ʧ, p^h, t^h, k^h, $ʧ^h$/). Nevertheless, this question resembles the subset problem in L2 perception [3], and is empirically important for the development of hypotheses about non-native speech perception involving such typologically marked contrasts.

The target stimuli consisted of 12 Korean consonants (/p*, t*, k*, p, t, k, p^h, t^h, k^h, ʨ*, ʨ, $ʨ^h$) combined with three vowels (/a, i, u/) to form 36 CVs, which were produced by four female native speakers of Seoul Korean, resulting in a total of 144 unique target CV stimuli. Listeners completed a perceptual assimilation task using all 144 stimuli, in which they indicated which L1 sound the stimulus was most similar to, and rated the category goodness on a scale of 1 to 5.

The native Hindi listeners (n=15) were tested in Delhi. Most (n=10) also spoke English as a second language, as is common in India. The Paite listeners (n=15) were tested in Churachandpur, a town in Manipur state, India, and all reported speaking English as a second language. The experiment was presented in a quiet room using headphones and a notebook computer running OpenSesame version 3.2.4. Responses were recorded on paper.

An assimilation fit index (FI) was calculated for each Korean/L1 sound pair by multiplying the mean goodness rating by the percentage of trials of the Korean sound that were assimilated to the L1 sound. Thus, the FI can range from 0 (= the L1 sound was never chosen for a given Korean sound) to 5 (= it was always chosen and always given a goodness rating of 5). As another example, an FI of 3 could represent an assimilation rate of 75% and a mean goodness rating of 4 (75% × 4 = 3), an assimilation rate of 60% and a mean goodness rating of 5 (60% × 5 = 3), or one of many other combinations. Due to space limitations, the FIs presented here (shown in Table 1) are collapsed into Hindi and Paite voicing/aspiration categories, rather than showing each consonant and vowel context individually.

Korean tense stops strongly assimilated to Hindi and Paite voiceless unaspirated categories, with FIs ranging from 3 to 4. Korean lax and aspirated stops generally assimilated to Hindi and Paite voiceless aspirated categories, but the FIs were slightly lower, ranging mostly from 2 to 3, and with some Korean categories weakly assimilating to a voiceless unaspirated category as well. In terms of place of articulation (not shown in Table 1), bilabial and velar stops strongly assimilated to homorganic Hindi and Paite categories. Hindi listeners assimilated Korean alveolar stops to both Hindi dental and retroflex categories, whereas Paite listeners strongly assimilated them to Paite dental categories. In summary, the results for stops were roughly in line with what we expected based on the Perceptual Assimilation Model.

Although phonologically analogous to the stops, the assimilation of Korean affricates exhibited quite a different pattern. First, Paite listeners assimilated all three Korean affricates, including unaspirated /tɕ*/, to a voiceless aspirated category. This result may be explained by the fact that Paite voiceless affricates consist of only /ʦ/ and /ʧʰ/. Korean /tɕ*/ may be more similar to /ʦ/ in terms of aspiration, but /ʧʰ/ in terms of place, suggesting that Paite listeners might have weighed the place cue more heavily than the aspiration cue.

In the case of Hindi, the highest FIs for the Korean affricates were found in the Hindi voiced affricate categories. Korean /tɕ*/ assimilated reasonably well to Hindi /ʤ/, but also to a lesser degree to Hindi /ʧ/. Korean /tɕ/ and /tɕʰ/, however, were split more between Hindi /ʤ/ and /ʤʰ/. These results were unexpected, as none of the Korean affricates are voiced, and none of the Korean stops assimilated even slightly to a Hindi voiced category. Thus, even though the Korean affricate contrast is usually described as laryngeally analogous to the well-studied three-way stop contrast, it does not perceptually map onto the Hindi four-way contrast in an analogous way.

These results suggest that although Hindi and Paite listeners may not differ in how they perceive Korean stops, they may differ in their perception of Korean affricates. We are currently analyzing discrimination data from these same listeners, and plan to extend this work into the production of these Korean contrasts by native Hindi and Paite L2 learners of Korean.

	Hindi				**Paite**		
Korean	**voiced aspirated**	**voiced unasp.**	**voiceless unasp.**	**voiceless aspirated**	**voiced unasp.**	**voiceless unasp.**	**voiceless aspirated**
/p*/ ㅃ			**3.73**			**3.57**	
/t*/ ㄸ			**3.47**			**3.57**	
/k*/ ㄲ			**3.70**			**3.51**	
/tɕ*/ ㅉ		**2.32**	**1.11**			0.87	**2.56**
/p/ ㅂ			**1.19**	**2.11**		0.98	**2.85**
/t/ ㄷ			0.86	**2.43**		0.85	**2.72**
/k/ ㄱ			**1.05**	**2.38**		0.64	**2.81**
/tɕ/ ㅈ	0.84	**1.04**	0.69	0.62	0.29	0.58	**2.59**
/pʰ/ ㅍ			**1.23**	**2.04**		0.74	**3.08**
/tʰ/ ㅌ			0.86	**2.69**		0.85	**2.99**
/kʰ/ ㅋ			0.85	**2.69**		0.79	**2.82**
/tɕʰ/ ㅊ	**1.33**	**1.13**	0.36	0.55	0.23	0.38	**2.91**

Table 1. Perceptual fit indices (FIs) for Korean stops and affricates for Hindi and Paite listeners. For clarity, FIs less than 0.2 were omitted, and FIs greater than 1.0 are in boldface.

References

[1] Nianglianmoi. (2016). Sentence phonology of Paite (Doctoral dissertation). Jawaharlal Nehru University.

[2] Best, C. T. (1995). A direct realist view of cross-language speech perception. Speech Perception and Linguistic Experience (pp. 171-206).

[3] Escudero, P., & Boersma, P. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In Proceedings of BUCLD 26 (pp. 208-219).

Tonal Patterns in the IP-final AP in Chonnam Korean

Mira Oh

Chonnam National University (Korea)
mroh@jnu.ac.kr

This study investigates tonal realization of the IP-final AP in Chonnam Korean under variation in AP length, AP-initial consonant and question endings. The typical tonal patterns of an Accentual Phrase in Seoul Korean and Chonnam Korean are THLH and THL, respectively [1]. The IP-final AP contours are conditioned by the AP length and boundary tones in Jun's framework. In Seoul Korean, the second and the penultimate tones do not occur if APs are fewer than four syllables. The IP boundary tone overrides the accentual phrase-final tone in Korean prosody [1]. Jun's theory predicts that the rest of the AP tones are intact when the boundary tone replaces the AP-final tone. Then we would expect the following tonal patterns of the IP-final AP when the boundary tone is H% for Seoul and Chonnam Korean according to the IP-final AP length.

(1) Expected tonal realization (T=L or H depending on the AP-initial consonant)

IP-final AP length	Seoul Korean	Chonnam Korean
2 tone bearing units	TH→ T-H%	TH→ T-H%
3 tone bearing units	THH→ TH-H%, TLH→ TL-H%	THL→ TH-H%
4 tone bearing units	THLH→ THL-H%	THL→ TH-H%

Previous studies report that the IP-penultimate syllable shows a pitch valley before H% in both Seoul and Chonnam Korean. As for Seoul Korean, [2] simply stipulates a constraint prohibiting high pitch in the IP-penultimate syllable and [3] analyzes the F0 valley as an Intermediate Phrase boundary tone. [1] analyzes the F0 valley immediately before the IP boundary as the AP-penultimate L tone in Seoul Korean. Then, in Seoul Korean, the penultimate L can be realized in 3 and 4 syllable IP-final APs since it is part of the AP tones. However, the F0 valley in 2 syllable-APs beginning with a H tone before H% cannot be explained by this analysis in Seoul Korean. As for Chonnam Korean, [4] explains the IP-penultimate L as the shift of the AP-final L to the IP-penultimate mora. Then this account is inconsistent in that the IP-penultimate L tone is the AP-penultimate tone in Seoul Korean but the AP-final tone in Chonnam Korean.

This study investigates the following questions. First, does the IP-penultimate L occur in Chonnam Korean regardless of IP-final AP length? Second, given that the sentence-final boundary tones differ depending on the sentence-ending forms [5], does the rate of the IP-penultimate L vary according to the boundary tone? Third, are there any other factors which affect the tonal contours of the IP-final APs besides the AP length and boundary tones? By analyzing the production of statements and interrogatives, we argue for the effects of the initial consonant and length of a sentence-final IP, and the sentence-final ending forms on the prosody of Chonnam Korean. Based on the results, we propose a reanalysis of the AP tonal pattern in Chonnam Korean as THLL instead of THL.

Production experiment 3,600 sentences recorded from 17 young Chonnam Korean speakers (10 female and 7 male) and 8 old Chonnam Korean speakers (5 female and 3 male) (48 sentences x 3 repetitions x 25 speakers) were analyzed. The results for *yo*-ending interrogatives are summarized in Fig. 1.

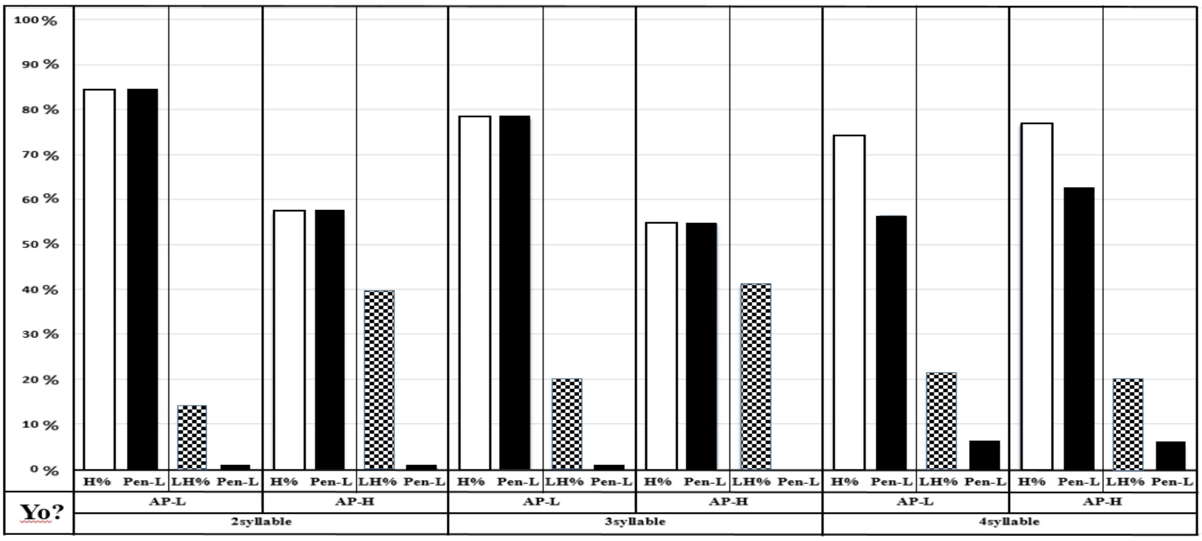


Fig.1 Rates of boundary tones and IP-penult L according to AP-initial tone and IP length

Three major factors affecting the IP-final AP contours for interrogatives (*–yo* ending and *–nya* ending) were found: i) *Lexical ending forms*: *–yo* ending was mostly realized with H% (68.6%), while *–nya* ending was realized as either LH% (42.3%) or H% (53.5%) in 2 and 3 syllable-IPs. Likewise, boundary tones were determined by lexical ending forms (*p*=6.48e-07); ii) AP-initial tone: the rate of H% is decreased but the rate of LH% is increased when the IP-final AP begins with a laryngeal consonant in 2 and 3 syllable-IPs (*p*=2e-16); iii) IP-penultimate L: The penultimate L-lowering is constantly present regardless of the number of syllables within IP and it occurs more before H% than before LH% (*p*=2e-16). It occurs less when the IP-final AP begins with a laryngeal consonant in 2 and 3 syllable-IPs (*p*=1.41e-11).

Discussion and conclusions The IP-final AP contours vary due to tone crowding by which multiple target tones cannot all be realized. We show that the IP-penultimate L occurs regardless of the length of the IP-final APs in Chonnam Korean. To explain the IP-penultimate L, we propose that the tonal pattern of the AP is LHLL instead of LHL in Chonnam Korean. It was shown that the rate of IP-penultimate L is affected by the types of boundary tones and the AP-initial tone. It suggests that the tonal patterns of the IP-final AP cannot be explained by simple replacement of the AP-final tone by a boundary tone. The high rate of the IP-penultimate L-lowering in two-syllable IP in this study coincides with Kim's (2014) suggestion that the lack of tone bearing unit does not necessarily lead to tonal deletion in contrast to Jun's (1996) analysis.

References

[1] Jun, Sun-Ah. (1996). *The Phonetics and Phonology of Korean Prosody*, Garland, New York.
[2] Lee, Ho-Young. (1990). The Structure of Korean Prosody, PhD dissertation, University of London.
[3] Kim, Kyunghee. (2014). Tone, Pitch Accent and Intonation of Korean, PhD dissertation, Universität zu Köln.
[4] Jun, Sun-Ah. (1998). The accentual phrase in the Korean prosodic hierarchy, *Phonology* 15, 189-226.
[5] Yun, Jiwon. (2017). Sentence-final intonation in Korean, presented at *ICKL* 20.

Game Theory Choice on Salient Phoneme-Realisation (SPR): Strategies on Variance in Speech Production and Perception

Taeeun Kim

Pukyong National University (Korea)
taeeungo@gmail.com

The sound of salience comes in two forms (Rácz, 2013) [1]: salience of phonemes with lexical references and/or social indices. Though the notions of salience in phonemic alterations remains undefined, the appearance and perception of salience within speech is obvious when it appears in systematically occurring alternating speech patterns; which are perceived as "robust", dominant phonemes within phonological structures. In order to develop a deeper understanding of salience, this paper examines Game Theory choice of phonological variation in English specifically the plosive phonemes, the consequences for the perception of speech, and the distinction between cognitive and sociolinguistic salience.

The selection process on Salient Phoneme-Realisation (SPR) by a population of the speech community is especially in line with the static characterisation of Evolutionarily Stable Strategies (ESS) by John Maynard Smith (1982) [2]. A language speaker's choice on SPR is analogous to an evolutionary trajectory towards increased adaptation to the environment of a target sound within the language community, in which interactions between language interlocutors can be considered strategic games. In Game Theory, players have something to win or to lose in the interaction with other participants and the payoff for each player depends upon their strategic choice against said players. Maynard Smith [2] investigated various species and saw that participants within a population adopted strategies even if their genes were no to the benefit of future generations causing alterations or mutations that led to differences. The innate strategic aims of the participants according to Maynard Smith meant the populations evolved an evolutionarily stable strategy (ESS), the behaviour of individual participants would lead to certain dominant genetic traits in later generations throughout the population. Thus, individual interlocutors within a speech community inadvertently establish their SPR strategies upon the speech community altering the choice of salient sounds available to the speech community.

A language player's choice on Salient Phoneme-Realisation (SPR) from their own phonemic inventory in the ESS setting, the operationalisation mechanism of salience, can provide the player with two major sets of adaptation strategies in speech production: *Imitation* or *Mutation*. And here were eluted to measure the utility matrix of the speaker's strategy between imitation and mutation for SPR regarding its stability (cognitive salience) and dynamics (sociolinguistic salience) by focusing on /p/, /t/, /k/. If the speaker's strategic choice for robust phonemic realisations of the plosives in a game-like condition is evolutionarily stable, the strategies can be regarded as something intrinsically congenial with (linguistic) gene configurations [3].

In other words, continuous repetition and the payoffs of each game under the interlocutors' strategic interactions are accumulated as *fitness*, i.e. evolutionarily stable reproduction of SPR. If a certain strategy for *salient variants* of linguistic variance yields on average a payoff that is higher than the population average, its reproduction rate will be higher than the average and its proportion within the total population increases, while strategies resulted in *non-salient variants* with a less-than-average expected payoff decreases in frequency (Jäger, 2004) [3]. For example, the utterance of voiceless stop /p/ word-initially in English, if produced as aspirated [p^h], it is a salient variant; while the production of it as tensed fortis [p'] is a *non-salient variant*, i.e. evolutionarily unstable.

The game-theoretic analysis on Salient Phoneme-Realisation (SPR) under the framework of Evolutionarily Stable Strategies (ESS) requires five basic elements: players, rules of the game, information, strategies, and payoffs. This investigation examined a speaker's choice of phonological variation, i.e. the consequences for the perception of speech, and the distinction between cognitive and sociolinguistic salience. In line with the static characterisation of ESS, to pinpoint the speaker's strategies between imitation and mutation for SPR in the production of

English readings by Native speakers (NS, *n=16*) and late bilingual Korean learners (KL, *n=14*) [5], the plosives /p/, /t, /k/ were tested within the exemplar utility matrix for SPR. Strategically, the production of a voiceless stop /p/ word-initially, for instance, the interlocutors from the two populations (player 1: NS and player 2: KL) are expected to choose among the three options of sound variants; aspirated as in [p^h], unaspirated/lenis [p], or fortis [p']. The concept of SPR, i.e. congenial to ESS, is strongly related to the rationalistic approaches of a Cognitive Salience Stability (CSS) and a Socio-linguistic Salience Dynamics (SSD). The CSS generates a Nash Equilibrium (NE) of a target phoneme, e.g. /p/ word-initially in English, if it is a best response to itself, resulting in [p^h]; while the SSD entails a Strict Nash Equilibrium (SNE) of a target phoneme if it is the unique best response to itself, resulting in [p].

The outcome of the two players' choice on the same strategy can be generalised as an evolutionarily stable strategy since all individuals within a population adopt a strategy, e.g. imitation, will never leave this choice unless mutations occur. A set of *imitation* strategies (e.g., ±aspiration, ±neutralisation, ±voicing to the plosives as per their syllabic positional condition in English) are evolutionarily stable if they are resistant against small amounts of non-imitation mutant variant, e.g. fortis [p] in the syllabic onset position. The strategy of imitation sets are minimal evolutionarily stable sets, i.e. have not evolutionary stable proper sets [3, 4]. Therefore, it is predictable that each population of the speech community will approach a minimal evolutionarily stable set if the level of mutation variance is sufficiently small as the production of the target sound utterances by all the players converges to a CSS. However, the existence of Nash equilibrium SSD, the unique best response of [p] within both NS and KL players, illustrates the possibility of a social index that leads the operationalisation mechanism of salience variance.

In conclusion, the SPR Utility Matrix for the plosives realisations by Native English speakers and bilingual Korean-English speakers of the target phonemes revealed that speakers make lexical and social choices based on their phonemic inventories, i.e. the consequences for the perception of speech sounds results from both cognitive and sociolinguistic salience.

References

[1] Rácz, P. (2013). *Salience in Sociolinguistics*. Berlin & Boston : Walter de Gruyter GmbH.
[2] Maynard Smith, J. (1982). *Evolution and the therory of games*. Cambridge : Cambridge University Press.
[3] Jäger, G. (2004). Evolutionary Game Theory for Linguists : A primer. *Mnuscript*, Stanford University & University of Potsdam, available from http://www.sfs.uni-tuebingen.de/~gjaeger/lehre/egtTuebingen/index.shtml
[4] Pierarinen A. V. (eds.). (2007). *Game Theory and Linguistic Meaning*. In Jaszczolt, K.M. & Turner, K. (eds.), *Current Research in the Semantics/Pragmatics Interface, 18*. Elsivier Ltd.
[5] Weinberger, S. (2015). *Speech Accent Archive*. George Mason University. Retrieved from http://accent.gmu.edu

Production and perception of English vowel length depending on the following consonant voicing by Korean learners of English

Juyeon Chung

Indiana University (USA)
juychung@iu.edu

A vowel followed by a voiced consonant is consistently longer than that followed by a voiceless consonant in native English productions (Flege & Hillenbrand [1], de Jong [2]). Previous studies of ESL learners indicate that learners from various language backgrounds do not show English-like patterns in these vowel duration differences. Complicating the durational effects of voicing, some studies of second language learners demonstrate that learners from languages that have no tense/lax distinction have been shown to have problems with the English tense/lax pair, in which part of the phonemic identity has a strong durational component (Flege, Bohn & Jang [3], Ingram & Park [4], Kim [5]). These studies suggest that ESL learners might have transferred the phonetic difference from their L1 to L2 English performance.

This paper investigates how L2 Korean learners of English produce and perceive English vowel duration associated with different voicing contexts and quality differences (tense vs. lax). It was expected that they will show different patterns when the triggering consonant is in different positions (medial vs. final), since Korean has no coda consonant voicing contrast in final position, but does preserve the voicing contrast in medial position. Also, it was expected that quality differences will affect subjects' vowel duration production and perception since Korean does not have tense/lax contrast. The stimuli included nonce words of /bVC/ and /bVCa/ containing /i, ɪ, u, ʊ, e, ɛ/ preceding (/p, b/, /t, d/, /k, g/), which were spoken by two American Midwest speakers. The vowels chosen for the experiment are English tense/lax vowel pairs that cause high confusion to Korean learners of English because Korean has no tense/lax distinction. The consonants selected for the experiment consist of sets of plosives contrasting in voicing. The nonce words were chosen to avoid lexical effects.

16 Participants who were from the regions of Seoul and Geyonggi province participated in the experiment. 8 of them were recruited from universities in Korea and had less than 6 months of exposure to an English speaking country. The other 8 were recruited from a university in the U.S. and had more than three years of exposure to English speaking environment. The subjects' average duration of studying English at school in Korea was around 16 years. The Korean native speakers' year of birth ranged from 1973 to 1991.

Each subject was recorded digitally using Audacity on a laptop in a sound-dampened room, individually. In the production task, the target English nonce words including 36 monosyllabic words and 36 disyllabic words (2 structures x 6 vowels x 6 consonants = 72 tokens) were randomized and presented to the subjects in a reading list. English samples of real words were given at the right side of the target words that they could refer to. The participants were asked to read the target words in a carrier sentence, "say ___ soon" once. In an ABX discrimination task, the subjects were asked to identify the different words from among minimal triples. The target words were excised from the carrier sentences and they were presented to subjects aurally. First two tokens were spoken by one speaker and the third token was spoken by another speaker.

For the production task, preceding vowel duration was measured using Praat software and the vowel length was submitted to an ANOVA in which the amount of L2 exposure to English speaking country (more experienced group vs. less experienced group) was a between-subjects factor and final consonant voicing (voiced vs. voiceless) and prosodic structure (monosyllabic vs. disyllabic) were within-subjects factors. A series of independent t-tests was run on tense/lax vowel duration for each group, respectively. For the perception task, accuracy was calculated, and a regression was run to test for a correlation across subject between vowel length differences in production and perception scores.

According to the results, none of the speakers exhibited different patterns in monosyllabic and disyllabic structures in productions or perceptions. All speakers did exhibit durational correlates to the voicing contrast, and to the tense-lax distinction. Formant frequency differences were found for both voicing and the tense-lax distinction. The effect of L2 English experience was expected, but inexperienced Korean speakers showed better performance than experienced Korean speakers on discriminating final consonant voicing contrasts in /bVC/ structure in perception task, but there was no statistically significant difference between the two groups. Overall, there was no correlation between vowel length differences in production and perception scores.

NE: Native English speakers / ExpK: Experienced Korean speakers / InexpK: Inexperienced Korean speakers

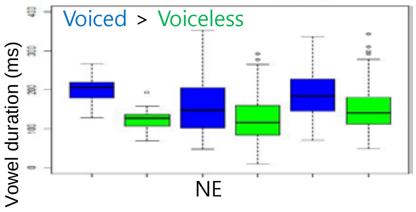


Fig.1 Vowel duration depending on the following consonant voicing in the monosyllabic structure

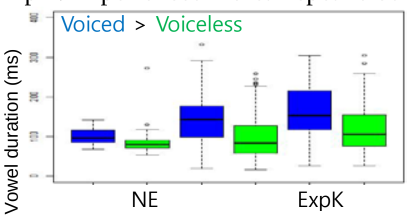


Fig.2 Vowel duration depending on the following consonant voicing in the disyllabic structure

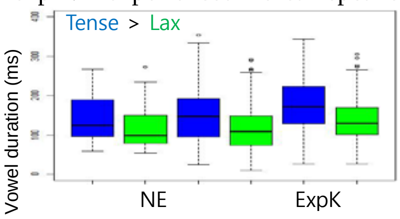


Fig.3 Vowel duration depending on tense/lax

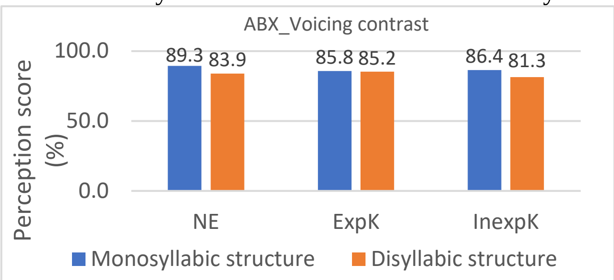


Fig.4 Scores of ABX task of discriminating vowel duration depending on the following consonant voicing

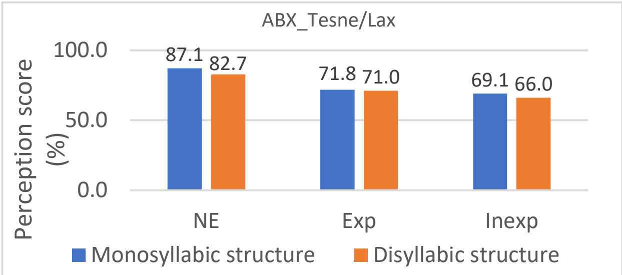


Fig.5 Scores of ABX task of discriminating tense/lax vowel

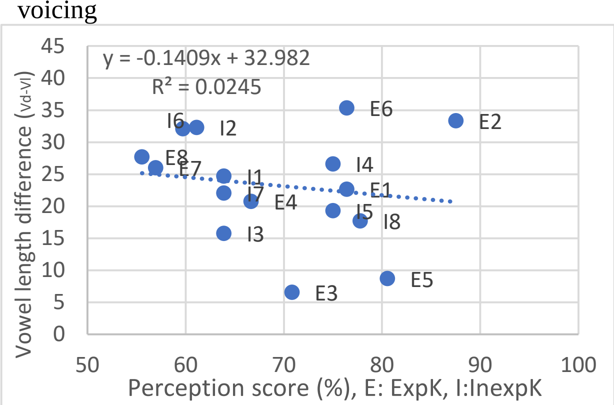


Fig 6. Correlation between vowel length differences (Vd-Vl) in production and perception scores

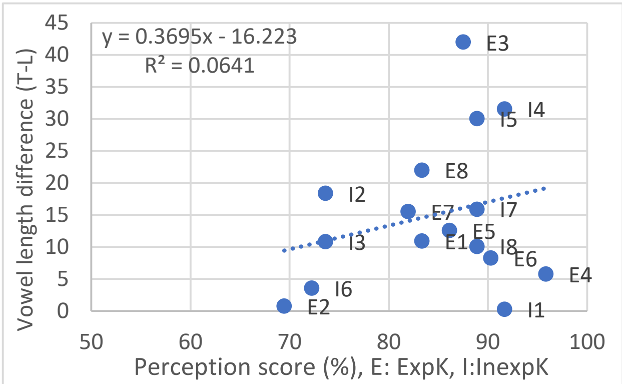


Fig 7. Correlation between vowel length differences (Tense-Lax) in production and perception scores

References

[1] Flege, J. & Hillenbrand, J. (1986). Differential use of temporal cues to the [s-z] contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America*, *79*, 508-517.

[2] 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.

[3] Flege, J., 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.

[4] Ingram, J., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343-370.

[5] Kim, J.-E. (2010). Perception and production of English front vowels by Korean speakers. *Phonetics and Speech Sciences*, *2*, 51-58.

A comparison of Mandarin and English palatal fricatives with articulatory data

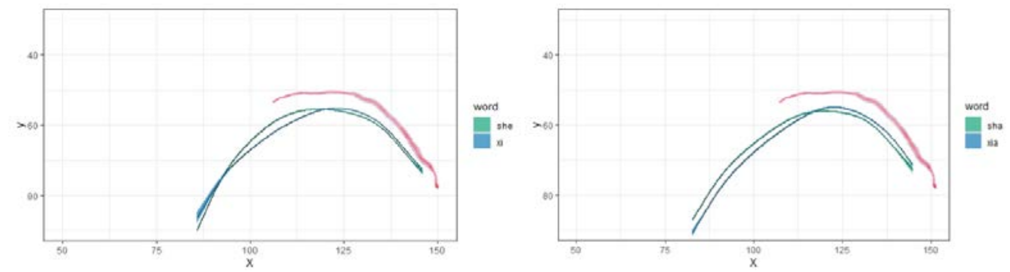
Yung-hsiang Shawn Chang

National Taipei University of Technology (Taiwan)
shawnchang@mail.ntut.edu.tw

Mandarin palatals [ɕ, tɕ, tɕʰ] are post-alveolars or pre-palatals articulated with the blade or the anterodorsum of the tongue [1]. They are found to be among the most difficult consonants for second language learners (e.g., [2, 3, 4]). When taught to English-speaking learners, these three Mandarin consonants are often compared to English palatals /ʃ, dʒ, tʃ/, which are traditionally described as labialized palato-alveolars produced with the front of the tongue [5]. These comparisons, however, have resulted in different pedagogical suggestions. Chin [2] argued that, in teaching, the distinction of these two palatal series should focus on the height of tongue front. Li [6] and Lin [4], on the other hand, suggested that the Mandarin palatals can be taught as their English counterparts *with spread lips.* The current study sought to resolve these discrepant pedagogical solutions, potentially due to the researchers citing articulatory descriptions from different sources, by conducting a detailed comparison of the articulations of Mandarin and English palatals (specifically, palatal fricatives) with ultrasound tongue imaging, linguographic, palatographic, and lip video data.

The stimuli included Mandarin [ɕi, ɕa, ɕoʊ] carrying Tone 4, English /ʃi, ʃa, ʃo/, and non-palatal-initial syllables serving as distractors. All the stimuli were real words. Four native Taiwan Mandarin speakers participated in the experiment. For data collection through the ultrasound and lip camera, the participant wore a head stabilizer, to which the ultrasound transducer, lip camera, and head-mounted microphone were attached. The stimuli were randomized and repeated four times in isolation. Palatographic and linguographic data were collected following [7]. The stimuli were read once in isolation for palatography and linguography respectively. All English productions were verified by two native English listeners to be highly accurate.

The ultrasound data showed that three out of the four subjects' tongue shapes of Mandarin [ɕ] and English /ʃ/, across all three vowel contexts, were nearly identical. One subject (S01) seemed to distinguish the two articulations, in that his Mandarin [ɕ] had a slightly higher and more fronted tongue front than English /ʃ/ (see Figure 1) (cf. Mandarin [ɕ] having considerable raising of the tongue front than English /ʃ/ in [8]. Palatographic and linguographic data confirmed that the three subjects had similar lingual and palatal contact locations for [ɕ] and /ʃ/—laminal post-alveolars. S01's palatograms of [ɕ] (see Figure 2) showed that the narrowest constriction channel for [ɕ] fell within the denti-alveolar region. On the other hand, the constriction channel of S01's /ʃ/ was mainly in the alveolar area. S01's linguograms showed that the lateral contacts of [ɕ] and /ʃ/ on the tongue body were similar, with the narrowest part of the constriction channel being on the blade. Therefore, S01's [ɕ] should be a laminal denti-alveolar, whereas his /ʃ/ was a laminal alveolar. Lip video data showed that for all four subjects the horizontal lip distance between the /ʃ/ production and the resting state was significantly greater than that between [ɕ] and the resting state (see Figure 3 for S01's lip data), indicative of robust lip protrusion in /ʃ/. Taken together, the preliminary findings of this study suggest that the primary and consistent feature distinguishing between Mandarin [ɕ] and English /ʃ/ is lip rounding, instead of lingual and palatal contact locations. These results therefore support [4] and [6]'s assertion that, in pedagogical practices, Mandarin [ɕ] can be considered to be an English /ʃ/ without lip rounding.



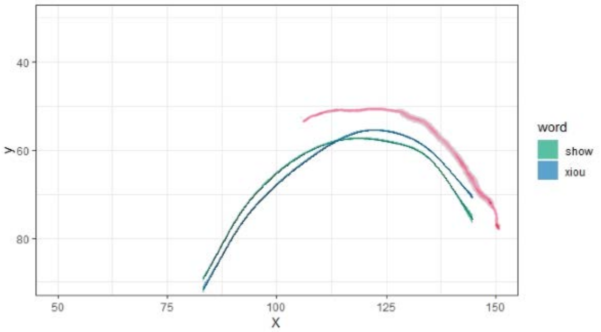


Fig.1 Tongue shapes of [ɕ] and /ʃ/ for S01 (the right side of the image is anterior; palatal contour is traced in red)

Fig.2 S01’s palatograms and linguograms (upper panel: [ɕi, ɕa, ɕo]; lower panel: /ʃi, ʃa, ʃo/)

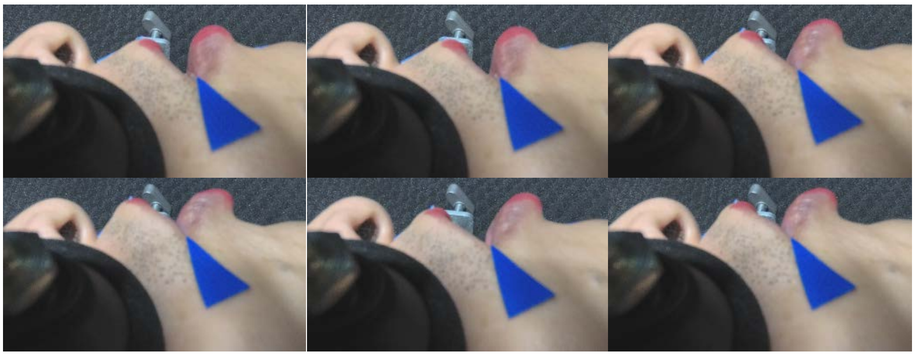


Fig.3. S01’s lip images of [ɕi, ɕa, ɕo] (upper panel) and /ʃi, ʃa, ʃo/ (upper panel)

References

[1] Lee, W. S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109-112.

[2] Chin, Y. L. C. (1972). How to teach Mandarin retroflex and palatal sounds to native speakers of English. *Journal of the Chinese Language Teachers Association*, *7*(2), 77-81.

[3] Chung, R.-F. (2009). *Duibi Fengxi Yu Huayu Jiaoxue [Contrastive Analysis and Chinese Teaching]*. Taipei: Cheng Chung Book.

[4] Lin, H. (2005). Understanding problems in learning Mandarin consonants by monolingual speakers of English. *Journal of Canadian Teachers of Chinese as a Second Language*, *15*(1), 1–18.

[5] Ladefoged, P. (2006). *A Course in Phonetics* (5th ed.). Wadsworth Publishing.

[6] Li, C. W. C. (2003). Phonetic detail in the teaching of Mandarin pronunciation. Presented in the *Chinese Language Teachers Association 2003 Spring Workshop*.

[7] Dart, S. N. (1998). Comparing French and English coronal consonant articulation. *Journal of phonetics*, *26*(1), 71-94.

[8] Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Accented vs native exposure in child second language intelligibility

Margarethe McDonald[1], Eon-Suk Ko[2]

[1]*University of Wisconsin-Madison (USA),* [2]*Chosun University (Korea)*
mlmcdonald3@wisc.edu, eonsukko@chosun.ac.kr

It is generally accepted that *more* exposure to language leads to improvements in child language abilities [1]. However, it is still not clear to what degree quality of the input affects language development. In the case of acquisition of the phonological system of a second language, children often have exposure to both native and accented speakers of the second language. Especially in Korea, there has been a push for more native exposure in children to aid phonological acquisition [2]; however, the improvement of English-speaking abilities as a result of such exposure is difficult to quantify. There have been studies showing children as young as 5-years-old can attune their speech to be more similar to an exposed stimulus in the short term [3]. Therefore, any effects found in the long term may also be evident after short term exposure. This study focuses on both quantity and quality (native vs accented) of English input in the development of English phonological systems in children. Further, we examine if the effects of input are detectable in both the short- and long-term.

Sixty-eight children aged 6-9 years old were recruited to participate in this study. Children were native Korean speakers residing in Korea, with varying degrees of English language abilities. Children participated in an experimental task in addition to multiple language tests in English and Korean. Parents filled out information about the child's language development and detailed information about the child's weekly exposure to both accented and native English. The experimental task was a series of exposure and test trials under two conditions. In the first condition, the exposure stimuli were recorded by a native English speaker and in the second condition, the stimuli were recorded in English by a Korean accented speaker. For exposure trials, participants performed a picture identification task which included 2 images on a screen and an auditorily presented English word. The English words alternated between having an initial phoneme from a set of 2 that are not easily distinguishable in Korean ([b] vs [v]; [p] vs [f]; [l] vs [r]). Following exposure to 4 English words, participants performed 2 test trials which included a picture naming task where they named words from a set of minimal pairs with an initial phoneme from the same difficult to distinguish set (e.g. bee vs vee). Children spoke the words into a microphone on a headset they were wearing, and each participant provided 72 single word recordings to be judged. Condition and order of phoneme sets were counterbalanced.

Recordings were judged, blind to condition, for intelligibility by a native English speaker (first author). The data for four words that we expected our participants to have trouble producing (fan, vee, lock, rock) were used for all analyses. Five participants were excluded due to various reasons. Three analyses were performed to answer our research question. The first analysis looked at the effect of short-term native vs accented English exposure during the course of the experimental task. Intelligibility was assessed with a logistic mixed effect model regressing intelligibility on condition and including a random intercept and slope for condition using the afex package [4] in R. Average intelligibility of the words in the accented condition was 0.74 (SD = .44) and in the native condition was 0.76 (SD = .43) and the model indicated that the addition of condition did not significantly improve the model (χ^2 (1) = .01, p = .94). The second analysis examined how long-term experience with accent affected performance on the task. Here intelligibility of all difficult items regardless of condition were regressed on the overall percentage of weekly English exposure provided according to parent report. The proportion of English exposure was square root transformed to account for the skew of the data. The logistic regression indicated that addition of English exposure significantly improved the model (χ^2 (1) = 13.79, p < .001). Finally, percentage of weekly English exposure was divided into exposure from a native vs accented English speaker. These two variables were entered into logistic regression. Both variables were again square root transformed to account for skew. The addition of native English exposure improved the model significantly when controlling for accented exposure (χ^2 (1) = 10.15, p < .01) and the addition of accented English exposure significantly improved the model when controlling for native exposure (χ^2 (1) = 8.53, p < .01). See Figure 1 for a depiction of the last model. Model coefficients can be used as quasi effect size measures in logistic regressions; the accented exposure (b = .51, SE = .17) had a higher coefficient than native exposure (b = .43, SE = .13).

Our results clearly show that in both short and long-term, differences in native vs accented exposure do not lead to large differences in intelligibility. If anything, accented exposure seemed to be a stronger predictor of intelligibility than native exposure. This is likely the result of differences in how the exposure is received (i.e. in person vs recordings) rather than the type of exposure. Overall amount of exposure played a large role in intelligibility, such that whether or not the exposure is accented or native, more exposure led to higher intelligibility. However, intelligibility is only one way of rating foreign accent. Data coding is ongoing with ratings of accent of each recording being the next item to be analyzed. Ratings of accent are expected to vary much more than ratings of intelligibility since an item can be highly foreign accented and yet still very intelligible. Results with regard to accent will also be discussed.

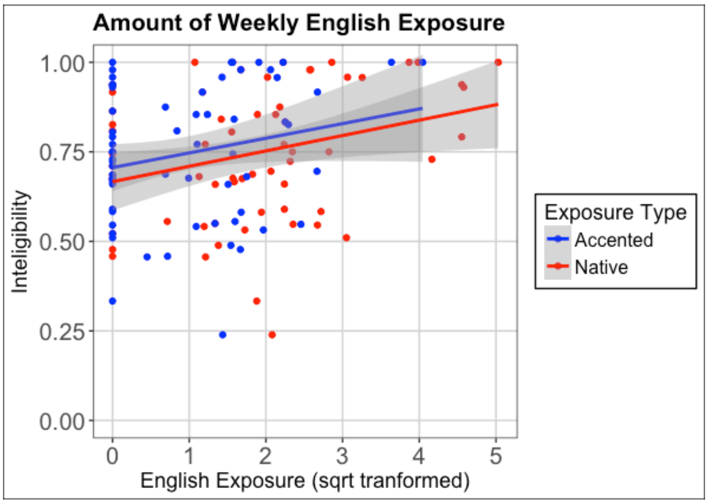


Fig.1 Amount of average weekly exposure to accented vs native English speech. Each point represents one child, grey represent standard errors. Exposure is square root transformed such that a 5 represents approximately 25% exposure on an average week.

References

[1] Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American educator*, *27*(1), 4-9.

[2] Jeon, M., & Lee, J. (2006). Hiring native-speaking English teachers in East Asian countries. English Today, 22(4), 53-58.

[3] Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. Journal of Speech, Language, and Hearing Research, 57(6), 2065-2075.

[4] Singmann, H., Bolker, B., Westfall, J. and Aust, F. (2017). afex: Analysis of Factorial Experiments. R package version 0.18-0. https://CRAN.R-project.org/package=afex

Prosodic marking of neutral and non-neutral refusal in Russian: an identification experiment

Pavel Duryagin

Ca' Foscari University of Venice (Italy)
pavel.duryagin@unive.it

As Standard Russian belongs to the group of languages in which the intonational contrast between yes/no questions and statements is marked by pitch accent, the prosodic means of marking this difference were repeatedly studied experimentally. In particular, [1] showed that the perceptual cues for Russian polar questions are steep f0 rise and late peak alignment, while [2] described them as higher f0 peak, peak alignment around the offset of stressed syllable and presence of the low turning point at the onset of the accented syllable. Our earlier experimental study of Russian utterance "*Да ну"* /daˈnu/ (an idiomatic expression with the illocutionary meaning of disagreement or refuse) showed that speakers of Russian consistently used similar prosodic features to mark different forms of negation. Namely, later f0 peak alignment, higher f0 peak frequency and longer stressed vowels were used by the speakers to mark polite, non-categorical negation (e.g., in the context of refusal in response to an offer of help or in a context of disagreement due to positive reasons).

To test whether these means of marking semantic differences can be generally perceived by the speakers of Russian an identification experiment was conducted. Russian phrase "*Не надо*" /nʲiˈnada/ (*'there is no need to'*) was chosen as a stimulus because of its segmental and syllabic structure. First, three different productions of this phrase by native Russian male speaker were recorded: a neutral statement, a "polite" refusal and a yes/no question. The former two tokens served for training session and as control stimuli. The "neutral production" of the utterance served as a base for overlap-add manipulation in *Praat* [3].

Three acoustic parameters were manipulated: f0 peak frequency, f0 peak alignment and stressed vowel duration. As both [1] and [2] reported that the configuration of the slope of the pre-accentual rise contributes to the opposition between different accent types, the temporal distance between the slope low turning point position and f0 peak was fixed (however, minor steepness differences caused by peak height and vowel length manipulations were ignored). The three peak height levels chosen were 150, 180 and 210 Hz, the five peak alignment points were stressed syllable onset (1), stressed vowel onset (2), ⅓ and ⅔ of the stressed vowel duration (points 3 and 4) and stressed vowel offset (5). In addition, every contour was recreated with modified duration of the stressed vowel (with 33% and 66% greater duration).

All 45 experimental stimuli, as well as control stimuli, were presented in random order to 25 native speakers of Russian (18 F, 7 M). A short training preceded every performance. Participants were asked to listen twice to a stimulus and identify it as a question or a statement. In case the utterance was identified as a statement, the listener was asked to evaluate its "politeness" on a 5-point Likert-type scale, where 1 referred to a "very categorical, almost rude refusal", 2 – "rather categorical, rude refusal", 3 – "neutral response", 4 – "rather polite refusal", 5 – "very polite refusal". Informants also could mark phrases as "unnatural" and not interpret them. The experiment was designed in *PsychoPy* software [4].

The results for "question" judgments generally replicate the findings of [1]. Figure 1 illustrates the interaction between manipulated parameters and the number of "question judgments". Chi-square tests of independence confirmed statistically significant relations between the number of "question" judgments and peak alignment (χ^2 (2) = 413.05, $p < .01$) and peak height (χ^2 (2) = 80.09, $p < .01$). Later and higher f0 peaks conditioned more "question" judgments. No significant interaction between vowel duration and the number of "question" judgments was found (χ^2 (2) = 1.24, $p = .54$).

The main scope of the study was to estimate the relations between the three manipulated factors and "neutral" vs. "non-neutral" statement (Likert scale 3 vs. 1, 2, 4, 5) and "rude" vs. "polite" statement judgments (Likert scale 1, 2 vs. 4, 5). The results for the non-interrogative judgments are presented in Figure 2 (diverging stacked bar charts were created by means of “HH” package in R [5]; responses for late alignment points, 4 and 5, are omitted here).

For “neutral” vs. “non-neutral” opposition, chi-square tests showed significant interaction between the judgments and all three manipulated factors ($p < .01$; χ^2 (2) = 17.72 for vowel length, χ^2 (2) = 32.19 for peak height, χ^2 (2) = 35.76 for peak alignment with points 4 and 5 excluded from the analyses). Post hoc comparisons with Bonferroni adjustments revealed that significantly larger number of “neutral” responses was conditioned by the stimuli with the earliest peak (point 1, as opposed to points 2 and 3) and medium (133%) vowel duration, as opposed to short (100%) and long (166%) vowels. Increasing peak frequency consistently significantly reduced the number of “neutral” judgments.

As for the “rude” vs. “polite” dichotomy, post-hoc tests revealed significant effects of vowel duration (χ^2 (2) = 65.9, $p < .01$) and peak alignment (χ^2 (2) = 20.69, $p < .01$) and marginal effect of peak height (χ^2 (2) = 9.0891, $p = .01062$). Post hoc comparisons with Bonferroni adjustments show strong interaction between vowel duration and “politeness” (stimuli with short vowels were more often considered “rude” and long vowels caused “polite” judgments). The earliest peak position significantly more often caused “rude” judgments than peak points 2 and 3.

The results of the experiment partly support the hypothesis that “polite refusal” in Russian is marked by the combination of peak alignment, peak height and vowel duration. The only parameter that consistently increased the number of “polite” judgments was vowel duration. However, the obtained judgments based on early vs. medial peak alignment and f0 peak height fall in line with the earlier experimental findings showing the common informational interpretations of the “frequency code” [6].

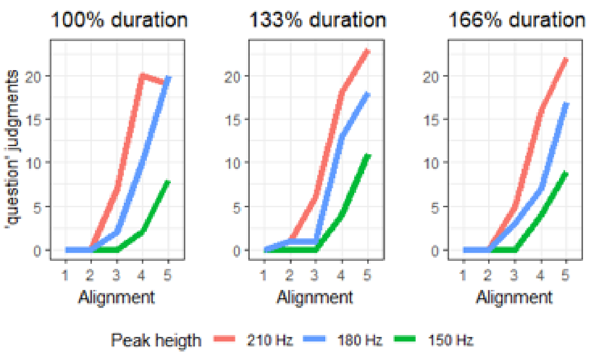


Fig.1 Interaction between 'question' judgments and three manipulated parameters

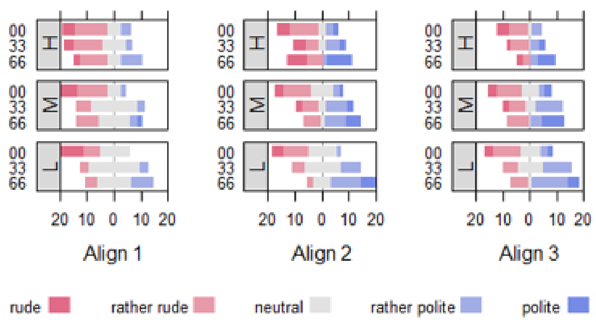


Fig.2 Number of Likert-scale 'statement' responses plotted against the three manipulated parameters

References

[1] Rathcke, T. V. (2006). A perceptual study on Russian questions and statements. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK), 37*, 51–62.

[2] Igarashi, Y. (2006): Intonation patterns in Russian interrogatives - Phonetic analyses and phonological interpretations. In: Y. Kawaguchi et al. (eds): *Prosody and Syntax, Usage-Based Linguistic Informatics 3*. Amsterdam, Philadelphia: John Benjamins, 175-196.

[3] Boersma, P. & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.49, retrieved 2 March 2019 from http://www.praat.org/.

[4] Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*.

[5] Heiberger, R. M. & Robbins, N. B. (2014). Design of Diverging Stacked Bar Charts for Likert Scales and Other Applications. *Journal of Statistical Software*, 57(5), 1-32.

[6] Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: CUP.

Articulation and Neutralization: Inherent and Derived Palatals in Korean

Jae-Hyun Sung

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

Korean exhibits a lexicalized palatalization process in which an alveolar noncontinuant obstruent (/t, t^h/) becomes palatalized (i.e., becoming [ʧ] and [$ʧ^h$] respectively) before a high front vowel or a palatal glide, but only across a morpheme boundary [1,2,3]. As shown in Table 1, /mat+i/ ‘the eldest’ and /mati/ ‘joint, node’ show very similar underlying representation in Korean. However, only /mat+i/ ‘the eldest’ undergoes lexical palatalization because of its morphological structure showing a compound boundary between /mat/ ‘eldest’ and /i/ (nominalizer), while /mati/ ‘joint, node’ does not.

Word	Effect	Result
/mat+i/ ‘the eldest’	/t/ → [ʧ]	lexical palatalization
/mati/ ‘joint, node’	/t/ → [t] or [t^j]	no lexical palatalization; potential coarticulation

Table.1 Palatalization in Korean.

This lexical palatalization creates a sound change in which the resulting palatals from the palatalization process are perceptually the same as the underlying palatals. This fact raises the question of whether true inherent palatals (i.e., underlyingly /ʧ/ or /$ʧ^h$/), and apparent, derived palatals (i.e., derived from /t/ or /t^h/) are truly neutralized in the production of Korean speakers. Given the previous claim that multiple articulations create perceptual neutralization [4], it is possible that inherent and derived palatals in Korean may be articulated differently in a covert fashion, but perceptually indistinguishable. Using ultrasound imaging, our study investigates the articulation of inherent and derived palatals [ʧ] and [$ʧ^h$] in Korean to see if palatals from the two different sources truly create a phonetic merger. Examples of test words are illustrated in Table 2.

Word	Effect	Result
/pata/ ‘sea’ /paʧi/ ‘pants’ /mat+i/ ‘the eldest’	/t/ → [t] /ʧ/ → [ʧ] /t/ → [ʧ]	no palatalization (control) inherent palatal derived palatal
/ot^ha/ ‘typo’ /ka$ʧ^h$i/ ‘value’ /pat^h#i/ ‘field+NOM’	/t^h/ → [t^h] /$ʧ^h$/ → [$ʧ^h$] /t^h/ → [$ʧ^h$]	no palatalization (control) inherent palatal derived palatal

Table.2 Examples of test words.

Representative tongue contours from 11 native speakers of Korean (shown in Figures 1 and 2) were compared using SSANOVA [5,6]. The results show that although speakers show individual variation in the relationship between their non-palatal and palatal/palatalized tongue contours, the majority of speakers do maintain distinction between inherent and derived palatals. The results reported here provide articulatory evidence of contrast between two types of palatals.

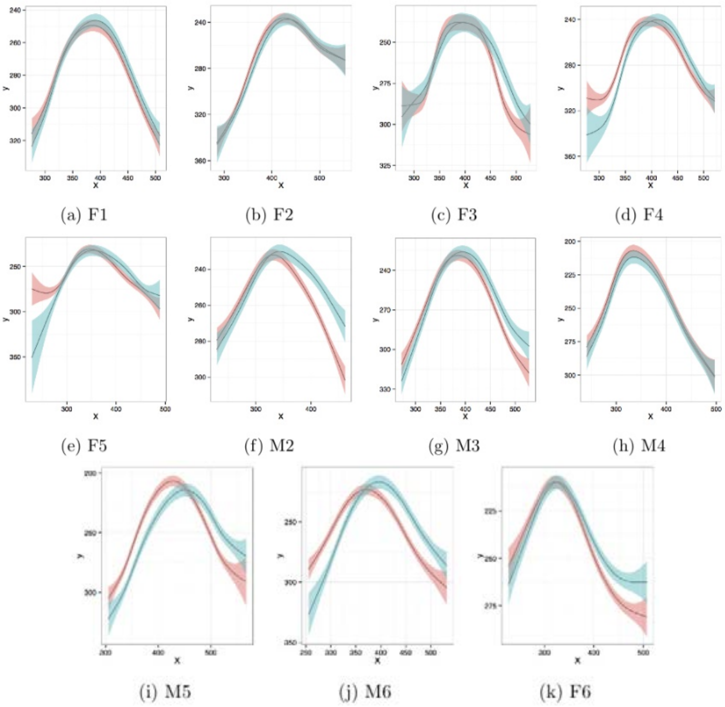


Fig.1 Ultrasound tongue contours: inherent (red) and derived (blue) [ʧ]; Tongue tip is to the right, and shares represent 95% confidence interval. Axis values correspond to pixels.

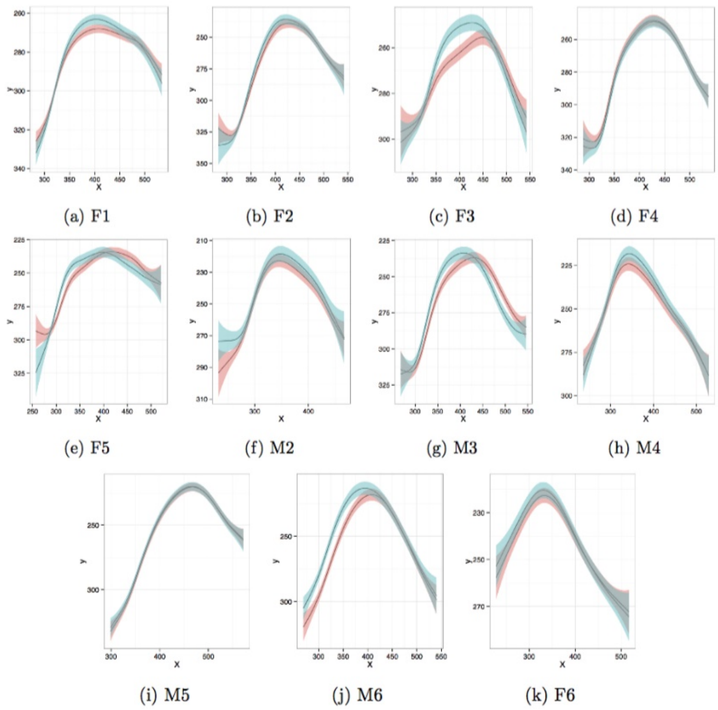


Fig.2 Ultrasound tongue contours : inherent (red) and derived (blue) [$ʧ^h$].

References

[1] Kim, S-G. (1976). *Palatalization in Korean* (Unpublished doctoral dissertation). University of Texas at Austin.
[2] Ahn, S-C. (1985). *The Interplay of Phonology and Morphology in Korean* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
[3] Sohn, H-M. (2001). *The Korean Language*. Cambridge : Cambridge University Press.
[4] Twist, A., Baker, A., Mielke, J., & Archangeli, D. (2007). Are ‘covert’ /ɹ/ allophones really indistinguishable ? *University of Pennsylvania Working Papers in Linguistics*, *13*, 207-216.
[5] Gu, C. (2002). Smoothing Spline ANOVA Models. New York : Springer.
[6] Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of Acoustical Society of America*, *120*(1), 407-415.

Poster Presentations
(Day 2)

Do Japanese learners distinguish prosodic levels in French?

Rachel Albar, Hiyon Yoo

UMR 7710 Laboratoire de Linguistique Formelle, Paris Diderot University (France)
rachel.albar@univ-paris-diderot.fr, yoo@linguist.univ-paris-diderot.fr

French is generally described as a phrase language, with at least two prosodic levels above the word: AP (accentual phrase) and IP (intonational phrase) (among others, [1], [2], [3]). Michelas [4] argues the need for an intermediate phrase (*ip*) between the AP and the IP, containing more than two APs and corresponding to some specific syntactic boundaries. She shows that a pitch reset occurs at the end of the *ip* while the final stressed vowel is statistically longer compared to a simple AP position, and that these differences are perceptually distinguished.

One question that arises is how learners deal with these different prosodic levels. The aim of this study is to focus on the production and the perception of these prosodic levels by L2 learners whose mother tongue's prosodic structure is quite different. Indeed, Japanese, unlike French, is a language with a tonal accent [5]. The prosodic structure has only APs and IPs, related to the presence of a tonal accent, since an AP carries the tonal accent while an IP is mostly characterized by downstepping [6]. The prosodic IP level corresponds to the XP syntactic category [7] which is similar to the French *ip*. [8] Moreover, in Japanese there is no notion of continuation at prosodic boundaries since each AP generally ends with a low tone [9]. If the notion of continuation lacks, and if the prosodic structure is different in the two languages, are Japanese learners able to deal with the prosodic structure of French?

Our main hypothesis is that Japanese learners will not be able to distinguish the *ip* unlike native speakers. Our second hypothesis is that despite this difficulty, learning these prosodic levels is possible and that learners with a higher level of French will obtain better results.

In order to test our hypotheses, we built a sentence completion task, based on Michelas [4]. This task highlights, among other things, the use of phonological structures during the syntactic analysis of a sentence [10]. In our experiment, it shows subjects' ability to perceive the difference between the AP and the *ip* prosodic boundaries, and to associate the corresponding syntactic structure. After listening to the beginning of a statement, participants were asked to select the sequence of words they thought most likely to follow what they had just heard from two options (see Figure 1). 9 stimuli for each prosodic condition (AP and *ip* boundary) and 18 fillers were constructed by taking account that participants were learners (use of accessible vocabulary). Stimuli were recorded by native speakers of French and were truncated at the end of the second AP.

38 Japanese learners of French and 12 native French speakers (as a control group) participated in this experiment. Results were analysed using a generalised linear mixed model (GLMM) with proficiency and type of prosodic boundary as fixed factor, item and subject as random effects. They show that learners' score is significatively lower than French participants in the sentence completion task. However, contrary to our initial hypothesis, there is no correlation between the learners' level and their rate of correct answers.

In fact, for Japanese learners, associating different continuous contours to a specific syntactic structure seems to be a difficult task. Prosody is not taught, and a possible explanation for the lack of improvement in results with language level could be that their exposure to French takes place almost exclusively in the classroom and would not be sufficient for the acquisition of subtle prosodic parameters. We plan to repeat the experiment with learners living in France for a significant period of time, in order to show whether immersion in the L2 context gives better results in distinguishing this intermediate prosodic level.

The sentence completion task was complemented with the analysis of production read data of Japanese speakers. We analysed the prosodic parameters of F0 rise and syllable duration in a corpus that consisted of the production of small excepts of texts (4 texts ~50 words per text). 17 Japanese learners were recorded and we used pre-existing recordings of the same texts read by 5 French speakers as control group [11]. As for perception, we observed that Japanese speakers had difficulties to differentiate the prosodic boundaries and to produce them accordingly. For example, AP boundaries were frequently produced with a greater F0 rise than IP boundaries. Our results show the different prosodic levels of French seem to be challenging for Japanese learners in both perception and production.

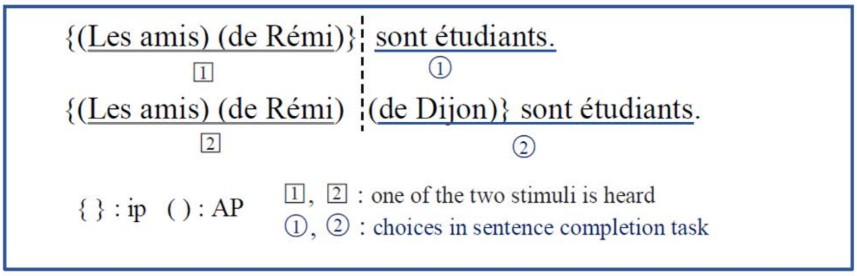


Fig.1 Sentence completion task's stimuli example "Rémi's friends (from Dijon) are students"

References

[1] Delais-Roussarie, E., & Post, B. (2008). Unités prosodiques et grammaire de l'intonation: vers une nouvelle approche. *Actes des Journées d'étude sur la Parole JEP-TALN*, *8*.

[2] Delais-Roussarie et al. (2015) Intonational Phonology of French: Developing a ToBI system for French In S. Frota & P. Prieto (Eds.), Intonation in Romance. OUP: Oxford. 63-100.

[3] Jun, Sun-Ah & Cécile Fougeron (2000) A phonological model of French intonation. In Intonation: Analysis, Modeling and Technology, ed. by Antonis Botinis. Kluwer Academic Publishers. 209-242.

[4] Michelas, A. (2011). *Caractérisation phonétique et phonologique du syntagme intermédiaire en français: de la production à la perception* (Thèse de doctorat, Université de Provence-Aix-Marseille I).

[5] Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In *Prosodic typology: The phonology of intonation and phrasing*, S.-A. Jun (Ed.), Oxford: Oxford University Press. 172-200.

[6] Igarashi, Y. (2015). Intonation. In Kubozono, H. (ed.) *Handbook of Japanese phonetics and phonology,Berlin: De Gruyter Mouton*. 525-568.

[7] Selkirk, E. (2009). On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Gengo Kenkyu*, *136*, 35-73.

[8] D'Imperio, M., & Michelas, A. (2010). Embedded register levels and prosodic phrasing in French. In *Speech Prosody* (p. 4).

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

[10] Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. Language and speech, 51(1-2), 61-75

[11] Santiago, F. & E. Delais-Roussarie. (2015). The acquisition of Question Intonation by Mexican Spanish Learners of French. In E. Delais-Roussarie, M. Avanzi & S. Herment (eds), *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations.* Heldelberg: Springer, 243–270.

The representation of variable tone sandhi in Shanghai Chinese

Hanbo Yan[1], Yu-Fu Chien[2], Jie Zhang[3]

[1]*Shanghai Intl. Studies Univ. (China),* [2]*Fudan Univ. (China),* [3]*Univ. of Kansas (USA)*

yanhanbo@shisu.edu.cn, chien_yc@fudan.edu.cn, zhang@ku.edu

Introduction. Variation in phonological patterns challenges models of spoken word recognition. One crucial question is how listeners process and represent variant forms in the mental lexicon. It has been shown in behavioral experiments that more frequent forms were responded to more quickly than less frequent forms (e.g., American English word-final *t/d* deletion, Deelman & Connine, 2001). Ranbom & Connine (2007) further showed that in *nt*-flapping in English (e.g., [sɛɾɚ] for *center*), the less frequent form [nt] also had a strong phonological representation in the lexicon, suggesting that both forms of a phonological alternation contribute to lexical access. These studies, however, mainly focused on variation of phonetic reduction processes, which can be interpreted as processes of late phonology (Coetzee & Pater, 2013). The processing of morpho-syntactically conditioned phonological alternations, which presumably occur earlier in the derivation, has received considerably less attention. The current study investigates one such alternation — a variable tone sandhi pattern in Shanghai Chinese.

The majority of disyllabic verb-noun (V-N) combinations in Shanghai Chinese can undergo a tonal extension tone sandhi, whereby the base tone of the first syllable is spread onto the entire disyllable. But the sandhi applies variably. For example, /tsʰɑ̃24/ “to sing” + /ku53/ “song” can be realized with either the sandhi form ([tsʰɑ̃33 ku44]) or the non-sandhi form ([tsʰɑ̃24 ku53]). In a series of subjective rating experiments, Yan (2016) found that native speakers generally preferred the non-sandhi form over the extension sandhi form in V-N, and that more frequent V-N items preferred the non-sandhi form even more.

However, how V-N items with variant sandhi forms are processed during spoken word recognition remains unclear. To this end, we designed an auditorily primed lexical decision experiment in which disyllabic V-N targets were primed by a monosyllable that shared the segments of the first syllable with different tonal conditions. Based on Yan (2016), we expect the prime with the non-sandhi tone to elicit facilitative priming in general, suggesting that the non-sandhi form is represented in the mental lexicon. Moreover, we also predict that the priming effect may be modulated by the familiarity ratings of the V-N items, with the non-sandhi tone prime facilitating the recognition of more frequent V-Ns, and the sandhi tone prime facilitating the recognition of less frequent ones. These results would suggest that both variant forms of a V-N item are represented in the mental lexicon and activated to different degrees depending on frequency of the item.

Methods. Thirty-six native Shanghai speakers performed a lexical decision task. Each disyllabic target was preceded by a sandhi tone prime, which shared the same tone with the initial syllable of the extension sandhi, a non-sandhi tone prime, which had the same underlying tone as the first syllable of the non-sandhi form, and an unrelated tone prime, which was not related to the first syllable of the target in tone (see Table 1). Given that the extension sandhi form applies variably in V-N disyllables, the stimuli-recording speaker was asked to produce the disyllables as naturally and comfortably as possible without considering variation. She applied the non-sandhi form to all V-N critical targets, which is consistent with Yan (2016). Hence, the non-sandhi form was used for all the V-N targets in the current study. All primes shared the same segments with the first syllable of the target. After the priming experiment, a familiarity rating task on the disyllabic targets was conducted with the same participants.

Table 1. Examples of the three prime types sandhi target

53 + 13 44 + 13 (non-sandhi) or 55 + 31 (tonal extension)		
Prime Type		Target
Sandhi Prime	[sɔ55]	[sɔ44 vɛ13] /sɔ53 vɛ13/ “to cook a meal”
Non-sandhi Prime	[sɔ53] “a little”	
Control Prime	[sɔ24] “to sweep”	

Results and discussion. The overall accuracy rate for the V-N stimuli, including both critical and filler words, is 85% (1662/1944 trials). For the reaction time analyses on the critical stimuli, inaccurate responses (6%, 42/648 trials) and responses over two standard deviations from the mean reaction time of critical words (5%, 35/648 trials) were excluded. To reduce the skewness of the raw data, reaction times were log-transformed, and then modeled with a series of Linear Mixed-Effects models with *participant* and *item* as random effects, *prime* and *familiarity* as fixed effects, and by-subject random slopes for *prime*.

Model comparisons under likelihood ratio tests showed that the best model is the model with *prime***familiarity*. The interaction can be seen in Fig.4. Although both Sandhi and Non-sandhi primes facilitate V-N recognition in general compared to the Control prime (Fig.4a), *familiarity* regulates the priming effects of different *prime* types (Fig.4b). The model shows that the slope of *familiarity* regression line for the Control primes is significantly less negative than that for the Non-sandhi primes (EST.=-0.035, SE=0.014, t=-2.440, p=0.015) and marginally significantly less negative than that for the Sandhi primes (EST.=-0.027, SE=0.014, t=-1.886, p=0.060).

(a) (b)

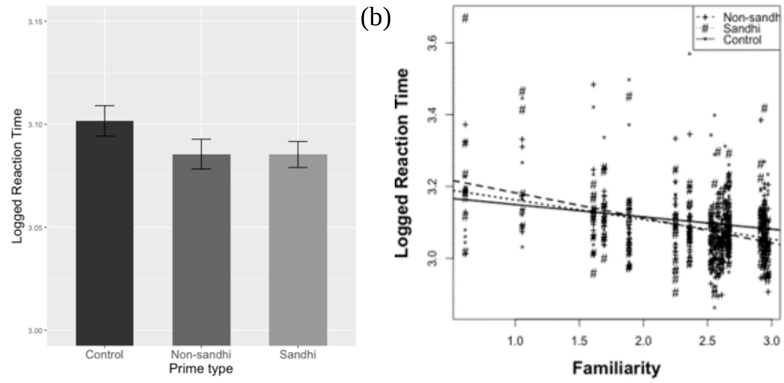


Fig.4 (a) Logged reaction times and standard error bars for control, non-sandhi, and sandhi primes in the lexical decision task. (b) Reaction time elicited by the non-sandhi (+), sandhi (#), and control (*) prime as a function of familiarity for Shanghai V-N items (each point represents a target word primed by a certain type of prime, which was differentiated by symbols).

These data suggest that both the underlying and tonal-extension forms are represented in the mental lexicon due to tone sandhi variation. The finding that the more frequent variant — the non-sandhi form — has an effect on V-N recognition is in accordance with previous findings (e.g., Connine, 2004). The fact that the Non-sandhi primes have the most negative slope for *familiarity* is consistent with our prediction that Non-sandhi primes are more likely to have a facilitation effect for V-N items with higher frequency. But our prediction that less frequent V-N items should be more facilitated by the Sandhi primes was not borne out, and it is unclear to us why we obtained an inhibition effect for both Sandhi and Non-sandhi primes for items with lower familiarity ratings.

Taken together with Yan et al.'s (in revision) earlier finding that for modifier-noun (M-N) combinations in Shanghai, in which the extension tone sandhi applies obligatorily, their recognition is primarily primed by sandhi primes, these results indicate that both grammatical factors, such as syntactic structures, and usage factors, such as frequency, contribute to the representation and processing of tone sandhi forms in Shanghai. Like variation as the result of reduction in late phonology, variation that is syntactically sensitive should be represented in the lexicon as well.

References

[1] Deelman, T., & Connine, C. M. (2001). Missing information in spoken word recognition: Non-released stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 656-663.

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

[3] Coetzee, A. W., & Pater, J. (2011). The place of variation in phonological theory. In J. Goldsmith, J. Riggle, & A. Yu (eds.), *The Handbook of Phonological Theory: 2nd Edition*, 401-434. Cambridge: Blackwell.

[4] Yan, H. (2016). *The nature of variation in tone sandhi patterns of Shanghai and Wuxi.* Dissertation, University of Kansas.

[5] Connine, C. M. (2004). It's not what you hear, but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review,* 11, 1084-1089.

[6] Yan, H., Chien, Y.-F., & Zhang, J. (in revision). Priming the representation of left-dominant sandhi words: A Shanghai dialect case study.

Perturbation effects in Chongming Chinese with and without focus

Si Chen, Eunjin Chun, Bei Li & Yike Yang

The Hong Kong Polytechnic University (Hong Kong)
qinxi3@gmail.com, eunjin.chun@polyu.edu.hk, 199122libei@gmail.com, yi-ke.yang@connect.polyu.hk

Focus is a linguistic means used to highlight or emphasize part of an utterance [1]. Several phonetic cues have been reported in the realization of prosodic focus, including variations in f0 patterns, intensity, and duration. According to [2], voiceless consonants interrupt the first 40% of the target syllables, and the onset f0 of the vowels following the voiceless aspirated bilabial stop is lower than other consonants. [3] investigated the consonant-f0 interaction in Shanghai Chinese, where there is a three-way laryngeal contrast (voiced, voiceless aspirated and voiceless unaspirated), and speakers voluntarily exaggerated the contrasts among three types of consonants, which the author attributed to both phonological contrasts and phonetic consequence on f0 perturbation.

Chongming Chinese is a northern Wu dialect, with eight contrasting tones as shown in Table 1. Although Chongming Chinese is reported to have a three-way contrast among onset obstruents (voiceless aspirated, voiceless unaspirated and voiced), tonal contrasts only exist for voicing distinction but not for aspiration distinction [4,5]. According to [6,7], Chongming Chinese showed post-focus compression in certain tones, but the perturbation effects under focus remain to be investigated. The current study aims to investigate the phonetic perturbation effects induced by different types of onset consonants under no-focus, on-focus and post-focus conditions.

Because the vowel [æ] bears the largest range of tones for different onsets in Chongming Chinese, twelve monosyllabic words, all in combination of different stop onsets and the vowel [æ], were selected as the target words. The target words were then embedded in the same sentence structure under three focus conditions: no focus, on-focus and post-focus. Different preceding and following syllables were also manipulated, so under each focus condition, the target word appeared in four sentences. There were 1,728 sentences in total (12 target monosyllables (8 tones) * 3focus conditions * 12 speakers * 4 sentences). The vowel portions of the recordings were first segmented, and f0 values were extracted at 20 normalized time points using the ProsodyPro Praat script [8]. After normalizing f0 values, we applied functional data analysis [5] to model f0 contours and compare pairs of contours, such as contours of the voiceless aspirated syllable versus the voiceless unaspirated syllable under the post-focus condition.

We listed differences found in f0 contours across tones after three contrasts of consonant onsets (i.e., voiceless unaspirated, voiceless aspirated, and voiced) under no focus, on-focus and post-focus conditions. In order to determine whether f0 contours after two different pairs of consonantal onsets are different, we performed functional t-test on each pair (i.e., voiceless unaspirated ([tæ]) vs. voiceless aspirated ([t^hæ]), voiceless unaspirated ([tæ]) vs. voiced ([dæ]), and voiceless aspirated ([t^hæ]) vs. voiced ([dæ])). According to the comparison, native speakers of Chongming Chinese showed significant perturbation effects after three contrasts of consonant onsets in the no-focus condition. However, under focus and post-focus conditions, there were less consonant perturbation effects. We also explored whether this perturbation effect can differ under on-focus and post-focus conditions. The results of this analysis revealed that perturbation effect in Chongming significantly changed, regardless of tones or consonant types in these conditions.

Our findings were contrary to the findings of Shanghai Chinese reported by [3]. The phonetic consonant perturbation effects of Chongming Chinese are suppressed under the on-focus condition for all tones. In the post-focus position, more consonant perturbation effects were found only for T5, while less effects were observed in other tones. The results indicate that speakers can voluntarily control articulatory settings not only to enhance the contrasts, but also to suppress them if f0 was used to indicate focus at the same time. It remains to be investigated whether speakers adjusted the weight of acoustic cues under focus. They may have enhanced a different acoustic cue, such as voice onset time, to signal the consonant contrasts under focus and suppressed the use of f0, since f0 curves were changed to indicate where the focus lay in. Also, due to post-focus compression, it is likely that consonant perturbation effects were compressed in Chongming Chinese.

Middle Chinese categories		**Ping (Level) Even**	**Shang (Rising) Oblique Oblique**	**Qu (Departing)**	**Ru (Entering) Even**
Chongming allotones	High register	1 (53) H	3 (435(424)) HMH	5 (33) M	7 (55(5)short) Hʔ
	Low register	2 (24) LM	4 (241(242) LML	6 (213(313)) MLM	8 (23(2)short) Lʔ

Table 1. Eight tones in Chongming dialect (H, M, L representations)

Comparison	**No Focus**		**Focus**		**Post-focus**	
	Difference	Significant proportion	Difference	Significant proportion	Difference	Significant proportion
ta1 vs tha1	different	36-70%	Not different	N/A	Not different	N/A
ta3 vs tha3	different	1-22%	different	1-15%	different	1-20%
ta5 vs tha5	different	1-14%	different	1-8%	different	1-24%
ta7 vs tha7	Not different	N/A	Not different	N/A	Not different	N/A

Table2. Results of contour comparisons in the focus condition

References

[1] Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of linguistics*, *19*(2), 377-417.
[2] Wong, Y. W. & Xu, Y. (2007). Consonantal perturbation of f0 contours of Cantonese tones. In *Proceedings of The 16th International Congress of Phonetic Sciences* (pp. 1293-1296). Saarbrücken, Germany.
[3] Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics, 39*(4), 612-625.
[4] Chen, M. & Zhang, H. M. (1997). Lexical and postlexical tone sandhi in Chongming. In: J., Wang, & N. Smith (eds.), *Studies in Chinese Phonology* (pp. 13–52). Mouton de Gruyter, Berlin and New York.
[5] Chen, S., Zhang, C., McCollum, A. G., & Wayland, R. (2017). Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal languages. *Speech Communication, 88*, 17-38.
[6] Yang, Y., Chen, S., & Li, K. (2018). Pitch realization of post-focus components in Chongming Chinese. *The Journal of the Acoustical Society of America*, *144*(3), 1938-1938.
[7] Yang, Y., Chen, S., & Li, K. (2019). Effects of Focus on Duration and Intensity in Chongming Chinese. In *Proceedings of The 19th International Congress of Phonetic Sciences*. Melbourne, Australia.
[8] Xu, Y. (2013). ProsodyPro — A tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody 2013* (pp. 7-10). Aix-en-Provence, France.

Prosodic cues in the perception of Cantonese sarcasm

Chen Lan, Peggy Mok

The Chinese University of Hong Kong (Hong Kong)
lchsapphire@gmail.com, peggymok@cuhk.edu.hk

Verbal irony has generally been described as a rhetorical device for either implying the opposite of what the content is literally [2], or expressing a different meaning from what is said [7]. Sarcasm is referred to as one type of irony, the ironic criticism, that uses positive contents to deliver negative meanings [6].

A previous study suggested that prosodic properties such as pitch, duration, and amplitude were essential cues to distinguish sarcasm and non-sarcasm in Cantonese [3]. More recently, a perceptual study [4] compared how Cantonese and English speakers interpreted the sarcastic speech in their native and non-native language, indicating that acoustic markers such as the fundamental frequency (F0) played an essential role in recognizing sarcasm for Cantonese speakers. Except for [4], to our knowledge, no published research studied the prosodic cues of sarcasm in Cantonese from the perspective of speech perception. Furthermore, previous studies about the perception of the sarcastic intonation usually made use of the stimuli being produced by speakers with professional acting experience (e.g. [8]) or being validated to have the intended meanings (e.g. [3, 4]). Little is known about whether the listeners are able to distinguish sarcasm and non-sarcasm in more natural conversational condition. This study examined how prosodic features signal the sarcastic tone produced spontaneously in Cantonese and how well native Cantonese speakers can perceive them.

Six native Hong Kong Cantonese speakers, who were undergraduate students, were recruited to record the stimuli. They were provided with short scenarios with positive or negative situations commonly happened in daily life and produced the target utterances according to the contexts. There were two sets of target utterances, and within each set, the sentence structure was controlled. As exemplified in Table 1, the first set contained the utterances with a degree modifier, an adjectival phrase, and a sentence final particle. An intensifier /ʦɐn˥hɐi˧/ 'really' was added to create the second set of the target utterances. To sum up, 96 target utterances including two attitudes (sarcastic and sincere) and two sentence sets (with and without the intensifier) were recorded. Twenty-six native Cantonese speakers participated in a perception task individually. The participants were asked to listen to the stimuli and rate for each target utterance on a 6-point Likert scale from 1 to 6 whether they perceived the sentence as being produced with a very sincere (1) or very sarcastic (6) tone of voice. In total, the rating scores from 2496 responses (96 target utterances × 26 participants) were analysed statistically. Two-way ANOVAs with repeated measures were conducted considering two factors: Attitude and the sentence Set.

The recorded utterances were analysed acoustically in Praat [1], and the results showed that, compared to the sincere production, sarcasm was significantly marked by a reduction in the speech rate, a lower mean F0, a lower mean amplitude, an enlargement of the amplitude-range, and an increase in the HNR (see Fig. 1). Fig. 2 summarizes the mean scores of the perceptual ratings for two attitudes across sets. A significant interaction between Attitude and Set was found ($F(1,623) = 8.314$, $p = .004$), together with main effects for the two factors: Attitude ($F(1,623) = 1096.497$, $p < .001$); Set ($F(1,623) = 57.344$, $p < .001$) in listeners' ratings. Similar significant differences in listeners' ratings between sarcasm and sincerity were found across sentence sets. This finding suggests that listeners were able to discriminate sarcasm from sincerity cueing by the prosodic features alone, since the sentence structure was controlled, and the only difference between the sarcastic and sincere utterances should be the speakers' tone of voice. Additional examination was conducted across two sentence sets, showing that sarcastic utterances with the intensifier were rated significantly more sarcastic than those without the intensifier ($p < .001$). Possible explanation is that, in Cantonese, this intensifier was used frequently for expressing criticism [5], which may be considered as a semantic cue by the listeners in addition to the acoustic cues.

To conclude, prosodic features cueing the interpretation of Cantonese sarcasm by the native speakers were investigated. Extensive examination on how each of these acoustic variables is responsible for the perception of Cantonese sarcasm is currently underway. More participants are also needed to corroborate our current findings.

Table 1. Example of the contexts (1. negative; 2. positive) and the target utterances with English translations (a. sentence without intensifier; b. sentence with a target intensifier).

Contexts	
1.	**What? Was yesterday the deadline for course registration? I thought it would be due today.**
2.	**It's raining. I know you have not taken your umbrella with you, so I bring one for you.**
Target utterances	
a.	**你好醒呀 [You are so smart]**
b.	**你真係好醒呀 [You are really (so) smart]**

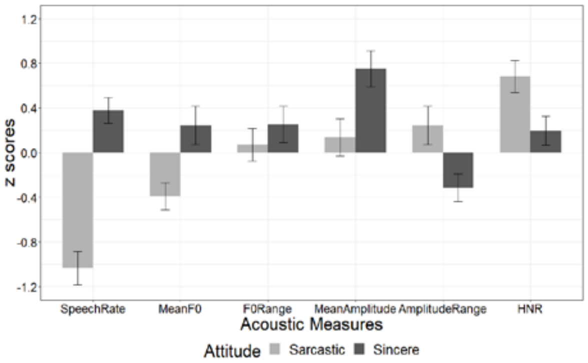


Fig.1 Mean values (z-scores) of the six acoustic variables across two attitudes. Error bars indicate the standard errors.

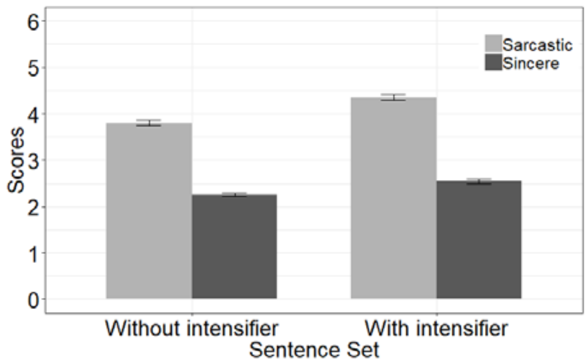


Fig.2 Mean rating scores of two attitudes across sentence sets. Error bars indicate the standard errors.

References

[1] Boersma, P., Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved from http://www.praat.org/.

[2] Brown, P., Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In *Questions and politeness strategies in social interaction*.

[3] Cheang, H. S., Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America, 126*(3), 1394–1405.

[4] Cheang, H. S., Pell, M. D. (2011). Recognizing sarcasm without language: A cross-linguistic study of English and Cantonese. *Pragmatics and Cognition, 19*(2), 203-223.

[5] Fung, R. S.-Y. (2000). Final Particles in Standard Cantonese: Semantic Extension and Pragmatic Inference. PhD dissertation. Ohio State University.

[6] Mauchand, M., Vergis, N., Pell, M. (2018). Ironic tones of voices. *9th International Conference on Speech Prosody 2018*, (June), 443–447.

[7] Myers, R. A. (1978). *Irony in Conversation*. Ann Arbor: University of Microfilms International.

[8] Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research, 29*(5), 483–495.

The Effects of Sentential Context on Compensation for English Assimilated Speech by L2 Listeners

Eunkyung Sung[1], Youngeun Kim[2] & Sooyeon Lee[2]

[1]*Cyber Hankuk University of Foreign Studies (Korea)*, [2]*Hankuk University of Foreign Studies (Korea)*
eks@cufs.ac.kr, youngeun.kim01@gmail.com, salee0916@naver.com

Speech is inherently variable, and phonological processes such as assimilation, deletion, and substitution render many relevant distinctions obscure in connected speech. Perceptual ambiguity caused by phonological processes must be compensated for by the perceptual system. Previous studies have shown that surface variations in speech were perceptually tolerated when modifications occurred in phonologically viable contexts, and that the phonological compensation is moderated by language experience with native assimilation rules [1, 2, 4, 5, 6]. Previous research on L2 acquisition of phonological assimilation has presented the significant effect of L2 proficiency level in perception [2, 3].

The aims of this study were to examine the effect of sentential context on compensation for English place assimilation and to compare compensation patterns between native English listeners and native Korean listeners with a high proficiency level in English. To these ends, we conducted two experiments (i.e. discrimination and identification tasks) for English and Korean listeners. In the discrimination experiment, two types of stimuli (i.e. compound words and sentences) were presented involving English coronal place assimilation. In the discrimination task involving compound words, a target token was embedded in one of three phonological contexts (i.e. no change, unviable change, and viable change), and in the task involving sentences a target token was presented in the viable change context. In compound words, the participants listened to 90 test items and 20 filler items, and were asked to indicate whether a priming target token and the first syllable of a compound word were the same or different (e.g. beat, bea[t] sound/ bea[$^{t}{}_{p}$] note/ bea[$^{t}{}_{p}$] box). In sentences, the participants listened to 60 test items and 20 filler items, and were asked to discriminate between a priming target token and the first syllable of a compound word presented in a sentence (e.g. bea[t], The bea[$^{t}{}_{p}$] box battle finished). In the identification experiment, the participants listened to 80 test items and 40 fillers, and were requested to identify target tokens in sentences. A target token including a coronal or non-coronal consonant in coda was embedded in a sentence that provided the semantically neutral context for the two types of target tokens (e.g. Jane found the ca[t]/ca[p] by the front door).

The results of the discrimination experiment showed a strong effect of the sentential context and listener group. When comparing the three contexts in compound words, the two listener groups demonstrated the highest mean detection rates in the no change context and the lowest rates in the unviable change context. When the detection rates in the viable change context in words were compared with those in sentences, both listener groups demonstrated higher detection rates for sentences than for words. Thus, the two listener groups showed a similar compensation pattern for English coronal place assimilation. However, statistical differences in the factor of listener group were also revealed. The Korean listeners' higher detection rates in the unviable context in words than the English listeners' detection rates indicated that Korean listeners had difficulty perceiving phonetic differences between the original form and the assimilated form of coronal codas in the context in which assimilation was not conditioned. In addition, the boxplots displayed more variations in detection rates for Korean listeners than for English listeners. The results of the identification experiment showed a significant effect of coda type of target tokens (i.e. coronal vs. non-coronal coda consonants). Both listener groups revealed a strong sensitivity to phonetic differences between the assimilated form of coronal consonants and original bilabial or velar consonants in coda. In addition, the results of the identification experiment did not find any effect of listener group. The general results from the two experiments indicated that the Korean listeners with a high level of proficiency in English showed similar perceptual compensation patterns to native English listeners' patterns. A strong effect of sentential context was found in terms of compensation for assimilation for both L1 and L2 listeners.

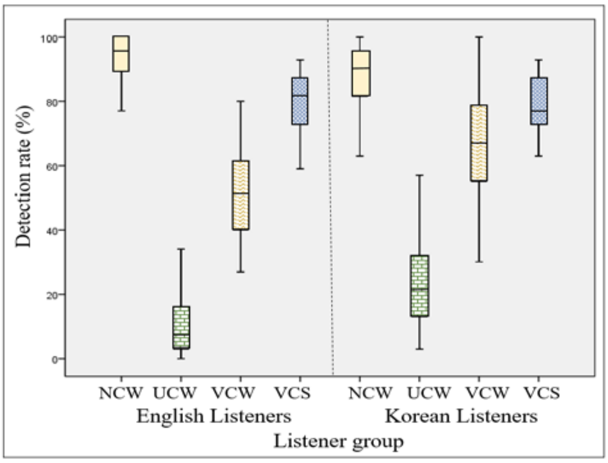


Fig. 1 Mean detection rates in three contexts of words and one sentential context by two listener groups

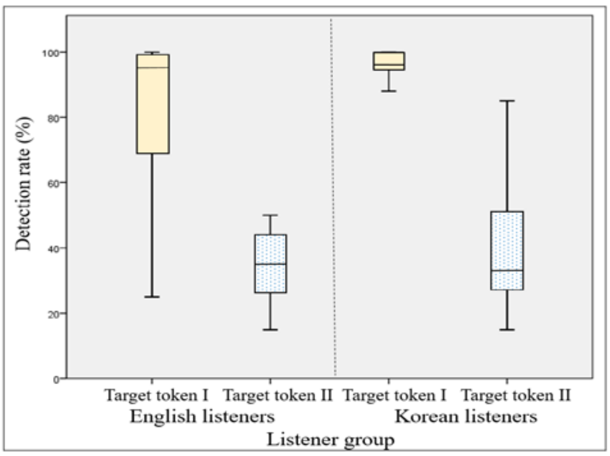


Fig. 2 Mean response rates in the identification task by two listener groups

Note. NCW= no change_word, UCW=unviable change_word, VCW=viable change_word, VCS=viable change_sentence

Table 1 Statistical results in the discrimination task

Factor type	Post-hoc comparisons		
Phonological context (UCW vs. VCW vs. VCS) $p<.0001$	English listener	UCW vs. VCW	$p<.0001$
		VCW vs. VCS	$p<.0001$
Listener group (English vs. Korean) $p<.05$	Korean listener	UCW vs. VCW	$p<.0001$
		VCW vs. VCS	$p<.05$

Note. NCW= no change_word, UCW=unviable change_word, VCW=viable change_word, VCS=viable change_sentence

Table 2 Statistical results in the identification task

Factor type
The effect of coda type of target tokens (coronal vs. non-coronal) p<.0001
The effect of listener group (English vs. Korean) p>.05

References

[1] Darcy, I. (2002). Online processing of phonological variation in speech comprehension: The case of assimilation. In S. Hawkins and N. Nguyen, eds., *Temporal integration in the perception of speech*. ISCA.

[2] Darcy, I., Peperkamp, S. and Dupoux, E. (2007). Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. *Laboratory Phonology* 9(9), 411-442.

[3] Darcy, I., Ramus, F., Christophe, A., Kinzler, K. and Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Kugler, C. Fery and R. van de Vijer, eds., *Variation and gradience in phonetics and phonology*. Berlin: Mouton de Gruyter.

[4] Gaskell, M. G. and Marslen-Wilson W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22, 144-158.

[5] Gaskell, M. G. and Marslen-Wilson W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology* 45(2), 220-266.

[6] Mitterer, H. and Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics* 65(6), 956-969.

Effects of L1 and L2 experience on the perception of Korean three-way stop contrast: evidence from Chinese and Korean listeners

Shengyi Wu[1], Ara Cho[2]

[1]*Shaoxing University (China),* [2]*Fudan University (China)*
wushengyi88@gmail.com, car10@daum.net

Cross-linguistic studies on the perception of Korean stops have demonstrated that listeners'perceptual strategies for the three-way contrast are modulated by their language experiences. For example, Seoul Korean listeners used both VOT and F0 as perceptual cues, while Kyungsang Korean listeners rely less on the F0 cue (e.g. Lee & Jongman, 2012; Lee et al., 2013). Likewise, English listeners used VOT as a reliable cue for Korean stops, but they were not sensitive to F0 changes (e.g. Kim & Lotto, 2002).

However, few studies have investigated the cue weighting strategies on Korean stops by tone language (e.g. Mandarin) listeners. Based on a perceptual assimilation task, Holliday (2014) demonstrated naïve Mandarin listeners were able to use VOT as a reliable cue for the Korean obstruents, but fail to use F0 effectively. Holliday (2015) further indicated that, with more L2 experience, some Mandarin learners were able to use F0 as a cue more effectively. These results suggested the perception of a non-native contrast was affected by both L1 and L2 experience.

In this study, we investigated how Korean three-way stops are perceived by Mandarin listeners with varied L2 experiences (elementary learner vs. advanced learner) using stimuli with manipulated F0 and VOT. Both VOT and F0 are reliable cues in Mandarin, but they play different roles in maintaining phonological contrasts (i.e. F0 for lexical tones and VOT for stops), thus it would be interesting to discuss Mandarin listeners' performance in perceiving Korean stops.

45 Chinese subjects (native Mandarin speakers) were recruited. 22 of them are elementary learners of Korean (about 8 months of leaning experience), 23 of them are advanced learners (4-5 years of experience). 22 Seoul Korean subjects were recruited as a control group. A speech continuum was created by manipulating F0 and VOT of the base token "phada". The VOT duration ranged from 86 ms to 6 ms in 10 ms steps. The F0 ranged from 260 Hz to 180 Hz in the step of 10 Hz. A total of 81 stimuli were created (9 levels of VOT * 9 levels of F0). In total 243 stimuli (81 stimuli * 3 repetitions) were randomly presented to listeners in Eprime. Subjects were instructed to identify the stimuli from three options (lenis, fortis, aspirated).

As Figure 1 showed, the Chinese and Korean groups used different perceptual strategies for the three-way stop contrast. Korean listeners relied on F0 to perceive lenis stops, and primarily on VOT for fortis responses, and F0 and VOT for aspirated responses, which are in line with previous studies. While for Chinese listeners, F0 plays a less important role in the perception of the fortis and aspirated stops. What's more, Chinese listeners relied more on VOT in distinguishing the aspirated from the lenis than Korean listeners did. For the long VOT stimuli (e.g. above 46 ms), Chinese listeners preferred aspirated responses than lenis responses, but this pattern was less salient for Korean listeners.

Figure 2 indicates that the cue weighting strategies were basically the same across the two groups of Chinese listeners (elementary learner vs. advanced learner). However, the two groups showed inter-level differences in terms of categorical perception of the three-way stop contrast. Comparing to the elementary group, the advanced group showed more salient categorical perception for two contrasts (aspirated vs. fortis, aspirated vs. lenis), indicated by the steeper identification curve across the categorical boundaries.

In this study, we observed both cross-linguistic and inter-level differences for the perception of three-way stop distinction. We argued that the performance of the Chinese group is affected not by L1 phonology, rather by their L1 experience on cue weightings for stops. First, the lexical tone status in Mandarin does not help them in distinguishing stops by F0, because F0 does not serve as a reliable cue for stop laryngeal distinction. Second, the Chinese listeners' dependency on VOT for the two-way stop distinction in L1 has been transferred to the perception of non-native stop contrast.

On the other hand, listeners' L2 experience does not affect the cue weighting for the Korean stops but play a role in the enhancement of non-native contrasts. The Chinese advanced learners demonstrated more salient categorical patterns for the Korean stop contrasts than the elementary learners. Specifically, for the advanced learners, the fortis and aspirated stops were better categorized by VOT dimension, and the lenis and aspirated stops by F0 dimension.

Taken together, this study demonstrated the different roles of L1 and L2 experience in modulating the perception of non-native contrasts.

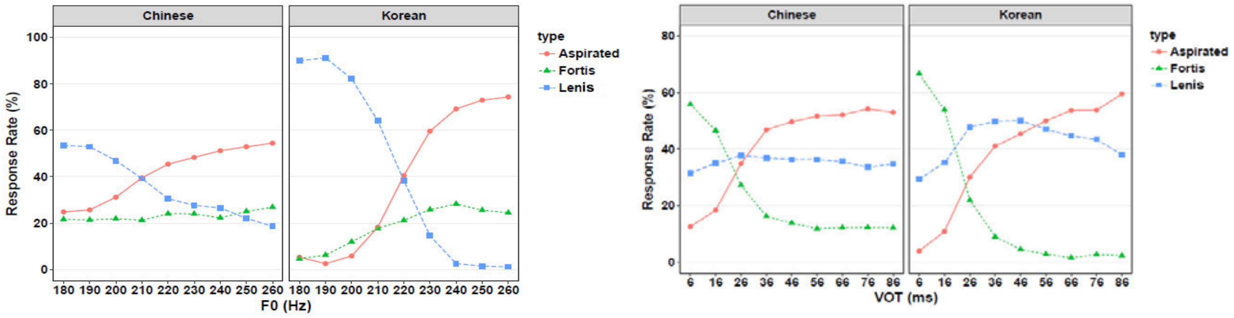


Fig. 1 Identification rate for the three-way distinction as a function of F0 (left) and VOT (right) for the Chinese and Korean group

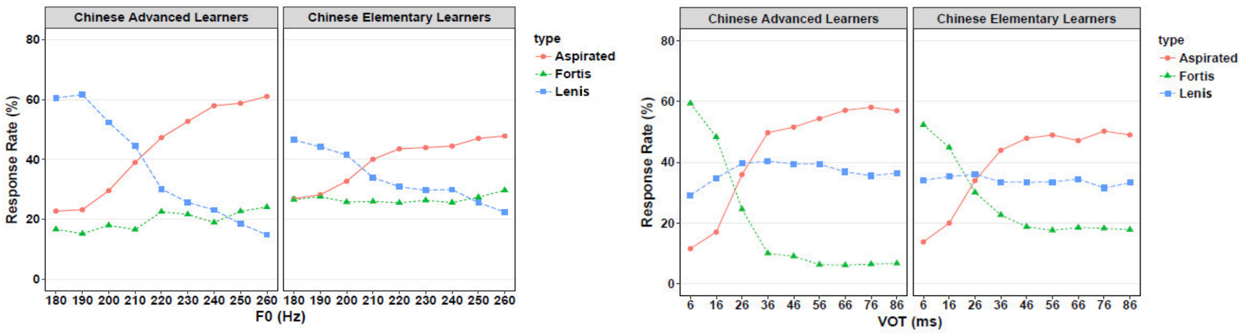


Fig. 2 Identification rate for the three-way distinction as a function of F0 (left) and VOT (right) for the Chinese advanced and elementary learners

References

[1] 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*(2), 117-132.

[2] Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. Journal of the International Phonetic Association, 42(2), 145-169.

[3] Holliday, J. J. (2014). The perceptual assimilation of Korean obstruents by native Mandarin listeners. *The Journal of the Acoustical Society of America*, *135*(3), 1585-1595.

[4] Holliday, J. J. (2015). A longitudinal study of the second language acquisition of a three-way stop contrast. Journal of Phonetics, 50, 1-14.

[5] Kim, M. R. C., & Lotto, A. J. (2002). An investigation of acoustic characteristics of Korean stops produced by non-heritage learners. The Korean Language in America, 7, 177-187.

Prosodic Realization of Multiple Accusative Construction in Korean

Dong Jin Kim, Okgi Kim & Hanyong Park

University of Wisconsin-Milwaukee (USA)
kim252@uwm.edu, okgikim@uwm.edu, park27@uwm.edu

Multiple Accusative Construction in Korean (henceforth MAC) involves two different accusative Noun Phrases (NPs) despite the fact that the verb (e.g., *manna* 'meet') requires only one direct object, as illustrated in (1) (cf. [1]).

(1a) a specificational MAC

Mimi-ka	UWM.haksayng-ul	John-ul	manna-ss-e
Mimi-NOM	UWM.student-ACC	John-ACC	meet-PAST-DECL

'Mimi met a UWM student: John.'

(1b) a predicational MAC

Mimi-ka	John-ul	UWM.haksayng-ul	manna-ss-e
Mimi-NOM	John-ACC	UWM.student-ACC	meet-PAST-DECL

'Mimi met John who is a UWM student.'

In this paper, we would like to suggest that despite no overt copular verb, the semantic relationship between the two direct objects in (1a) and (1b) is interpreted as specificational and predicational relations, respectively. Based on the information structure of canonical specificational and predicational clauses suggested by [2], we hypothesize that only the second direct object in specificational MAC receives a focus, but either the first or second direct object can be freely focused in predicational MAC, as shown in (2a) and (2b).

(2a) Mimi-ka [$^{?}_{Focus}$ UWM.haksayng-ul] [$_{Focus}$ John-ul] manna-ss-e

(2b) Mimi-ka [$_{Focus}$ John-ul] [$_{Focus}$ UWM.haksayng-ul] manna-ss-e

If our hypothesis is correct, we expect that only the second NP (NP2) in specificational MAC exhibits focus effects such as longer word duration, longer voice onset time (VOT) and higher F0 [3], [4], [5] than the first NP (NP1). In predicational MAC, such focus effects would be shown in either NP1 or NP2. In order to see whether this expectation was borne out, we conducted a production study in which six native speakers of Seoul Korean in their twenties participated. The target items (Table 1) were included in carrier sentences and the speakers were asked to read the prompted sentences. Considering a phrasal tone of accentual phrase in Seoul Korean, THLH [5], the initial segments of the target items were controlled as voiceless lenis stops /t, k/ making their phrasal-initial tones to L. The results showed that specificational NP2 exhibited longer word duration and VOT with higher F0 than predicational NP2 (Fig.1). These results indicate that focus effects were shown on specificational NP2 as we anticipated. This prosodic realization thus suggests that MACs can be classified into two types: specification and predication.

In order to see whether such prosodic information facilitates listeners' interpretation, five additional native listeners of Seoul Korean in their twenties participated in a perception experiment. We made three different types of stimuli: silence reading, contrastive-focused NP1, and contrastive-focused NP2. After listening to one of the stimuli, the listeners responded "Yes" or "No" to questions asking whether the sentence they heard has a specificational/predicational relation. The results revealed that when specificational MACs with the silence condition were presented, 69% of the responses were "Yes" to questions about the specificational relation. However, when specificational MACs with the contrastive-focused NP2 were presented, all the listeners chose "Yes" (100%) to the same question. Meanwhile, when predicational MACs with the silence condition were given, 79% of the responses were "Yes" to questions about the predicational relation, and the same percentage of "Yes" (79%) was also obtained when predicational stimuli with the contrastive-focused NP2 were presented. These perception results demonstrate that there are facilitation effects on the interpretation of MACs when the prosodic realizations correspond with the information structure given in (2a) and (2b).

Taken together, the results of the current study indicate that phonetic details that reflect the information structure of MACs are utilized in order for speakers to encode and decode linguistic information (e.g., specification and predication). Such phonetic information gives rise to the facilitation effects on listeners' interpretation of the encoded linguistic information [6].

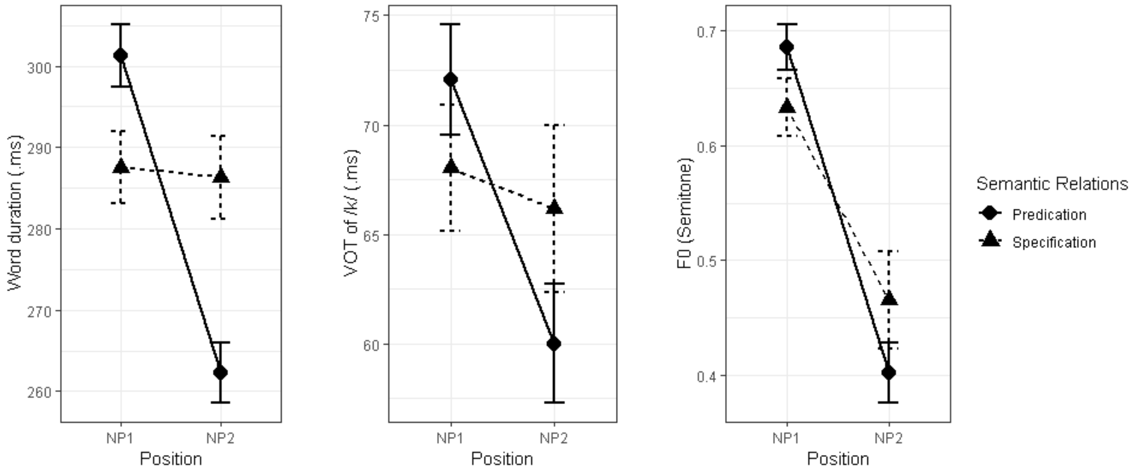


Fig.1 Comparison of word duration (ms), VOT of /k/ (ms), and F0 (semitone) regarding the position of NPs. Error bars represent standard errors.

Table 1 Target items in specificational and predicational MAC conditions.

Specification		Predication	
NP1	NP2	NP1	NP2
kogi 'meat/fish'	koni 'swan'	koni 'swan'	kogi 'meat/fish'
kogi 'meat/fish'	$tots^hi$ 'lumpfish'	$tots^hi$ 'lumpfish'	kogi 'meat/fish'
kogi 'meat/fish'	tomi 'snapper'	tomi 'snapper'	kogi 'meat/fish'

References

[1] Chae, H. R. & Kim, I. (2008). A clausal predicate analysis of Korean Multiple nominative constructions. *Ene [Language]*, *33*(4), 869-990.

[2] Mikkelsen, L. (2005). *Copular clauses : Specification, predication and equation*. Philadelphia, PA: John Benjamin Publishing Company.

[3] Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, *39*(3), 344-361.

[4] Choi, J., Lee, J., Kim, S., & Cho, T. (2018). Prosodically-conditioned phonetic cue in production of Korean aspirated vs. lenis stops. In T. Cho, S. Kim, J. Choi, J. J Kim, S. Y. Kim., & K. J. Lee (eds.), *Proceedings of Hanyang International Symposium on Phonetic and Cognitive Sciences of Language (HISPhonCog) 2018, Volume 1* (pp. 121-122). Seoul, Korea: HIPCS (Hanyang Institute for Phonetics and Cognitive Sceiences of Language).

[5] Jun, S. A. (2006). Intonational phonology of Seoul Korean revisited. In T. Vance & K. Jones (eds.), *Japanese/Korean Linguistics, Volume 14* (pp. 15-26). Stanford, CA: CSLI Publications.

[6] Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass*, *10*(3), 120-141.

Effects of prosodic structure on voice quality associated with Korean three-way stop contrast

Jiyoung Jang[1], Sahyang Kim[2] & Taehong Cho[3]

[1]*University of California Santa Barbara (USA),* [2]*Hongik Univ. (Korea),* [3]*Hanyang Institute for Phonetics and Cognitive Sciences of Language, Hanyang University (Korea)*

jiyoung@ucsb.edu, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Korean has a three-way voiceless stop contrast between aspirated, fortis, and lenis stops [1]. While VOT was regarded as the primary cue which distinguished the three stop categories, recent studies have suggested that Korean is undergoing a tonogenetic sound change, showing that the VOT distinction between the aspirated and lenis stop is merging and that the contrast is now distinguished by F0 (higher F0 for aspirated and lower F0 for the lenis) in younger Koreans' speech [2]. Traditionally, however, the three-way stop contrast in Korean has also been known to be phonetically characterized by the difference in voice quality of the following vowel, such that the breathiness of the vowel is greatest after the lenis, intermediate after the aspirated, and least after the fortis [1]. The purpose of this study is to examine the voice quality differences associated with the three-way contrastive stops, and to explore to what extent these differences contribute to the three-way stop contrast and how the effect may be conditioned by two prosodic-structural factors.

Regarding the effect of prosodic strengthening on the voice quality of the three stops, it is possible to hypothesize that the strengthening would be phonetically manifested by increased glottalization across the board as sounds are known to be glottalized (or become creakier) in prosodic strengthening environments [3,4]. The prosodic strengthening effect may thus increase the laryngeal muscular tension, which in turn would augment the degree of glottalization. If the three-way contrastive stops are produced with an increase in laryngeal muscular tension, the breathiness of the following vowel across the three stop categories would be reduced. However, studies on prosodic strengthening have also indicated that prosodic strengthening is not a mere low-level phonetic effect, but it refers to the phonological system of a given language, often enhancing the phonological contrast [5]. This leads to an alternative hypothesis that prosodic strengthening would increase the breathiness of the vowel only after the lenis stop while the creakiness of the vowel is reinforced after the fortis stop, enhancing the three-way stop contrast. This alternative hypothesis may not be borne out, however, if VOT and F0 are the only primary phonetic cues of phonological contrast in Korean, as has been argued in conjunction of a recent development of the tonogenetic sound change among young speakers [2]. To explore these possibilities, this study investigates whether the three-way phonological contrast of word-initial stops manifest itself in the voice quality of the following vowel in Korean in prosodic strengthening contexts.

Twelve native Seoul Korean speakers participated. Example sentences are given in Table 1. Two measurements of breathiness, H1*-H2* and H1*-A1*, were taken as acoustic indexes of the voice quality [4], obtained by VoiceSauce [7] at the 25% and 50% points of the vowel.

Results are summarized in Fig. 1. One of the basic findings of the present study is that the difference in voice quality of the following vowel, as indicated by breathiness measures (H1*-H2* and H1*-A1*), makes a three-way phonetic distinction among the (Seoul) Korean stop series produced by young speakers. The amount of breathiness is largest for the lenis stop, intermediate for the aspirated stop, and smallest for the fortis stop. The three-way laryngeal distinction is consistent with what was reported 16 years ago [1], indicating that the voice quality difference has continued to underlie the three-way stop contrast. Another important finding of the study is that the three-way distinction in voice quality is further conditioned by prosodic strengthening factors: focus-induced prominence and boundary. The results showed that the three-way stop contrast is indeed enhanced under focus, with the stops being substantially dispersed along the breathy-creaky phonetic continuum. The voice quality difference as a function of stop categories has also been found to be influenced by prosodic boundary (IP-initial vs. IP-medial). The dispersion effect was observed in both H1*-H2* and H1*-A1*. The converging enhancement pattern under prosodic strengthening is that vowels become more breathy (less creaky) after the lenis stop, but less breathy (more creaky) after the fortis stop, contributing to the enhancement of the phonological contrast. Interestingly, the voice quality associated with the aspirated stop falls somewhere in between, which may be understood as an effort to retain the contrast by maintaining its intermediate position.

The results imply that variation in the voice quality difference as a function of prosodic strengthening is not a mere low-level phonetic effect that would otherwise have applied to all three stops in a collective way, but is an outcome of the phonetic-prosody interface in reference to the phonological contrast in the language. The results also suggest that understanding the nature of laryngeal (voicing) contrast that occurs in Korean as well as in other languages requires multidimensional approaches to explore the phonetic realization of both the primary and other non-primary phonetic features [e.g., 8]. It remains to be seen to what extent the voice quality difference is exploited by the listeners and how the voice quality cues may interact with F0 and VOT [cf. 8].

Table 1. Example sentences with the target word *pak* in different prosodic context. Focused words are in bold.

IP-initial	**Foc**	**[ani]. IP [pak satʃʲin twi]. IP [twɛssʌ]? "No. To the right of the picture of pak. Got it?"** **(as an answer to "This time, do I place the word (card) to the right of the picture of mak?")**
	Unfoc	[ani]. IP [paksatʃʲin **twi**]. IP [twɛssʌ]? "No. To the **right** to the picture of pak. Got it?" (as an answer to "This time, do I place the word (card) to the **left** of the picture of pak?")
IP-medial	Foc	[ani]. IP [a*pa **pak**satʃʲin twi]. IP [twɛssʌ]? "No. To the right of dad's picture of **pak**. Got it?" (as an answer to "This time, do I place the word (card) to the right of dad's picture of **mak**?")
	Unfoc	[ani]. IP [a*pa paksatʃʲin **twi**]. IP [twɛssʌ]? "No. To the **right** of dad's picture of pak. Got it?" (as an answer to "This time, do I place the word (card) to the **left** of dad's picture of pak?")

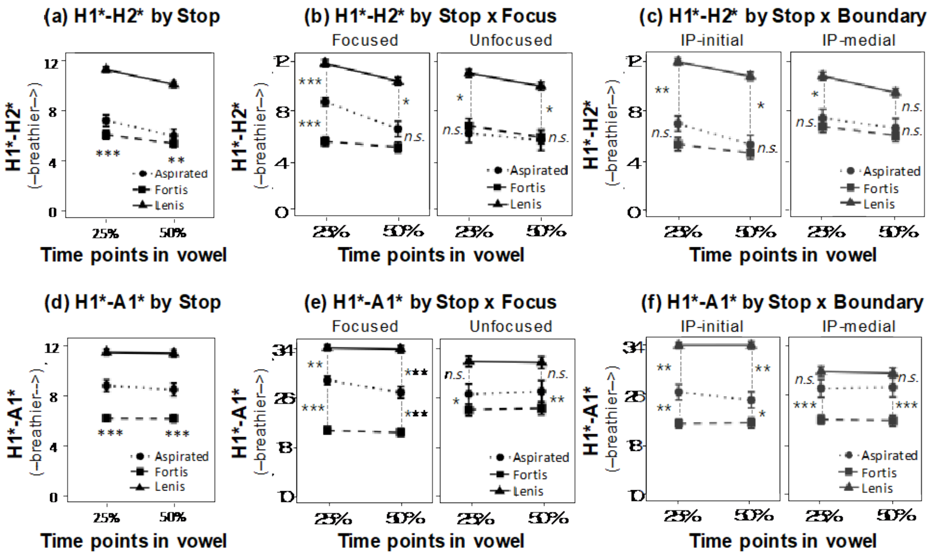


Fig.1 Effects of Stop and its interaction with prosodic factors (focus and boundary) on H1*-H2* (a-c) and on H1*-A1* (d-f). '*', '**', '***', 'tr.', 'n.s.' refers to p<.05, p<.01, p<.001, .05<p<.06, p>.09 respectively.

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] 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.

[3] Cho, T., Kim, D., & Kim, S. 2018. Prosodic strengthening in reference to the lexical pitch accent system in South Kyungsang Korean. *The Linguistic Review*.

[4] Garellek, M. 2014. Voice quality strengthening and glottalization. *Journal of Phonetics*, 45, 106-113.

[5] Cho, T., & McQueen, J. M. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33(2), 121-157.

[7] Shue, Yen-Liang 2010. *The voice source in speech production: Data, analysis and models*. Los Angeles: University of California. http://www.phonetics.ucla.edu/voiceproject/ Publications/shue_dissertation.pdf

[8] Kong, E. J., Beckman, M. E., & Edwards, J. 2011. Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics*, 39(2), 196-211.

Prosodic and syntactic boundaries in spontaneous English and Finnish speech

Anna Dannenberg[1] & Stefan Werner[2]

[1]*University of Helsinki (Finland)*, [2]*University of Eastern Finland (Finland)*
anna.dannenberg@helsinki.fi, stefan.werner@uef.fi

In this paper, we examine prosodic and syntactic boundaries in English and Finnish spontaneous speech corpora. Comparing prosodic and syntactic units can help to understand the interaction between prosody and syntax in both speech production and perception.

For analysing the prosodic structure of spoken language, especially of spontaneous speech, effective automatic applications have unfortunately been rare. A Continuous Wavelet Transform (CWT) based method [1] applies the weighted sum of f0, energy and segmental durations to represent prosodic signals in a two-dimensional time-scale plane akin to spectrograms. The results can be further enhanced with lines of maximum amplitude to produce a visual representation of the prosodic hierarchies of speech.

In this study we have used a CWT based tool to detect prosodic boundaries in spontaneous speech. The results have been compared with grammatical analysis of the same data to examine the relation of prosodic and syntactic units in spoken language.

Our English dataset is extracted from the Buckeye corpus (Ohio State University) [2]. Our corpus was comprised of a total of 10076 words of spontaneous English speech, consisting of samples from informal interviews of five female speakers of American English.

For comparison, we used data from Finnish dialect interviews conducted by the Institute for the Languages of Finland [3]. Our sample contains 2217 words of spontaneous Finnish speech from two speakers, one female and one male, both native speakers of different dialects of Finnish.

In preprocessing, comments of the interviewer were deleted, and the speech was divided into turns based on speaker changes and obvious pauses.

The data was prosodically segmented using a wavelet based tool. The weighted sum of normalized f0, energy and segmental durations was used as an input signal for the CWT. Prosodic boundaries were determined by tracking minima across scales in the resulting scalograms, lines of minimum amplitude. For the English data, the whole segmentation process was fully automatic and unsupervised. For the Finnish data, prosodic boundaries were tagged manually based on how the words were grouped into branches in the prosodic tree structure produced by the tool.

The syntactic segmentation of the English data was performed by 20 informants, all of them native Finnish speakers having good to excellent skills in English language and grammar. They were asked, demonstrated by some examples, to tag every syntactic sentence and clause boundary in the text, and in ambiguous cases, make their own interpretations according to the context. The informants had no access to the spoken data but only the transcription, so they had to perform the task without help of any acoustic cues. Each sample was segmented by four informants independently of each other. The Finnish corpus was syntactically segmented according to the same principles, but due to its smaller size, the segmentation was performed by only one skilled native speaker of Finnish.

The wavelet based prosodic segmentation resulted in 1700 prosodic boundaries in the English data and 703 in the Finnish data. The syntactic segmentation resulted in 2147 and 457 syntactic boundaries respectively. The mean length of a prosodic unit was thus 5.9 words in the English data and 3.2 words in the Finnish data, and for syntactic units, 4.7 and 4.9 words respectively. The remarkable difference in the length of prosodic units may be due to the fact that the Finnish speakers were rather aged, therefore speaking more slowly and taking more pauses that were interpreted as prosodic breaks.

Of all the boundaries marked in the English data, 906 were co-occurrences of prosodic and syntactic ones. This is 53.3% of all the prosodic boundaries and 42.2% of the syntactic ones. In the Finnish data, the number of co-occurrences was 313, the percentages being 44.5% and 68.5%.

It is thus clear that prosodic and syntactic boundaries tend to co-occur, since in both languages their co-occurrences are significantly more common than a random distribution. It is interesting, though, to take a look at the exceptions: prosodic boundaries without a syntactic one or vice versa.

For a syntactic boundary without a prosodic one, the most common instance in both English and Finnish data was a new syntactic main clause, either independent or coordinate, beginning in the middle of a prosodic unit. Almost 60% of solitary syntactic boundaries in the English data and a little less than 30% in the Finnish data were such occasions. They included a lot of half-grammaticalized syntactic elements like *I think* or *you know* produced as a part of a larger structure without need to separate them prosodically. A little less common were subordinate clauses beginning in the middle of a prosodic unit, with the percentages of ca. 30% and 20%.

(1) a solitary syntactic boundary after *y(ou) know* (**P** = prosodic boundary, **S** = syntactic boundary)
PS but it's n **P** ah **PS** yknow **S** it's not a backwater either **PS**

The most significant difference between English and Finnish data was related to cases where a solitary syntactic boundary preceded a conjunction, either a coordinate or a subordinate one, and a prosodic boundary followed the same conjunction. These cases corresponded to 42% of all the solitary syntactic boundaries in the Finnish data but only 5% in the English data; in English, it was much more common to have a prosodic break both before and after a conjunction. In spontaneous speech, a conjunction at the end of an utterance or a speaker turn is a common phenomenon since it often shows an intention to continue despite the prosodic break.

Prosodic boundaries in both English and Finnish were only seldom situated between a subject and a predicate, or between an auxiliary and a main verb. Only ca. 15% of solitary prosodic boundaries in both English and Finnish data were situated in either one of these locations. On the contrary, more than 50% of these boundaries in Finnish and almost 50% in English were situated either between a verb and its arguments other than subject, or between more remote constituents.

(2) a solitary prosodic boundary between a verb and its object (**P** = prosodic boundary, **S** = syntactic boundary)
PS oh they well they knew **P** all d all different things **PS**

It was also not uncommon to have a prosodic break in the middle of a noun phrase, especially in cases where an NP consisted of several words and the situation included some kind of hesitation.

Our results thus show that in both spontaneous English and Finnish speech, prosodic and syntactic boundaries typically tend to to co-occur. Even where they do not, prosodic boundaries only seldom break nuclear elements of a syntactic clause; rather they are situated on the peripheral areas of a clause. The fact that prosodic boundaries do not tend to break fixed syntactic elements corroborates the assumption that prosody and syntax serve a common purpose in structuring spoken language.

References:

[1] Suni, A., Šimko, J., Aalto, D., & Vainio, M. (2017). Hierarchical representation and estimation of prosody using continuous wavelet transform. *Computer Speech & Language*, 45, 123-136.

[2] Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech*. 2nd release. Columbus, OH: Department of Psychology, Ohio State University (Distributor), www.buckeyecorpus.osu.edu.

[3] *Samples of spoken Finnish: an online speech corpus* (2014). Version 1.0. Helsinki: Institute for the Languages of Finland. http://urn.fi/urn:nbn:fi:lb-1001100134.

Interaction between lexical tones and stress is affected by individual variation in language attitudes and L2 experience

Wai Ling Law[1], Olga Dmitrieva[2] & Alexander L. Francis[3]

[1]*The Chinese University of Hong Kong, Shenzhen (China),* [23]*Purdue University (USA)*
lawwailing@cuhk.edu.cn, odimitrie@purdue.edu, francisa@purdue.edu

The suprasegmental properties of bilingual speakers' first language (L1) and second language (L2) may influence each other, particularly in the direction from L1 to L2 [1]. A particularly interesting case is the interaction between lexical stress and lexical tone in learners of stress languages (e.g. English) whose L1 is tonal (e.g. Cantonese). Although both stress and tone use fundamental frequency (f0) as an acoustic correlate, the information value of f0 in cueing stress and tone is different [2]. Identification of tones is almost solely based on f0 height and the direction of f0 change (e.g. for Cantonese see [3]), while for stress f0 is but one of several correlates (e.g. for English see [4]) and vowel quality often provides a more robust cue to stress [5].

Previous studies have shown that speakers of tone languages recruit their L1 tonal systems for producing L2 stress contrasts. For example, Mandarin speakers produce significantly higher f0 in English stressed syllables compared to native English speakers [1]. Language experience factors such as L2 proficiency and amount of L2 use are likely to affect the probability and magnitude of such transfers. Greater L2 experience helps bilingual speakers better discern phonetically similar L1 and L2 segments and thus produce and perceive them more accurately [6, 7]. Therefore, we may expect similar effects for suprasegmental properties. Additionally, language attitudes may play a role in bilingual speakers' production of f0 in their L2 [8, 9]. However, little is known about how language attitudes might affect the interaction between L1 and L2 suprasegmental systems in bilingual speakers. Therefore, the present study was designed to investigate how Cantonese-English bilingual speakers' individual variation in language attitudes, L2 proficiency and L2 use affect their production of f0 in the service of L1 lexical tone and L2 stress.

Twenty Cantonese-English bilingual speakers living in Hong Kong participated in the study. Their attitudes toward both languages, their L2 proficiency and their amount of L2 use were quantified using a detailed questionnaire combining the Bilingual Language Profile [10] and the Language Experience and Proficiency Questionnaire [11]. In two reading tasks, participants produced monosyllabic and disyllabic cross-language near homophones under conditions emphasizing English or Cantonese language mode on separate days. Near homophones were used to control for factors that could affect suprasegmental properties such as word length, segmental composition and syllable structure. Examples are 咳 "cough" [kɐt55] and *cut* [kʌt], and 碩士 "Master's" [sɛk22si22] and *sexy* [ˈsɛksi]. The tonal properties of the tokens were acoustically quantified in terms of f0 range and slope and compared to the participants' language attitude scores, L2 proficiency and amount of L2 use.

Results indicate that the acoustic properties of Cantonese tones and English stress produced by the participants do vary according to individual differences in language attitudes and L2 proficiency and L2 use. Specifically, speakers with more positive attitudes toward Cantonese produced a higher f0 range in Cantonese compared to participants with less positive attitudes. English f0 range, however, was not affected by Cantonese attitudes. On the other hand, speakers with higher English proficiency and use produced significantly shallower f0 slopes in English than those with lower English proficiency and use, indicating less influence of the speakers' tonal system on their realization of English stress.

Taken together, these results suggest that higher L2 proficiency and use reduces the influence of the L1 suprasegmental system on that of the L2 in fluent bilinguals. This finding is analogous to the effect of L2 experience observed in production of L2 segments [6, 7]. Speakers with more L2 experience are more successful at realizing a distinction between L1 tones and L2 stress. Meanwhile, speakers with more positive attitudes toward their L1 implement a bigger acoustic difference between their two languages by rendering their Cantonese f0 more tone-like. These findings suggest that suprasegmental properties of bilinguals' L1 and L2 are flexible and subject to influence of language experience, proficiency, and attitudes.

References

[1] Zhang, Y., Nissen, S. L. & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of Acoustical Society of America*, *123*(6), 4498-4513.

[2] 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.

[3] Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, *9*, 20–36.

[4] Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, *3*(1), 1-11.

[5] Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, *38*(2), 260-271.

[6] Aoyama, K. & Flege, J. E. (2011). Effects of L2 experience on perception of English /r/ and /l/ by native Japanese speakers. *Journal of the Phonetic Society of Japan*, *15*(3), 5-13.

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

[8] Ohara, Y. (2001). Finding one's voice in Japanese: A study of the pitch levels of L2 users. In A. Pavlenko, A. Blackledge, I. Piller & M. Teutsch-Dwyer (Eds.), *Multilingualism, second language learning, and gender* (pp. 231-256). Berlin: Mouton de Gruyter.

[9] Dmitrieva, O., Law, W. L., Lin, M., Wang, Y., Conklin, J., & Kentner, A. (2015). Language attitudes and listener-oriented properties in non-native speech. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. http://www.icphs2015.info/pdfs/Papers/ICPHS0953.pdf.

[10] Birdsong, D., Gertken, L.M., and Amengual, M. (2012). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. https://sites.la.utexas.edu/bilingual/.

[11] Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language and Hearing Research*, *50*, 940-967.

The Acquisition of Xiamen Tone Sandhi by Children

Li Xiaolin[1], Peggy Pik Ki Mok[2]

The Chinese University of Hong Kong

1155068663@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

Previous studies on Cantonese and Mandarin showed that tonal acquisition starts early but finishes late [4][5]. However, these 2 languages either does not have a complex tone system or does not have (complex) tone sandhi. It is still largely unknown how children would acquire a complex tone system and complex tone sandhi. The Xiamen dialect is a good test case because of its complex tonal system. The Xiamen dialect belongs to the Southern Min dialectal group. It has 7 citation tones, including 5 free tones (44, 22, 24, 53 and 21) on open syllables or syllables with a nasal coda and 2 checked tones (4, 32) with a stop coda. The Xiamen dialect has very complex tone sandhi patterns. Every tone in the Xiamen dialect has a citation tone and a sandhi tone. The 5 free tones' sandhi patterns or the 5 free tones form a tone sandhi circle as shown in Figure 1a. Tone sandhi patterns for the 2 checked tones depend on the coda as shown in Figure 1b. Some studies show that syntactic structure decides the application of tone sandhi [1].

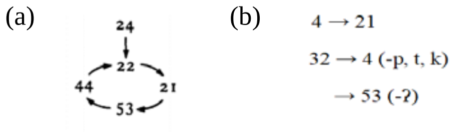


Fig.1 Xiamen tone sandhi patterns of (a) 5 free tones and (b) 2 checked tones

There were very few studies on the acquisition of Taiwanese tone (a Southern Min dialect which is very similar to the Xiamen dialect), and none on the Xiamen dialect. Hsu observed that although sandhi tones were more difficult for children, but they already had very few mistakes in production at 2 [2]. Acoustic analysis also revealed that children had acquired the checked tones well by the age of 3 [3]. However, these studies were all based on natural speech data in corpus without rigorous control on the speech materials. It is unclear if the high accuracy of tone sandhi production was because of children's successful repetition of the sandhi production in their input, or because children already had the abstract knowledge of the tone sandhi rules. It is particularly noteworthy that some previous studies [6][7] demonstrated that tone sandhi in Taiwanese is not productive in adults. Thus, it is important to investigate the productivity of tone sandhi during the course of acquisition by children.

The present study used a picture naming experiment with real words and wug words to investigate children's production of the citation tones and tone sandhi patterns in the Xiamen dialect. Three groups of participants were included: the young children group (Y)(age 5;6-8;1, M=6;8, N=8), the old children group (O)(age 9;2-12;0, M=10;7, N=8) and adults (A)(age 25-57, M=43, N=8). All speakers spoke the Xiamen dialect natively and used it as the usual language in daily life, although they all spoke Mandarin as well. They had very limited exposure to English or other Chinese dialects.

a) "面" (noodle) b) "面包"(bread) c) "面包"(bread) d) "面豆" (noodle and beans)

Figure 2. Examples of the 4 types of stimuli

According to the Xiamen tone sandhi rules, monosyllabic words and the second syllable of disyllabic words are in citation positions and preserve citation tone. The first syllable in disyllabic words is in sandhi position. Four types of stimuli were used: 1) monosyllabic real words (Fig. 2a); 2) disyllabic real words elicited by 1 picture (Fig 2b); 3), disyllabic real words elicited by two pictures (Fig. 2c); 4) wug words using real syllables elicited by 2 pictures (Figure 2d).

2 adult native speakers received training on the Xiamen tonal system beforehand. They listened to the recordings and decided which tone they heard and whether the production was correct according to the tone sandhi patterns. The following analysis was based on their judgement. Fig. 3a shows that the overall accuracy of young children (M=0.889, SD=0.308) was significantly worse than adults (M=0.971, SD=0.098) in citation position. In sandhi position, Fig 3b shows that the young children (M=0.678, SD=0.308) had

significantly lower accuracy than the old children (M=0.840, SD=0.243) and adults (M=0.971, SD=0.098). Old children also had significantly lower accuracy than adults. For monosyllabic words (Fig. 4a) and disyllabic words (Fig. 4b), young children were found to be less accurate than the other 2 groups. For disyllabic words elicited by 2 picture (Fig. 4c), both children groups were significantly more accurate in citation than in sandhi positions while there was no such difference for adults. For wug words, all three groups were significantly more accurate in citation than in sandhi position (Fig. 4d). There was a clear developmental pattern for the accuracy in tones sandhi position in wug words, and that adults had very high (>90%) accuracy. The old children reached adults' accuracy in all conditions except the sandhi position in wug words. Our data clearly demonstrated that tone sandhi in the Xiamen dialect is productive and follows the general development trend in child acquisition, though children as old as 10 years old still have not fully acquired the tone sandhi rules. Differences in experimental designs and naturalness of the materials might explain the discrepant findings in the current study and previous studies on the productivity of tone sandhi.

3a)

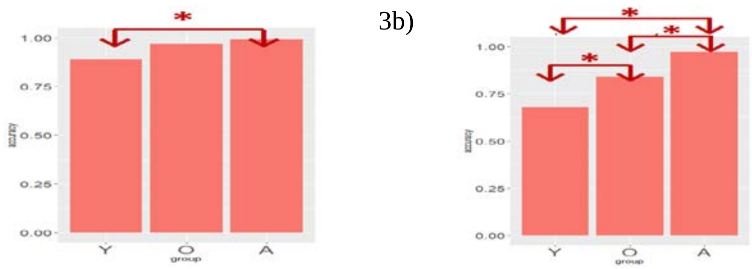


Fig.3 The overall accuracy on (a) citation position and (b) sandhi position

4a)

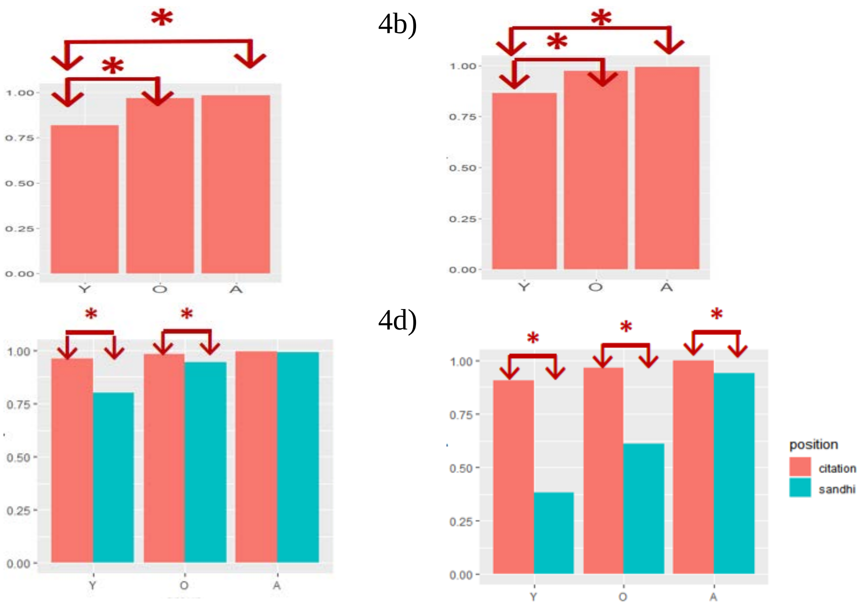


Fig.4 The accuracy of (a) monosyllabic real words, (b) disyllabic real words elicited by 1 picture, (c) disyllabic real words elicited by 2 pictures and (d) wug words elicited by 2 pictures

References

[1] Chen, M. Y. 1987. The syntax of Xiamen tone sandhi. Phonology, 4(1), 109-149.

[2] Hsu, H. C. 1989. *Phonological acquisition of Taiwanese: A longitudinal case study* (Doctoral dissertation, MA Thesis, National Tsing Hua University).

[3] Jane S. Tsay 2001. "Acquisition of Taiwan Tones" 89th Ministry of Science and Technology (general linguistics）research result paper collection page 237-255 (original paper is in Chinese)

[4] Lei, K. Y. M. 2007. *Tone Discrimination in Infants Acquiring a Tonal Language* (Doctoral dissertation, Chinese University of Hong Kong).

[5] Wong, P. 2013. Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, *133*(1), 434-443.

[6] Zhang, J., Lai, Y., & Sailor, C. (2009). Opacity, phonetics, and frequency in Taiwanese tone sandhi. In *Current issues in unity and diversity of languages: Collection of papers selected from the 18th International Congress of Linguists* (pp. 3019-3038).

[7] Zhang, J., Lai, Y., & Sailor, C. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua*, *121*(2), 181-206.

The Perception and Production of Cantonese Syllable-final Stops in Mandarin Speakers: An Analysis of Perceptibility Scale

Zhang Changwei, Chen Si & Leung Man-Tak

Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
changwei.zhang@connect.polyu.hk, sarah.chen@polyu.edu.hk, mantak.leung@polyu.edu.hk

Abstract Based on the Optimality Theory (Prince, A., & Smolensky, P., 1993), constraints related to Perceptibility Scale (Jun, 2004) are proposed as being PRES(pl([-k̚])) >> PRES(pl([-p̚])) >> PRES(pl([-t̚])). This universal ranking refers to the salience of unreleased syllable-final stops in the place of articulation. Targeting on Mandarin speakers' perception and production of Cantonese syllable-final stops ([-p̚], [-t̚] and [-k̚]), current study attempts to testify the applicability of this constraint ranking.

Stimuli and Experimental procedure In Cantonese phonology (Cheung, 2002), only [ɛ], [a] and [ɐ] can be collocated with all target segments ([-p̚], [-t̚] and [-k̚]). Since vowels [ɛp̚], [ɛt̚] and [ɛk̚] can only be used in limited syllables in spoken Cantonese, such tokens were not be considered. Thus, 12 tokens with [a]/[ɐ] as nucleus and [-p̚]/[-t̚]/[-k̚] as coda were selected as stimuli from Chinese Character Database: With Word-formations Phonologically Disambiguated According to the Cantonese Dialect[1]. The production of these 12 tokens by a male native Cantonese speaker were collected with an Audio-Technica AT2020 microphone in a soundproof booth of a phonetics lab. All syllables were digitised at 22050 Hz and 75 dB then normalised for peak intensity (99% of the full scale) and duration (390.46 ms) via Praat (Paul Boersma & David Weenink, 2018).

By adopting Bi-level Diagram (Perceptual Level and Operative Level) from Silverman (1992) and WEE Lian-Hee (2014), 19 Mandarin speakers (9 males and 10 females) were recruited and assessed through AXB Identification Test and Monosyllable Repetition Test. The experiment was conducted via E-prime (Version 2.0) and the order of stimuli was randomised for every subject. In AXB Identification Test, participants were instructed to focus on the coda of experimental syllables and were trained to read IPA transcription of targets. After hearing the recording of target ([CV-p̚], [CV-t̚] or [CV-k̚]), they were required to identify targets from the IPA transcriptions of [CV-p̚]-[CV-t̚], [CV-p̚]-[CV-k̚] or [CV-t̚]-[CV-k̚] displayed on a screen. In Monosyllable Repetition Test, participants were instructed to repeat each of 12 target syllables after hearing the stimuli. Participants' productions were scored independently by five trained phoneticians as target-like production (1) or non-target-like production (0) for further statistic analysis.

Statistic analysis The transformed scores (0 or 1) from raw identification and production data were fit in a generalised linear mixed effects model (Stefan Th. Gries, 2009) using the 'lme4' R package (Bates D, MaÈchler M, Bolker B & Walker S, 2015) in R (R Core Team, 2018). To identify the respective influence from different fixed effects on identification and production accuracy, post-hoc pairwise comparisons among different parameters of each fixed effect were conducted by releveling the reference level in the same model. Moreover, the correlation test was used to measure the relationship between identification accuracy and production accuracy.

Result and Discussion Results suggested that target [-p̚] represented the highest identification accuracy at 88.16% followed by the figures for target [-t̚] (63.82%) and [-k̚] (67.11%). For production results, target [-p̚] obtained the highest accuracy rate at 57.89%, while the figures for target [-t̚] and [-k̚] were 11.84% and 22.37% respectively. Moreover, a positive correlation (ρ=0.469, $p < 0.05$) was identified between identification and production accuracy.

At Perceptual Level, identification results can be described by the constraint ranking as PRES(pl([-p̚])) >> PRES(pl([-t̚]))，PRES(pl([-k̚])). The reasons for this new ranking could be:

---

[1] Chinese Character Database: With Word-formations Phonologically Disambiguated According to the Cantonese Dialect was developed by The Chinese University of Hong Kong and can be retrieved from http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/.

①Phonetically, target [-k˺] is perceptually less salient with proceeding central-low vowel [a] (Kochetov, A., & So, C. K., 2007). Stimuli used in current study took [a] and [ɐ], which are central-low vowel, as the nucleus. This could result in the lower perception rate of target [-k˺] and lead to the ranking of PRES(pl([-p˺])) >> PRES(pl([-k˺])). ② Phonetic variations of [-t˺] and [-k˺] is developing in current Cantonese. According to Chen (1999), Zee (1999) and Carol K. S. To, Sharynne Mcleod & Pamela S. P. Cheung (2015), [-t˺] and [-k˺] tend to merge and become phonetic variations especially when the nucleus is [a]/[ɐ]. This support the result of equal ranking between PRES(pl([-t˺])) and PRES(pl([-k˺])) in the current study.

At Operative Level, production of targets performs the same pattern as perceptual ranking and results in the ranking of PRES(pl([-p˺])) >> PRES(pl([-t˺])) , PRES(pl([-k˺])). The positive correlation between identification and production accuracy indicates that production accuracy increases as the identification figure goes up. This suggests that more perceptually salient segments are better preserved in production.

To conclude, the universal ranking of Perceptibility Scale can reflect language users' performance in unreleased syllable-final stops [-p˺], [-t˺] and [-k˺] only to some extent. Both identification and production rankings in the current study match partially such ranking. However, the result may suggest that vowel context, phonetic variations of targets and the perception-production relationship could affect the re-ranking of constraints related to Perceptibility Scale.

Keywords Cantonese syllable-final stops, perception and production, Perceptibility Scale, Functional Optimality Theory

References

[1] Prince, A., & Smolensky, P. (2004) [1993]. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

[2] Jun, J. 2004. Place assimilation. In Hayes, B., Kirchner, R., & Steriade, D. edited *Phonetically based phonology*, 58-86.

[3] CHEUNG Hung-nin (2002). Cantonese of the 21st Century: Formation of a New Phonological System in the Hong Kong Language. *Journal of Jinan University (Philosophy & Social Science Edition*), 24(2), 25-40.

[4] Paul Boersma & David Weenink. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.37. Retrieved from http://www.praat.org/.

[5] Silverman, D. (1992). Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology, 9*(2), 289-328.

[6] WEE Lian-Hee. (2014). Revisit optimality theory: Perspective of Chinese Phonology. *Contemporary Linguistics, 16*(03), 346-357.

[7] Psychology Software Tools, Inc. [E-prime 2.0] (2016). Retrieved from http://www.pstnent.com.

[8] Gries, S. (2009). *Statistics for linguistics with R : A practical introduction* (Trends in linguistics. Studies and monographs ; 208). Berlin ; New York: Mouton de Gruyter.

[9] Bates D, MaÈchler M, Bolker B, Walker S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

[10] R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

[11] Kochetov, A., & So, C. K. (2007). Place assimilation and phonetic grounding: a cross-linguistic perceptual study. *Phonology*, *24*(3), 397-432.

[12] Chen, N. L.-L. (1999). The velar coda variation in Hong Kong Cantonese. Unpublished Master's Thesis. City University of Hong Kong, Hong Kong.

[13] Zee, E. (1999). Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics*, *27*, 120–167.

[14] Carol K. S. To, Sharynne Mcleod & Pamela S. P. Cheung. 2015. Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical Linguistics & Phonetics*, *29*(5), 333-353.

Syntactic and Length Constraints on the Prosodic Phrasing of Second Language Sentences

Mengtian Chen

University of Iowa (USA)
mengtian-chen@uiowa.edu

Producing an intelligible and fluent utterance is a major goal of second language (L2) learning, but one with which most learners struggle. One of the difficulties is determining the location and the number of breaks in fast-flowing speech [1, 2]. There are usually two ways that non-native phrase boundaries can result in problematic utterances—inappropriate break assignment and too many breaks. Inappropriate break assignment damages the prosodic structure of a sentence so that it sounds awkward. Too many breaks, on the other hand, may induce grammatically correct phrases, but it will reduce the naturalness of articulating a sentence.

The assignment of boundaries is one of the prosodic phrasing issues in sentence production. Prosodic phrasing, or prosodic grouping, is the manner in which phonological units are organized to form larger meaningful units [3]. Traditional approaches to exploring prosodic phrasing in a first language (L1) emphasize the hierarchical representation of prosodic structures in both comprehension and production, including prosodic words, prosodic phrases, and intonational phrases [4, 5]. In L2 processing, however, the features of prosodic unit boundaries are more variable, especially in adult learners' speech, which makes hierarchical classification more difficult [6]. Even highly proficient learners may have difficulty articulating grammatically correct units at refined prosodic structure levels [7]. Moreover, many sub-processes, such as semantic and pragmatic construction, are not automatized, adding more pressure to learners' already burdensome speech production. Thus, learners prefer to produce smaller, more easily managed segments [8], which has a negative impact on L2 oral fluency.

Researchers have investigated the variety and complexity of constraints on prosodic phrasing, particularly the way in which these constraints interact. Syntactic constraints—constituency and dependency relations between words—have been claimed to be the most influential [9, 10], along with the impact of length and focus [11]. The location of a break should first be a syntactic word boundary, otherwise it would be perceived as awkward and inappropriate by native speakers [12]. Yet prosodic structures are much flatter and less branching than syntactic structures because speakers need to balance the length of constituents in their utterances [13]. This becomes more obvious in L2 production, since length is a more explicit and general factor than syntax across languages. Therefore, the interaction of syntactic and length constraints, rather than the establishment of hierarchical prosodic structures, is what learners need to resolve first in their output.

The current study explored the constraints of syntactic constituency, syntactic complexity, and length on the assignment of breaks in the sentence production of Chinese as an L2. Syntactic constituency was operationalized as having adjuncts or not in a sentence. Adjuncts, such as adverbials and prepositional phrases, are not treated as the main elements of an argument structure and are often separated by breaks, but may be bound to the sentence stem as long as they are short [14]. Thus, adding adjuncts to a sentence can influence the syntactic constituency among words, which in turn impacts the prosodic phrasing of the whole sentence. Syntactic complexity was operationalized as simple versus complex sentences, because syntactic complexity in Chinese is realized by sentence pattern [15]. Length was hypothesized to impact learners' phrasing performance by its interaction with syntactic complexity as well as on its own. The reason why the interplay between syntactic constituency and length was not probed was that sentence lengthening inevitably involved adding adjuncts or modifiers, making it difficult to disentangle these two factors.

Two dependent variables were selected to depict learners' phrasing performance. The sequential assembly of sentential components was denoted by phrasing rate—the lower the phrasing rate, the more fluent sentence production, and the more sophisticated learners' phrasing performance should be [16]. The linearization of a prosodic structure was represented by chunk size—the larger a chunk was, the smoother the structuring of prosodic units, and the more sophisticated learners' phrasing performance should be [17].

Phrasing rate and chunk size from sentence recall tasks showed asymmetric effects of syntax and length across groups of different Chinese language levels. Whereas the performance of lower level learners was largely affected by syntactic constituency that indicated word coherence within a small sentential component, the performance of advanced learners was more impacted by the complexity of syntactic structure. The influence of sentence length could be augmented when superimposed onto syntactic constraints. Neither group of learners attained *compressed chunking*, that is, to divide distinct sentences into a similar number of chunks of various lengths, like native Chinese did, but producing a fixed length of chunks instead. This strategy was particularly important in linearizing prosodic structures during the temporal unfolding of spoken language.

References

[1] Cao, J. (2011). Prosodic structure and speech sound variation. *Journal of School of Chinese Language and Culture Nanjing Normal University, 3*, 12–22.

[2] Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmental: Effects of L2 experience on prosody and characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1–30.

[3] Hirst, D. (1993). Peak, boundary and cohesion characteristics of prosodic grouping. *ESCA Workshop on Prosody* (pp. 32–37). September 27–29, Lund, Sweden.

[4] Chu, M., & Qian, Y. (2001). Locating boundaries for prosodic constituents in unrestricted Mandarin texts. *International Journal of Computational Linguistics and Chinese Language Processing: Special Issue on Natural Language Processing Researches in MSRA, 6*(1), 61–82.

[5] Frazier, L., Carlson, K., & Clifton Jr, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*(6), 244–249.

[6] Chen, M., & Wang, J. (2008). A case study of prosodic boundary features of spoken Chinese output. *Chinese Language Learning, 4*, 102–108.

[7] McCauley, S. M., & Christiansen, M. H. (2017). Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science, 9*(3), 637–652.

[8] MacWhinney, B. (2011). The logic of the Unified Model. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 211–227). New York, NY: Routledge.

[9] Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology, 3*(01), 371–405.

[10] Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry, 30*(2), 219–255.

[11] Selkirk, E. O. (2000). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: Theory and experiment: Studies presented to Gösta Bruce* (pp. 231–261). Dordrecht, Netherlands: Springer.

[12] Zhou, M. (2002). The relationships between syntactic structures and phrasing boundaries in Chinese. *Language Teaching and Linguistic Studies, 3*, 35–42.

[13] Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*(4), 411–458.

[14] Bachenko, J., & Fitzpatrick, E. (1990). A computational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics, 16*(3), 155–170.

[15] Gao, L., & Sun, H. (2007). A design of the quantification instrument for classes of teaching Chinese as a foreign language. *Chinese Teaching in the World, 4*, 105–117.

[16] Chen, M. (2012). Chinese oral fluency of CSL learners of American English speakers. *Language Teaching and Linguistic Studies, 2*, 17–24.

[17] Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*(1), 84–119.

A Comparison of Perceptual Compensation for Korean Regressive and Progressive Assimilations by L2 Learners

Eunkyung Sung[1], Youngeun Kim[2] & Sooyeon Lee[3]

[1]*Cyber Hankuk University of Foreign Studies,* [23]*Hankuk University of Foreign Studies*

eks@cufs.ac.kr, youngeun.kim01@gmail.com, salee0916@naver.com

Speech sounds are distorted in continuous speech due to various phonological processes, and listeners need to compensate for such variations in order to activate the correct word representation. This study investigates whether native Korean and Mandarin Chinese listeners learning Korean show similar perceptual compensation patterns involving two Korean assimilation processes (i.e. nasalization and lateralization). This study also examines the effect of assimilation direction (i.e. regressive and progressive assimilations) and the effect of sentential context on perceptual compensation.

To these ends, twenty-one Korean listeners and twenty Mandarin Chinese listeners with a high level of proficiency in Korean participated in four kinds of discrimination tasks (i.e. regressive assimilation in words, regressive assimilation in sentences, progressive assimilation in words, and progressive assimilation in sentences). In each type of discrimination tasks, half of the stimuli involved nasalization, while the other half involved lateralization. The target tokens in the stimuli consisted of monosyllabic nouns or adjectives with a CVC structure. In the discrimination task involving words, the target tokens were embedded into one of three types of phonological contexts (i.e. no change, unviable change, and viable change). The participants listened to 360 target tokens (30 tokens * 3 phonological contexts * 2 assimilation directions * 2 phonological rules) and compound words including the target tokens (e.g. no[n] and no[l].li 'logic' for regressive lateralization; [n]wun and kil.[l]wun 'sense of direction' for progressive lateralization). For the discrimination task involving sentences, the target tokens were embedded in words ending with either their original forms or assimilated forms. The participants listened to 200 target tokens (30 original forms * 2 assimilation directions * 2 phonological rules, and 20 assimilated forms * 2 assimilation directions * 2 phonological rules) and sentences containing the target tokens (e.g. ci[p] and kunun ci[m]mwunul talassta 'He hung the door of his house' for regressive nasalization; [l]yek and kunun nung[n]yeki cohassta 'He had a good ability' for progressive nasalization). For regressive assimilation, the participants were asked to discriminate between a priming target token and the first syllable of a compound word, whereas for progressive assimilation, the participants were requested to discriminate between the priming target token and the second syllable of a compound word. All of the stimuli were prompted by the *PsychoPy* software and randomly presented to each participant.

All the 'same' responses were collected to calculate the detection rates. The results showed that there were clearly different perceptual patterns between the two listener groups in both words and sentences. The Korean listeners revealed a sensitivity to context in both regressive and progressive assimilations involving words and sentences. When the detection rates in the three contexts (i.e. unviable change, viable change in words, viable change in sentences) were compared with each other, regardless of the type or direction of assimilation, the highest detection rates were found in the viable change context in sentences, followed by the viable change context in words, and the lowest detection rates were revealed in the unviable change context. The Korean listeners compensated for both nasalization and lateralization by consistently showing different detection rates in the three contexts. The Korean listeners also demonstrated the effect of sentential context regardless of assimilation directions. On the other hand, the Chinese listeners were not sensitive to context in all cases. Although the Chinese listeners in this study were all advanced learners of Korean, they were not able to perceive the phonetic differences in both word and sentence stimuli. In addition, the L2 advanced learners were not able to use sentential context in perceiving assimilated speech. The effect of L1 syllable structure seemed to strongly affect L2 perception. It was also revealed that although the two listener groups' general perceptual patterns were similar between regressive and progressive assimilations, their sensitivity to phonetic differences of coda consonants was higher in the regressive assimilation than in the progressive assimilation.

Table 1. Construction of stimuli and the statistical results

Phonological context and assimilation direction	Listener group	Context effect: Post-hoc test	Context effect: Regressive assimilation	Context effect: Progressive assimilation
Phonological context (UCW vs. VCW vs. VCS) p<.0001	Korean listener	UCW vs. VCW	p<.05	p<.01
Assimilation direction (Regressive vs. Progressive) p<.01		VCW vs. VCS	p<.01	p<.05
Phonological context (UCW vs. VCW vs. VCS) p>.05	Chinese listener	Not applicable		
Assimilation direction (Regressive assimilation vs. Progressive assimilation) p<.05				

Note. NCW= no change_word, UCW=unviable change_word, VCW=viable change_word, VCS=viable change_sentence

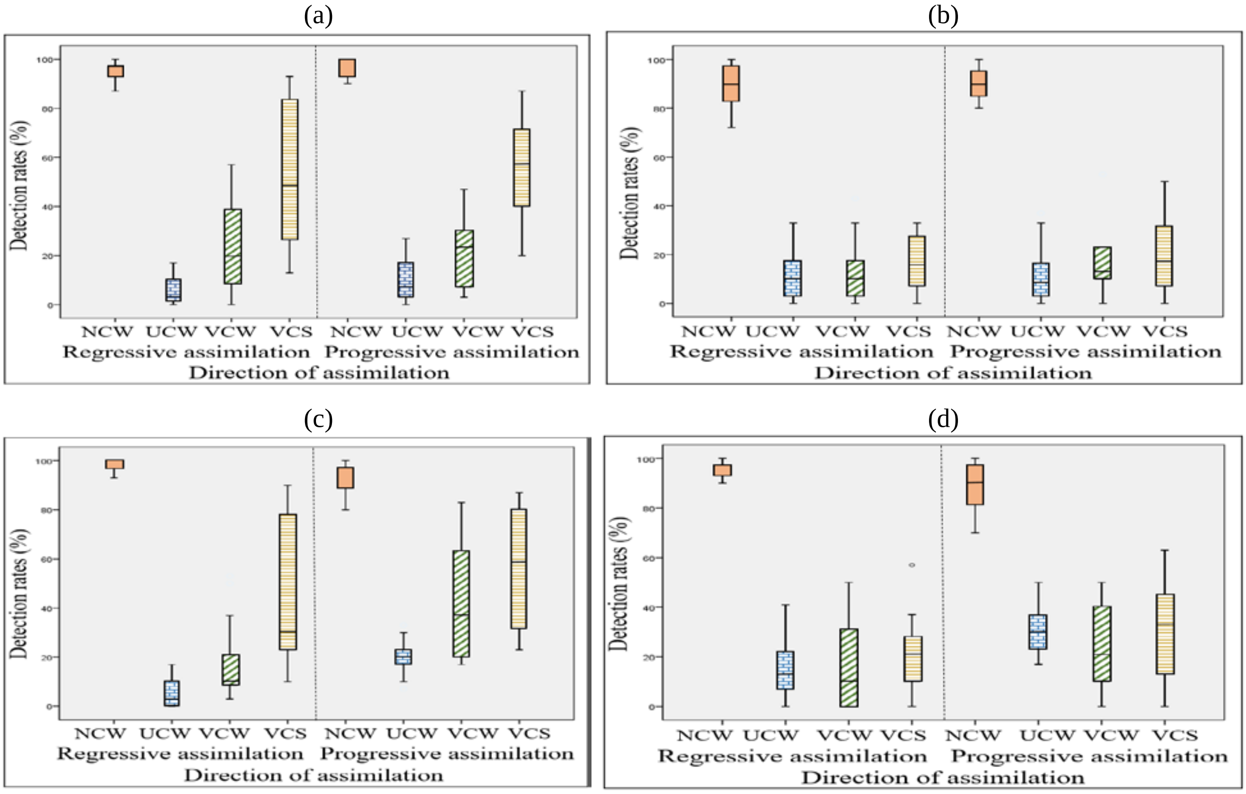


Fig.1 Detection rates in word contexts and one sentential context between regressive and progressive assimilation (a) Korean listeners for nasalization (b) Chinese listeners for nasalization (c) Korean listeners for lateralization (d) Chinese listeners for lateralization Note. NCW=no change_word, UCW=unviable change_word, VCW=viable change_word, VCS=viable change_sentence

References

[1] Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological Knowledge in compensation for native and non-native assimilation. In f. Kugler, C. Fery and R. van de Vijer (eds.). *Variation and Gradience in Phonetics and Phonology*. Berlin: Mouton de Gruyter.

[2] Darcy, I., Paperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. Hualde (eds.). *Papers in Laboratory Phonology 9*. University of Illinois, Champaign, IL.

[3] Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance, 34*(6), 1632-47.

[4] Kim, T. (2000). On the Mechanism of Nasalization and Lateralization in Korean. *Korean Linguistics*, *11*, 227-259.

[5] Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65*(6), 956-969.

The Effects of Syntactic Constraints on Prosodic Chunking of L2 Speech in Sentence Recall

Yanjun Wei[1], Mengtian Chen[2] & Jianqin Wang[1]

[1]Beijing Language and Culture University (China), [2]The University of Iowa (USA)
yanjun.wei@hotmail.com, cmt1923@126.com, wangjianqin@blcu.edu.cn

Human immediate memory capacities are severely limited in general. In contrast to non-native speakers, native speakers are more capable of overcoming some of such limits and demonstrating a stronger ability to talk fluently and continuously (Pawley & Syder, 1983). Second language (L2) learners usually find it difficult to achieve such native-like fluency. Prosodic chunking has the potential to unravel this puzzle. According to Jun (2003), prosodic chunking allows speakers of any given language to organize an utterance into meaningful chunks by combining small constituents (e.g., morphemes, words) into larger units (e.g., phrase, clauses). Failure to master prosodic chunking would lead to speech dysfluency, which is more noticeable in L2 speech production as a result of L2 speakers' limited chunking capability. During the chunking process, the syntactic structure construction is of great importance. However, previous studies mostly focused on the effect of syntax on speech production while paid little attention to the role of the number of syntactic constituents and syntactic structure complexity in L2 learning (Wang, Yang, & Lyu, 2004). The present study investigated the effects of the number of syntactic constituents and syntactic structure complexity on the chunking of words in the speech production of Chinese as a second language. The research questions were as follow:

1) *How does the number of syntactic constituents in one sentence affect the prosodic chunking of Chinese as L2?*

2) *How does the syntactic structure complexity of sentences affect the prosodic chunking of Chinese as L2?*

Two experiments were conducted to investigate the proposed questions. Fifteen English speakers with a low-level Chinese, 15 English speakers with a high-level Chinese and 15 native Chinese speakers were recruited as participants.

In Experiment 1, two groups of 10 subject-predicate sentences were generated for each of the short sentences (six or seven constituents in length) and the long sentences (10 or 11 constituents in length). In Experiment 2, another 10 subject-predicate sentences and 10 pivotal sentences were manipulated as simple and complex syntactic structure. A sentence recall task was used to elicit online oral production. Each sentence was visually presented on the screen for 5 seconds, and then participants were asked to orally repeat the sentence as accurately and fluently as possible within another 5 seconds.

We used chunk length as a measure of the prosodic chunking. Chunk length was defined as the number of words between each of two consecutive chunk boundaries. Chunk boundaries were obtained by two different ways: native speakers' perception and phonetic analysis method. Two native Chinese listeners were invited to mark the chunk boundaries by estimating the temporal speech information such as pause, final word vowel lengthening and pitch reset (Wang, Yang & Lyu, 2004). For phonetic analysis, any inter-syllable interval (ISI) more than 50ms was identified as chunk boundaries by using Praat. The analysis revealed that the phonetic analysis method had a strict criterion to measure the chunk length.

Our results suggested that the number of constituents and syntactic structure complexity affected L2 prosodic chunking, and that these effects were regulated by their L2 proficiency. Both groups of L2 learners have not acquired the chunking capability as the native Chinese speakers do. In specific, low-level Chinese speakers produced utterances in a word-by-word mode, while high-level speakers uttered in a phrase-by-phrase mode. Even the latter still failed to develop into a native-like sentence-by-sentence mode.

For low-level Chinese speakers, the prosodic chunking was constrained by both the number of constituents and syntactic structure complexity. Based on the *Perception* measurement, high-level Chinese speakers can be able to free from the constraint of increasing the number of constituents, but it is not this case for syntactic structure complexity. Based on *Phonetic Analysis* methods, both more constituents and complex syntactic structure decreased the chunking capability for high-level Chinese speakers. For native Chinese speakers, the prosodic chunking was not affected by both the number of constituents and syntactic structure complexity.

To conclude, for L2 speakers, most words were not chunked together; that is, a limited amount of chunks was stored in their long-term memory and these chunks were weak as well. Therefore, processing these small and weak chunks during speech production would consume more working memory resources, resulting in a less fluent speech than native speakers. As a result of the less well-established chunking capabilities, L2 speakers had difficulties in finding the right words, pronouncing them correctly, and manipulating syntax to construct word sequences in speech production. In contrast, for native Chinese speakers, those little pieces of information were chunked into higher-level meaningful units (Breen, Watson, & Gibson, 2011), and syntactic combination between words had achieved procedural automaticity. Therefore, native speakers were successfully freed from the activities of integrating words into chunks, and thus working memory resources could be reserved for high-level activities such as focusing on rhythm of the discourse, or the content and coherence of the speech.

References

[1] Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds). *Language and communication* (pp. 191-226). New York, NJ: Routledge.

[2] Jun, S. A. (2003). The effect of phrase length and speech rate on prosodic phrasing. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 483-486). August 3-9, Barcelona, Spain.

[3] Wang, B., Yang, Y., & Lyu, S. (2004). Acoustic analysis on prosodic hierarchical boundaries of Chinese. *ACTA ACUSTICA*, *29*(1), 29-36.

[4] Breen, M., Watson, D. G., & Gibson, E. (2011). Intonational phrasing is constrained by meaning, not balance. *Language and Cognitive Processes*, *26*(10), 1532-1562.

Sentence Final Particles and Wh-indeterminates in Beijing Mandarin

Yu-Yin Hsu[1], Ka-Wai Chan[1], Tsz-Shan Lo[1], Xia Wang[1] & Anqi Xu[2]

Hong Kong Polytechnic University[1], University College London[2]
yu-yin.hsu@polyu.edu.hk, ka-wai.kw.chan@connect.polyu.hk, tsz-shan.lo@connect.polyu.hk,
xuanqi@connect.polyu.hk, summer.x.wang@polyu.edu.hk

In this study, we used a speech production experiment to study whether and how Beijing Mandarin native speakers (n=6) use prosody to disambiguate *wh*-indeterminates, and to explore whether and how sentence final particles (SFPs) signal the clausal types and interact with focus marking. *Wh*-indeterminates refer to *wh*-phrases that are ambiguous between interrogative and indefinite readings, which has been attested in many languages (cf. Kuroda, 1965). For example, *wh*-phrases in Mandarin like *shenme* 'what' can be interpreted as a *wh*-interrogative in a *wh*-question (e.g., (1a)), or as an indefinite in a *yes/no* question (e.g., (1b)). Previous syntax-semantic studies have identified contexts that license the indefinite reading of *wh*-indeterminates, and the occurrence of SFPs such as *ma* for *yes/no* questions as in (1b) is one of such *wh*-indefinite-licensing contexts (Li, 1992; Cheng, 1994). While most linguistic studies have focused on sentences like (1a-b) for *wh*-indeterminates, we notice that in declarative sentences like (1c) with SFP *ba* indicating weak epistemic judgment, the *wh*-phrase also obtains an indefinite reading.

(1) a. Zhangsan mai-le shenme (ne)?
Zhangsan buy-ASP what WH-PARTICLE
'What did Zhangsan buy?'

b. Zhangsan mai-le shenme (ma)?
Zhangsan buy-ASP what Y/N-PARTICLE
'Did Zhangsan buy something/anything?'

c. Zhangsan mai-le shenme (ba).
Zhangsan buy-ASP what S-PARTICLE
'Zhangsan probably bought something/anything'

Examples in (1) show that *wh*-indeterminates in Mandarin not only are lexically ambiguous but are also relevant to structural ambiguity, given that SFPs often are not obligatory in Chinese.

Considering prosody as one of the disambiguation devices, some prior studies have reported that *wh*-indeterminates while functioning as *wh*-interrogatives manifest more acoustic prominence (due to the focus status) than *wh*-indefinites in languages such as Korean (Jun and Mira, 1996), Japanese (Ishihara, 2003; Kitagawa, 2007), and German (Truckenbrodt, 2013). Yet, some studies reported that no acoustic differences on the *wh*-indeterminates that distinguished interrogative and indefinite readings, and distinction may be found on the following units (e.g., Yon, 2018).

Different results were also reported for *wh*-indeterminates in Mandarin. Hu (2002) studied *wh*-subject *shei* 'who' and *shenme* 'what' and reported that Mandarin speakers expressed *wh*-interrogatives acoustically different from *wh*-indefinites: it was reported that *wh*-phrases had higher mean F0 in *wh*-questions, and the verb phrase of a sentence showed higher mean F0 in *yes/no* questions. In this study, only descriptive statistics results were reported for mean F0, duration and amplitude (with SD), and some inter-participant differences were found. For Taiwan Mandarin, Shyu and Tung (2018) reported two different findings; first, based on eight tokens from a speech corpus, they reported that some differences were found between *wh*-interrogatives and indefinites, however the syntactic and the phonetic contexts where these eight tokens occurred were different; second, their production study showed that participants did not acoustically disambiguate the two meanings of *wh*-indeterminates; since all their participants responded to the same two items for one context, it is difficult to draw a general conclusion for Taiwan Mandarin. Thus far, the findings about Mandarin *wh*-indeterminates remains inconclusive.

In this study, we examined the prosody of *wh*-questions and *yes/no* questions like those in (1) containing *wh*-indeterminates, and use declarative sentences as the baseline. Four *wh*-phrases (*shei* 'who', *shen.me* 'what', *na.li* 'where', and *shen.me-dong.xi* 'what thing') were used to constructed each type of sentences with five versions of the Tone4 verbs. In total, 60 target sentences (3 sentence types x 4 *wh*-phrases x 5 verb) and 40 filler sentences (in different sentence structure with no SPFs) were used in an experimental session. Each trial consisted of a pre-recorded leading context (25 character long) and a target sentence (like those in (1)) that participants used to respond. The leading contexts were pre-recorded by a female speaker of Beijing Mandarin. The acoustic measurements were generated by ProsodyPro 5.7.6 (Xu, 2013) for duration, fundamental frequency (F0) range and time normalized F0. Linear Mixed Effects models were conducted on duration and F0 range using the lme4 package in R (Bates et al. 2015). Example trails are shown below.

(2) *Wh*-question

A: 我等会儿要出去买饭。有没有人需要我顺路办事或者带饭的

'I am about to go to buy my meal. Does anyone need me to run simple errands or to buy a meal on the way?'

B:你可以帮忙带什么呢

'What can you help me bring?'

(3) *Yes/no* question

A: (The leading context same as the one in the *wh*-question)

B: 你可以帮忙带什么呢

'Can you help to bring something?'

Our results show distinction between declarative sentences and questions on the regions of *wh*-phrases and on the region of SFPs. This suggests that Mandarin speakers prosodically disambiguate *wh*-indeterminates. We also found very interesting and consistent patterns of duration and F0 range across four *wh*-phrase types. Unlike the on-focus lengthening effect reported in previous studies, in our study the last syllable of *wh*-phrases in both questions was shorter than that syllable in declarative sentences, especially when their *wh*-phrases were *what* (y/n: $p < .001$) and *where* (y/n: $p < .001$; wh: $p < .001$). However, the SPFs (which immediately following the *wh*-phrases) in both types of questions were significantly longer than the SFP in declaratives (*p*s<.001). Furthermore, the duration of SFPs was the longest in *yes/no* questions across four *wh*-phrase types rather than SFPs in *wh*-questions. The focus marking pattern in terms of F0 range was also consistent across four types of *wh*-phrases. When comparing with declaratives, the F0 range of statements tended to be wider than questions in the last syllable of *wh*-phrases, especially in *what* (y/n: $p = .004$) and *where* (y/n: $p = .008$; wh: $p = .076$).

These patterns seem to suggest that the occurrence of SFPs specifically defines the sentence types, and this information requires the prosodic organization to comply in expressing the differences between questions from statements. In sum, these results suggest that the internal organization at different structural levels may interact with the prosody system.

References

[1] Kuroda, S.-Y. (1965). Generative grammatical studies in the Japanese language. Doctoral Dissertation, MIT, Cambridge, MA.

[2] Li, Y.-h. A. (1992). Indefinite Wh in Mandarin Chinese. *Journal of East Asian Linguistics, 1*(2) 125-155.

[3] Cheng, L. L.-s. (1994). Wh-words as polarity items. In *Chinese Languages and Linguistics, vol. II: Historical Linguistics* (pp. 615-640).

[4] Jun, S.-a., & M. Oh. (1996). A prosodic analysis of three types of wh-phrases in Korean. *Language and Speech, 39*(1), 37-61.

[6] Kitagawa, Y. (2007). When we fail to question in Japanese. In *Proceedings of the 2nd workshop on prosody, syntax, and information structure (WPSI 2).*

[7] Truckenbrodt, H. (2013). An analysis of prosodic F-effects in interrogatives: Prosody, syntax and semantics. *Lingua, 124*, 131-175.

[8] Yon, J. (Accepted). Meaning and Prosody of Wh-Indeterminates in Korean. Massachusetts Institute of Technology.

[9] Hu, F. (2002). A prosodic analysis of wh-words in Standard Chinese. In *Speech Prosody*, Aix-en-Provence, France.

[10] Shyu, S.-i., & Tung, T.-l. (Accepted). Taiwan Mandarin Wh-phrases and Prosody. *Studies in Prosodic Grammar*.

[11] Xu, Y. (2013). A Tool for Large-scale Systematic Prosody Analysis. In *TRASP 2013*, Aix-en-Provence, France.

[12] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1) 1-48.

Perception and Analysis of Utterance-Final Lengthening of Cantonese maː^{33}

Chaak-ming Lau

The Chinese University of Hong Kong (Hong Kong)
chaakming@gmail.com

Nutshell This paper investigates the perception of the Cantonese sentence final particle (SFP) maː^{33}, and confirms that its two distinctive usages (formal question vs reasoning statement) can be perceived by native speakers in a response-choice task. Further analysis reveals that the duration of the statement particle is significantly longer, and can be analysed as the combination of SFP maː^{33} and boundary lengthening :%.

Background This paper explores the two distinct realisations of SFP maː^{33} [1]. This SFP is known to have two meanings: it can turn a statement into a formal yes-no question, as in (1); and it can also be used to make a comment about the reason of something, as in (2).

(1) question **maː^{33}**	(2) statement **maː^{33}**
nej^{23} hɵj^{33} maː^{33}	nej^{23} hɵj^{33} maː^{33}
you go SFP	you go SFP
Are you going?	That is because you are going.

The same two morphemes can also be found in Mandarin, and the two can be distinguished by pitch [2]. Similar strategy is impossible in Cantonese. The two functions are quite distinctive and must be confusing if native speakers cannot distinguish the two.

Experiment The goal is to establish that the two kinds of maː^{33} can be perceived by native speakers, i.e. whether it is a question or a statement is not judged purely by context.

Task Subjects (16 participants recruited) were asked to participate in a simulated instant messaging activity and their task was to choose the correct response to answer a friend's recorded message. In each trial, the subject would see a message written in Cantonese (displayed in Han characters) that was supposedly sent by the subject, and then the subject would hear a recorded message form a friend. Next the subject would be asked to respond to this friend by choosing the correct message from two available options.

Stimuli There were 24 groups of sentences, presented in one of the four conditions below. Conditions *ama_S* and *A-not-A* were controls. Another 48 groups of fillers were added.

(3)	(a) *ma_S*	kʰɵj^{23}	puːn^{25}sɐn^{55}	sɪk̚^{55}	nej^{23}	maː^{33} …	
	(b) *ama_S* (control)	kʰɵj^{23}	puːn^{25}sɐn^{55}	sɪk̚^{55}	nej^{23}	a^{55}maː^{33} …	
		3.SG	originally	know	you	SFP	
	That is because he knew you before.						
	(c) *ma_Q*	kʰɵj^{23}	puːn^{25}sɐn^{55}	sɪk̚^{55}	nej^{23}	maː^{33} ?	
		3.SG	originally	know	you	SFP	
	(d) *A-not-A* (control)	kʰɵj^{23}	puːn^{25}sɐn^{55}	sɪk̚^{55}m̩^{21}sɪk̚^{55}		nej^{23}	kaː^{33} ?
		3.SG	originally	know_AnotA		you	SFP
	Did he know you before?						

The two choices were *answers* ("Yes, he knew.", valid for questions) and *doubts* ("Oh really?", valid for reasoning statements). The statements were recorded by a female speaker who were given an appropriate context so that the resulting recordings would be in expected intonations.

Results Subjects were expected to choose *answers* in response to the questions, and *doubts* in response to reasoning statements. That describes the general trend of the controls and the *ma_S* condition. The response type for *ma_Q* (that is question usage of the particle) was less clear, probably because some subjects found *doubts* would also be valid responses.

Discussion One would expect chance-level performance (i.e. 48-48 for both *ma_S* and *ma_Q*) if the two were not distinctive. The experiment showed that this distinction is categorical and can be perceived by native speakers.

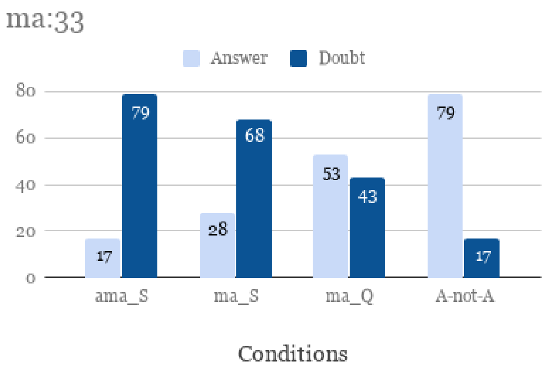


Fig.1 Answer versus Doubt in subject response

Post-hoc analysis of the distinction of the two particles Recorded stimuli (48 recordings) were analysed to confirm the difference between the two usages. Both question and statement usage of ma:^{33} show slight declination (a drop of less than 5Hz for both questions and statements), and the statement usage is significantly longer than question ma: (statement ma: μ=470ms, S.D.=39.2; question ma: μ=198ms, S.D.=28.4).

Proposal The difference between (1) and (2) is purely durational. SFPs are systematically lengthened as a result of the realisation of the boundary lengthening morpheme :**%**, which causes lengthening of the last syllable of the utterance. This can be controversial since boundary tones are usually not compatible with SFPs, e.g. interrogative **H%** cannot occur with any SFPs. I argue that the incompatibility is solely due to syntactic restrictions.

Implications Cantonese has a dense specification of tones, and it is likely that the temporal dimension is used as an alternative to F0 for intonation morphemes. This study confirms the impressionistic description about the "protracted intonation" [3], which should be added to the boundary tone inventory, detailed in Cantonese ToBI [4]. The grammaticality of the co-occurrence of boundary tones and SFPs can be used to ascertain the hierarchical structure of the left periphery [5]. If this lengthening (:%) is analysed as an intonation morpheme that heads a projection, then one needs to explain why it occurs after ma:^{33}, which is supposed to be the highest projection (AttitudeP)[6]. One either needs to analyse ma:^{33} as a syntactically lower element, or assume other higher projections. Syntactic analysis of the left periphery would certainly benefit from further investigation of boundary tones in other SFP languages.

References

[1] Leung, C.-S. (2005). *A Study of the Utterance Particles in Cantonese as Spoken in Hong Kong*. Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.

[2] Xiong, Z.-Y. (2005). Putonghua-zhong "ma"-zi Shifei Wenju de Yudiao Yin'gao Tezheng Fenxi [Pitch Feature Analysis of Intonation of "*ma*" yes-no questions in Putonghua]. *Technical Acoustics 2015.24* (3).

[3] Mai, Y. (2018). The Quick Intonation and the Protracted Intonation in the Sentence-final Position in Cantonese. *Current Research in Chinese Linguistics (CrCL)*, *91*(1), 3-16.

[4] Wong, W. Y. P., Chan, M. K. M., & Beckman, M. E. (2005). An Autosegmental-Metrical Analysis and Prosodic Annotation Conventions for Cantonese. In S.-A. Jun (Ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press.

[5] Sybesma, R., & Li, B. (2007). The dissection and structural mapping of Cantonese sentence final particles. *Lingua*, *117*(10), 1739-1783.

[6] Pan, V. J. (2019). *Architecture of the periphery in Chinese: Cartography and minimalism*. London & New York:Routledge.

A phonetic study of the Cantonese rising-falling intonation

Mei-ying Ki

The Chinese University of Hong Kong (Hong Kong)
kimeiying@gmail.com

Summary: This paper describes the phonetic realisation of the rising-falling intonation in Cantonese [1], which expresses an objection towards what the speaker believes about the hearer's belief. Based on the phonetic realisation, the intonation is analysed as a boundary tone HL%. It is further demonstrated how the intonation interacts with the six lexical tones of Cantonese. The realisation of lexical tones is affected by the intonation, but the tonal features are still preserved through various means. The observation may improve our understanding of the role that intonation plays in tonal languages like Cantonese in general.

Research questions: Two research questions are to be addressed: 1) Is the intonation only a boundary tone or it affects the utterance body? 2) How does the intonation interact with the lexical tone of the final syllable?

Method: In the production test, four native speakers of Hong Kong Cantonese (two males and two females) were invited to record the following sentences, according to the contexts provided: [1]

(1) is a statement, while (2) is a sentence with the intonation attached. The syllable "wai" was recorded in all six lexical tones. The total number of sentences tested is: 4 subjects × 2 intonations × 6 tones × 3 trials per subject = 144 sentences.

Results: The underlined parts (five syllables in each condition) were extracted for analysis in Praat [2]. For each syllable, 50 points of F0 data were extracted by ProsodyPro [3] and converted into Z-scores. Items with an absolute Z-score larger than 2 were discarded, which accounted for 3.20% of the original set of data. An SSANOVA (Smoothing Spline ANOVA) analysis was conducted. The data of the six tones are shown separately in Fig.1. The dotted line separates the last syllable.

The I curves (blue, rising-falling) are generally higher than the S curves (red, statement). This paper argues that this is not a global rise for the following reasons: the I curves are not higher than the S curves totally and steadily; variation was observed among different subjects; the use of the higher F0 for this intonation can be attributed to paralinguistic factors: the intonation is often used in denials, which may easily lead to a rise of F0.

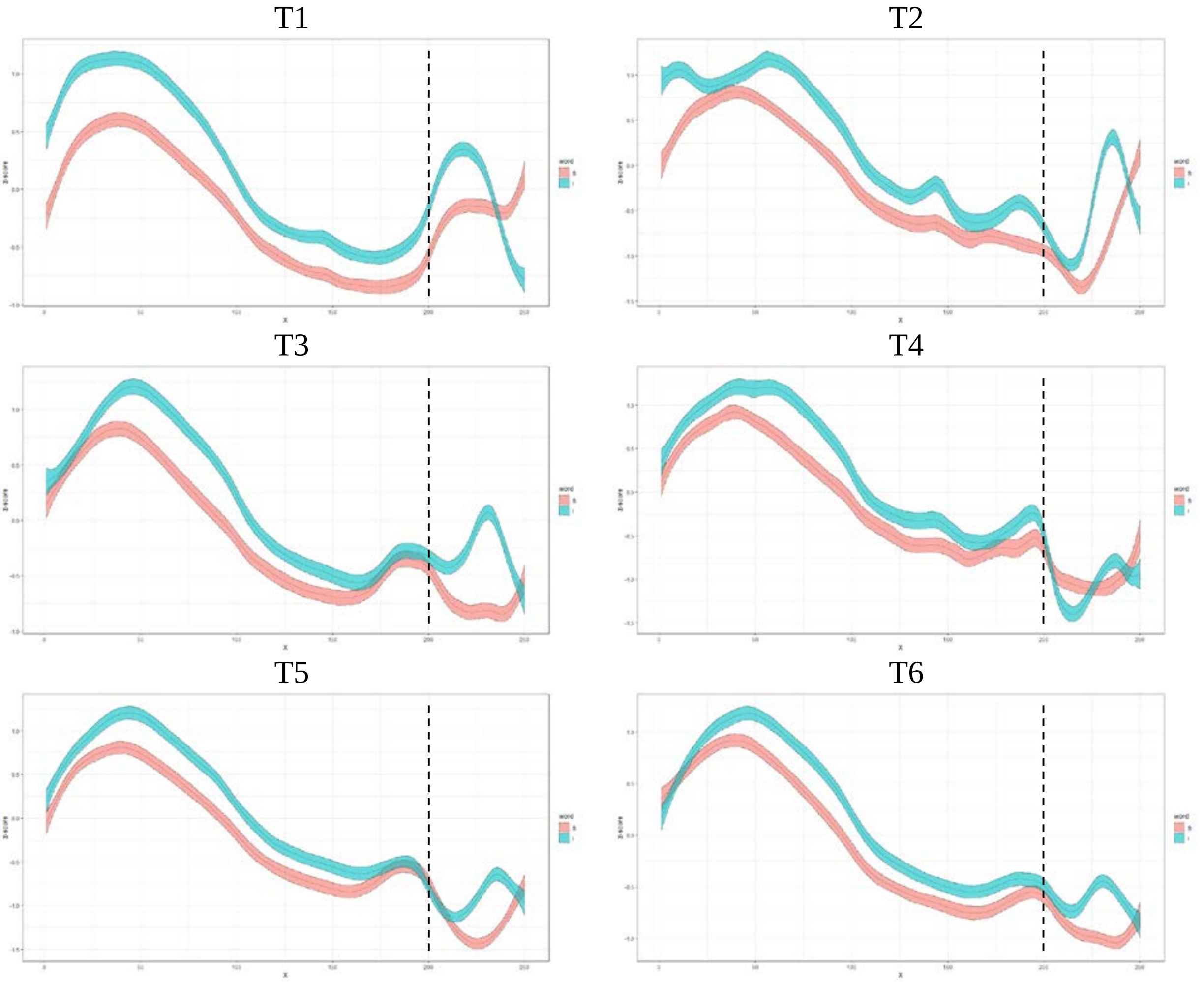
Regarding the effect on the lexical tones of the sentence-final syllables, except for T1 – which is at the upper limit of pitch range (so the realisation may be affected by ceiling effect), the realisations of the other five tones are similar. Simply speaking, instead of modifying the pitch of the whole syllable, the initial part of the syllable remains unaffected (the two curves are basically overlapping right after the dotted lines), with the rising-falling contour attached right after it.

Beside the initial part of the lexical tone, some other tonal features are also preserved. For example, for T2 and T5 (rising tones), the apexes of the two curves in the corresponding graphs are at the same height, showing that the pitch target is preserved. The end points of the I curves in T3 and T6 (level tones) are also at the same height as the level parts of the corresponding S curves,

[1] The symbol "↷" is used for simpler representation of the intonation here. Cantonese data are transcribed in LSHK Jyutping.

which is another cue of preservation of tone height. Also note that a much larger and longer falling pattern was observed in T1, while the rising pattern starts at an extra low position in T4 (the lexical tone that is at the lower limit of pitch range). These cues may also help tone perception (under the disturbance of intonation).

Conclusion: This study investigates the phonetic realisation of the rising-falling intonation (analysed as a boundary tone HL%) in Cantonese. The initial part of the final syllable remains unaffected, with the intonation contour attached at the end. Some other tonal features are also preserved, which may act as acoustic cues for tone perception. This will be left for further research.
Fig.1 Sentences with and without the rising-falling intonation in six lexical tones



References

[1] Kwok Chang, H. H., & Luke, K. (1986). An Initial Study of Intonation in Cantonese. *Yuwen Zazhi, 13*, 32-40.

[2] Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer [Computer program]*. Version 6.0.50, retrieved 31 March 2019 from http://www.praat.org/

[3] Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (pp. 7-10). Aix-en-Provence, France.

Cumulative usage effects in inflectional paradigm: Probability of being affixed

Daiki Hashimoto

University of Tokyo (Japan)
daiki@phiz.c.u-tokyo.ac.jp

Past usage patterns affect production of a linguistic unit in future speech. This effect is called a "cumulative usage effect." For example, words with lower informativity, which have often been affected by probabilistic reduction, are produced with shorter duration even when they occur in a less predictable context (Seyfarth [1]; Sóskuthy & Hay [2]). This effect can be neatly captured by Exemplar Theory (Pierrehumbert [3]), which consists of the following key hypotheses:

H1: Sounds encountered in speech are stored with detailed phonetic information as "exemplars" in the mental space.

H2: Exemplars with similar properties (e.g., phonetic, semantic, and social information) cluster together, and "categories" are formed.

H3: A speaker begins speech production by activating an intended category. Then, the speaker chooses a couple of exemplars belonging to the category, and average their phonetic values to form a production target.

The informativity-oriented reduction can be captured, by positing that words that are likely to undergo probabilistic reduction may be represented by a larger number of exemplars with reduced phonetic properties.

This raises a research question of whether cumulative usage effects are also observed in relation to inflectional paradigms of lexemes. The aim of this talk is to inform our understanding of how lexemes are represented in our mind by addressing the following concrete RQ:

RQ: "Is the duration of lexeme-final /s/ (e.g., *bus* and *buses*) affected by the likelihood of being followed by bound morphemes (i.e., *-es, -ed,* and *-ing*)?"

It is known that a coronal fricative phoneme /s/ is produced with shorter duration in word-medial position than in word-final position in American English (Umeda [4]): therefore, it can be expected that lexeme-final /s/ is produced with shorter duration in New Zealand English when they are attached by bound morphemes (e.g., *buses* and *iced*) in comparison with when they are not (e.g., *bus* and *ice*). We crucially hypothesize that exemplars encoding affixed forms and unaffixed forms of a lexeme are closely represented, because they share a variety of properties (e.g., phonological information, syntactic information, semantic information, and social information). The following prediction can be put forward on the bases of the hypotheses:

Prediction: Lexeme-final /s/ may be produced with shorter duration in lexemes that are in general more likely to be followed by bound morphemes (e.g., *-s*, *-ed*, and *-ing*).

This prediction is illustrated in Figure 1. According to our data, which is explained below, *chase* is a word that is likely to be affixed, while *base* is a word that is unlikely to be affixed. Namely, *chase* is likely to be produced with shorter lexeme-final /s/ because of the following bound morphemes, and consequently this lexical category may be represented by a larger number of exemplars with shorter lexeme-final /s/. Because of this representation, it can be predicted that the lexeme-final /s/ in *chase* may be produced with shorter duration in general even when it is not followed by bound morphemes. This is because an exemplar with shorter /s/ is more likely to be chosen in production of this word. The reverse is true for the lexical category *base*.

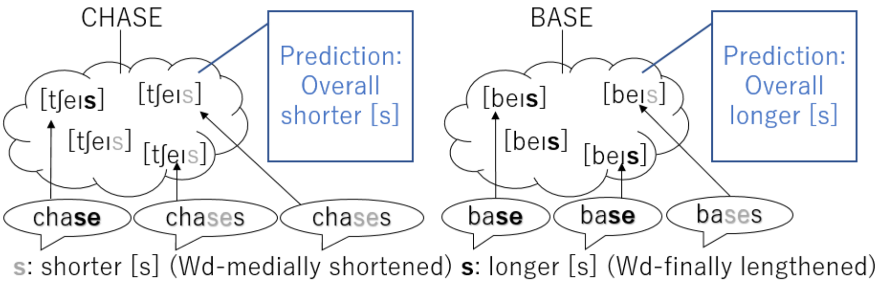


Fig. 1 Prediction about cumulative usage effects on lexeme-final /s/

Lexemes ending with non-morphemic /s/, which can be countable nouns or verbs, were collected from CELEX (Baayen et al. [5]). Then, the recorded tokens of the unaffixed and affixed forms (i.e., followed by bound morphemes) were collected from ONZE Corpus (Gordon et al. [6]). Some tokens were removed for acoustic or morphological reasons. Finally, 16,059 tokens with lexeme-final /s/ were collected: 9,120 unaffixed tokens and 6,939 affixed tokens.

The 16,059 tokens of lexeme-final /s/-sounds were hand-fitted into a mixed-effects linear regression model using the *lmer* function in the *lme4* library implemented in R. The model includes some control variables such as word predictability and prosodic factors. As expected, lexeme-final /s/ is shorter when they are affixed (p<0.05). As for the probability of being affixed, the effect is significant only in affixed forms (p<0.05), but it is non-significant in unaffixed forms (p=0.46), see Figure 2. This result is in line with our Prediction, and it suggests that cumulative usage effects are also observed in relation to inflectional paradigms of lexemes. This can be captured by positing that our exemplar space is developed in accordance with speech experience, that is, a category of a lexeme that is likely to be affixed is represented by a larger number of exemplars that are NOT affected by word-final lengthening.

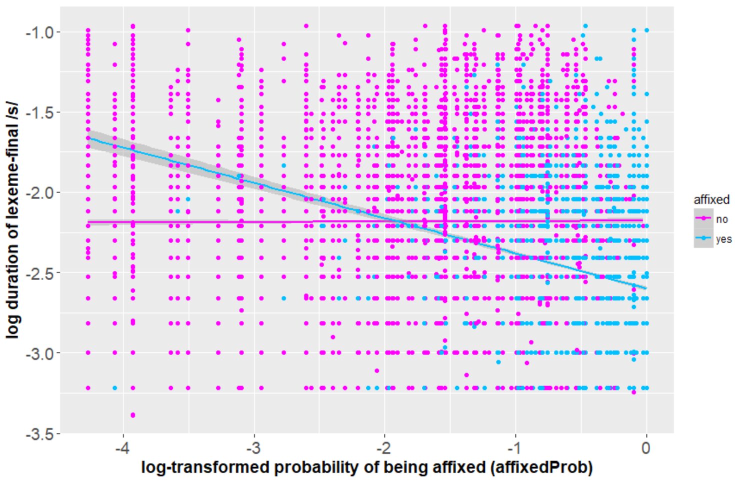


Fig.2 Relationship between log-duration of /s#/ and log-probability of being affixed

References

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

[2] Sóskuthy, M., & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition*, 166, 298-313.

[3] Pierrehumbert, J. B. (2001). Exemplar dynamics, word frequency, lenition, and contrast, In J. Bybee, & P. Hopper (Eds) *Frequency effects and the emergence of linguistic structure* (pp. 135-157). Amsterdam: John Benjamins.

[4] Umeda, N. (1977). Consonant duration in American English. *Journal of Acoustic Society of America*, 61, 846-858.

[5] Baayen, H., Richard P., Léon G. (1995). *The CELEX Lexical Database*. Philadelphia: Linguistic Data Consortium (University Pennsylvania).

[6] Gordon, E., Maclagan, M., Hay, J. (2007). The ONZE Corpus. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds), *Models and methodsin the handling of uncontroversial digital corpora. Diachronic corpora* (Vol. 2, pp.82-104). Basingstoke, Hampshire: Palgrave Macmillan.

Cross-linguistic differences in long-lag priming

Stephen Politzer-Ahles[1], Leon Ka Keung Lee[1], Lei Pan[1], & Jueyao Lin[2]

[1]*The Hong Kong Polytechnic University (Hong Kong),* [2]*City University of Hong Kong (Hong Kong)*
sjpolit@polyu.edu.hk, leon-leechk@hotmail.com, panlei.pl.pan@connect.polyu.hk, jueyao.lin@connect.polyu.hk

Long-lag priming ([1], among others) is a useful technique for investigating the nature of lexical relationships and lexical activation patterns while avoiding potentially problematic influences of task-based strategies by the participants or of stimulus-based confounding issues such as associative or form relationships between stimuli. In long-lag priming experiments, participants see many stimuli and conduct some task (such as judging whether the stimulus is a real word or not) for every stimulus they see. Unbeknownst to participants, some stimuli are "primed" by a related stimulus (such as a repetition of the same stimulus, or a morphologically-related stimulus) that has occurred many trials earlier; other stimuli are not primed (the stimuli that occur before these are all unrelated to the target stimulus). Unlike in traditional paired priming paradigms, where primes and target stimuli are presented together and their relationship may be obvious to participants, in long-lag priming participants generally do not notice that some stimuli have been primed.

One important feature of long-lag priming is that the response time to target stimuli is facilitated by primes that are morphologically related to the target (e.g., *hunt… hunter*), compared to primes that are unrelated to them (e.g., *book… hunter*). This pattern is not due to the meaning relationship between prime and target, but to the morphological relationship (as associatively related prime-target pairs without morphological relationship do not show facilitation). Another important feature is priming may combine lexical effects and episodic effects [1]. Specifically, if the "related" prime for a target is the same word as the target (e.g., *hunter… hunter*), then there are two factors which facilitate response to the target: the fact that the lexical entry for HUNTER has been activated already (lexical priming), and the fact that the physical stimulus (the visual representation *hunter* or the auditory stimulus that was presented) is fresh in the participant's memory (episodic priming). Some evidence for this dissociation is that episodic priming, but not lexical priming, is absent if the prime is a different recording of the same word (e.g., $hunter_{\text{male-voice}}$ … $hunter_{\text{female-voice}}$). In such a situation, lexical priming is still present (as the lexical entry for HUNTER is activated each time) but episodic priming is not (as the stimuli are physically different). Target stimuli in this situation are responded to more quickly than targets with unrelated primes (suggesting that lexical priming is present) but more slowly than targets with completely identical primes (suggesting that episodic priming is gone).

These patterns have been observed in languages like French. In the present study, we tested these features of long-lag priming using Mandarin. Whereas morphological relationships in previous long-lag studies were generally based on concatenative morphology (e.g., *hunt* and *hunter* share a root but differ because of a suffix), the present study used tone to manipulate morphology. Mandarin syllables with Low tone undergo phonological alternations which, in certain contexts, cause them to be pronounced with Rising tone instead. Thus, segmentally identical syllables with Low and Rising tone (e.g., shi^{L} and shi^{R}) could be allomorphs of the same morpheme; in other words, they are morphologically related.

In Experiment 1 ("same voices"), 153 native Mandarin-speaking participants completed a long-lag priming task following the design of [1]. Participants heard 96 target monosyllables, each of which was preceded by either an unrelated prime (e.g., hua^{F}…shi^{L}), a morphologically related prime (e.g., shi^{R}…shi^{L}), or an identical prime, which was the exact same recording of the same monosyllable (e.g., shi^{L}…shi^{L}). There were 18-56 other syllables in between each prime and target. Mixed in with the critical prime-target pairs were 56 filler prime-target pairs, and 288 nonwords (phonotatically legal accidental gap syllables). Participants performed an auditory lexical decision for every stimulus they heard. Stimuli were arranged into three lists in a Latin square design. After removing incorrect responses and outliers (trials more than 1.5IQR away from that participant's or item's median), priming effects were calculated by subtracting the response time for targets with unrelated primes from the response times for the other two targets. As shown in Figure 1, there was robust facilitation from identical primes (b=-65.78, t=-13.14 in a maximal mixed-effects model), but not from morphologically related primes (b=9.83, t=1.71), contrary to what has been observed in other languages (e.g., [1]).

Experiment 2 ("different voices", with 120 participants, out of a planned 153) followed the same design as the same-voices experiment, except that primes and targets were different recordings, spoken by different people (one woman and one man), and the stimuli organized into six Latin square lists instead of three. As shown in Figure 2, there was again robust identity priming (b=-61.36, t=-11.75) and not morphological priming (b=5.74, t=0.89), contrary to what was observed in other languages but replicating the pattern from the first experiment. Contrary to what has been observed previously [1], the same-word priming effect is about the same (61-65 ms) regardless of whether the words are spoken in the same voice or different voices.

The results challenge current understanding of the mechanisms underlying priming. The fact that morphological priming did not occur in Mandarin suggests that either there is something different about how lexical representations work in Mandarin compared to other languages that have been tested, or that some other moderating property of the stimuli interacts with morphological priming. One possibility is homophony: most of our monosyllabic stimuli correspond to multiple morphemes, so it is possible that morphological priming only occurs when one morpheme can be uniquely identified and activated; this is a question for future study. The fact that episodic priming played a smaller role (or none at all) compared to in previous studies suggests that current understanding of the role of episodic memory in priming experiments must be refined.

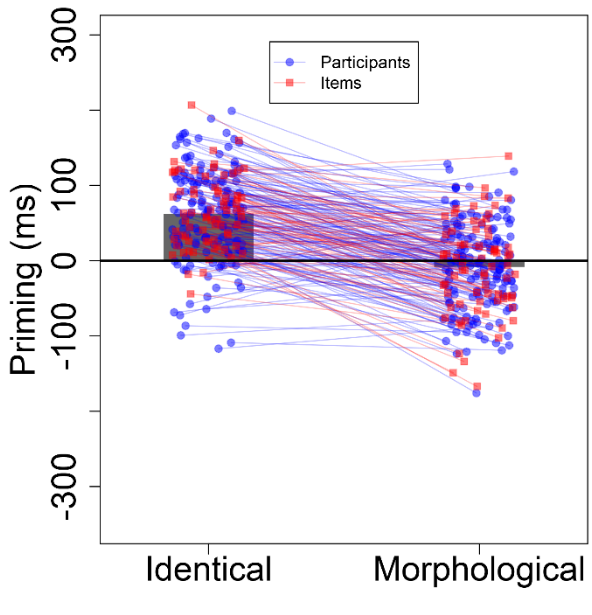


Fig.1 Experiment with same voices for prime and target

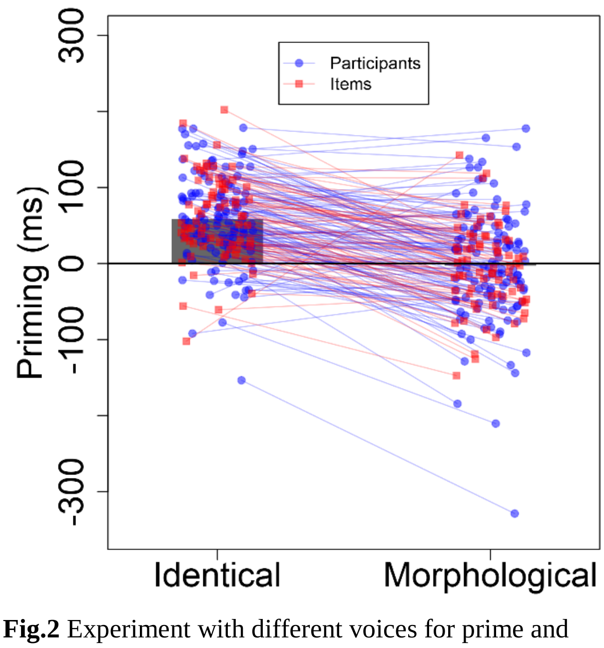


Fig.2 Experiment with different voices for prime and target

References

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

Spectral properties of diphthongs in two varieties of Assamese

Leena Dihingia & Priyankoo Sarmah

Indian Institute of Technology Guwahati (India)
leenadihingia@gmail.com, priyankoo@gmail.com

Several studies have evidenced the presence of an articulatory undershoot in diphthongs [1, 2, 3, 4]. An element of the diphthong, usually the less prominent one often fails to achieve the endpoint of the vocalic element. Bladon [5] opines that one reason for this phenomenon lies in the perceptual probability that listeners pay more attention to the trajectory of the diphthong rather than the actual target. Another assumption comes from Hu [6] who opines that the distribution of the diphthongs in comparison to the corresponding monophthongs in an acoustic plane determines the presence of a spectral target. If the diphthong element has a more dispersed distribution than the monophthong, it tends not to have target and if the distribution is more comparable, the diphthong tends to have spectral target. Considering the variations in monophthongs in two varieties of Assamese, the paper examines the spectral properties of diphthongs in the two varieties.

Assamese is an Indo-Aryan language spoken by about 13 million people in the North-east Indian state of Assam and it has several varieties based on the geographic location of the speakers. Eastern Assamese (EA) and Western Assamese (WA) have distinct variations in several phonological, morphological and lexical aspects [7]. In case of monophthongs, the canonical 8 vowel system undergoes two different mergers in the varieties whereby EA becomes a seven vowel system and WA, an eight vowel system [8].

20 native speakers of Assamese, 5 males and 5 females each speaking EA and WA varieties were recorded for the study. Six Assamese falling diphthongs /ai, au, oi, ou, ei, eu/ and two level diphthongs /iu, ui/ appeared in a list of words. Diphthongs were annotated as composed of onset element, transition and offset elements. Onset of a steady state formant was considered vowel beginning and the termination of the steady state was considered vowel end. Formant frequencies F1 and F2 were extracted at midpoint of the elements in Hertz (Hz). Hz units were then converted to Mel scale using an inbuilt function on Praat 5.4. Plots of vowel group means with 1 standard deviation ellipses were generated on F1-F2 space. The mean durations of diphthong elements were calculated. In order to check the spectral dynamics of the diphthongs, the mean durations and the rate of spectral change for F1 and F2 was calculated following Tsukada [9].

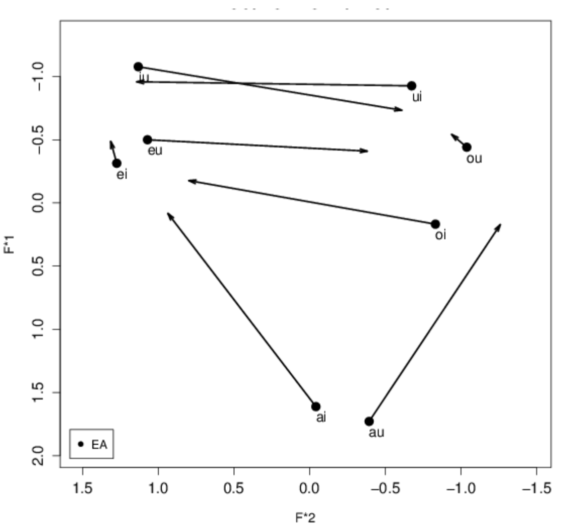


Fig 1: Diphthongs in EA variety in F1-F2 plane.

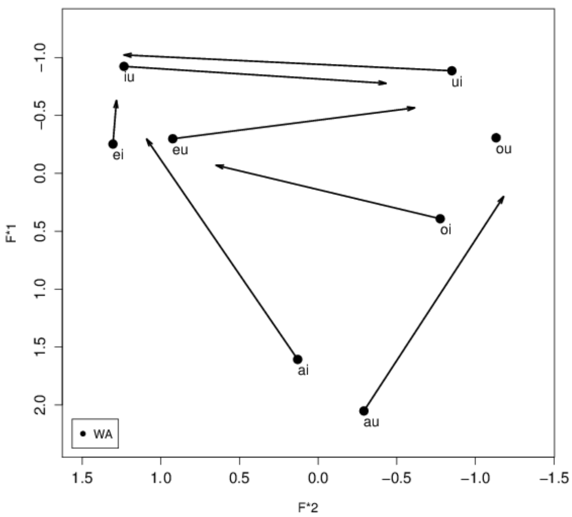


Fig 2: Diphthongs in WA variety in F1-F2 plane.

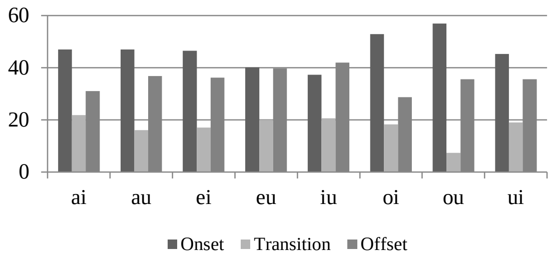


Fig. 3: Mean durations of diphthong elements in percentage for EA variety.

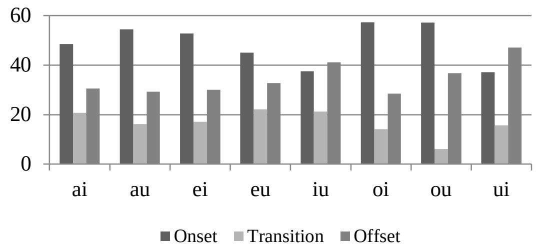


Fig. 4: Mean durations of diphthong elements in percentage for WA variety.

The diphthongs analyzed in both the varieties have shown similar patterns. In case of temporal properties, onset elements in both the varieties have been found to be longer than the offset for all the falling diphthongs. Transitions are the shortest elements of the diphthongs. Spectral results show that in both varieties falling diphthongs never reach the target, evidencing the presence of an articulatory undershoot, the offset in this case. However, the results also show that the onsets are more dispersed than the offsets. This suggests that falling diphthongs are composed of a single target in both EA and WA. The low transition duration in case of /ou/ and the high overlap of ellipses of the elements also suggest a possible monophthongization of this diphthong. The level diphthongs /iu/ and /ui/ have a clear target, however, it is the vowel /i/ rather than an element that has been seen as the target. /u/ in both the cases have larger dispersion and is more fronted than the monophthong. This suggests that in case of level diphthongs, although the targets may be different, it is the trajectory that is more important.

References

[1] Solomon, I & Sara, S.J. (1984). “English diphthongs, [ai, oi, ou],” M.P.R. Van den Broecke, A. Cohen (Eds.), *Proceedings of the 10th International Congress of Phonetic Sciences*, Foris, Dordrecht, pp. 411-414. 1984

[2] Jones, D. (1960). *Outline of English Phonetics* Heffer, Cambridge.

[3] Lindblom, B. & Studdert-Kennedy, M. (1967). “On the role of formant transitions in vowel recognition,” *Journal of the Acoustical Society of America*, 42, pp.840-843.

[4] Gurlekian, A. & Franco, H.E. (1984). “Recognition of a Spanish VV sequence,” M.P.R. Van den Broecke, A. Cohen (Eds.), *Proceedings of the 10th International Congress of Phonetic Sciences*, Foris, Dordrecht, pp. 237-242.

[5] Bladon, A. (1985). “Diphthongs: A case study of dynamic auditory processing.” *Speech Communication*, *4*(1), pp.145-154.

[6] Hu, F. (2013). “Falling diphthongs have a dynamic target while rising diphthongs have two targets: Acoustics and articulation of the diphthong production in Ningbo Chinese,” *Journal of the Acoustical Society of America*, Vol. 134, No. 5, Pt. 2, 4199, Nov 2013; [full paper in Chinese] in: Collected Papers in Language Studies, 10: pp.12-37, Shanghai: Shanghai Lexicography Press, 2013.

[7] Goswami, G.C & Tamuli, J. (2003). “Asamiya” Cardona and Jain, eds, pp. 391–443.

[8] Tsukada, K. (2008) “An acoustic comparison of English monophthongs and diphthongs produced by Australian and Thai speakers.” *English world-wide*, *29*(2), 194-211.

[9] Dihingia, L & Sarmah, P. (2017). “Vowel perception in two Assamese varieties”. *Proceedings of Seoul International Conference on Speech Sciences 2017*, Seoul, South Korea.

On the phonetic contrast between fortis and non-fortis fricatives in Korean: interaction with prosodic position effect and vowel coarticulation

Boram Lee[1], Hiyon Yoo[2] & Cécile Fougeron[1]

[1]*Laboratoire de Phonétique et Phonologie, UMR7018, CNRS/ Université Sorbonne-Nouvelle (France),*
[2]*Laboratoire de Linguistique Formelle, UMR7710, CNRS/ Université Paris Diderot (France)*
lee.boram@sorbonne-nouvelle.fr, hiyoni@gmail.com, cecile.fourgeron@sorbonne-nouvelle.fr

The three-way contrast between stops and two-way contrast between fricatives typical of Korean has attracted a great deal of attention in the phonetics and phonological literature. Regarding the fricatives, on which this study is focused, much of the debate is concerned by the definition of the counterpart of the fortis fricative /s*/: a lenis fricative for some [1, 2], an aspirated fricative for others [3, 4, 5, 6], or an aspirated-lenis fricative for others [7, 8, 9]. Our interest in the present study is not in the definition of this fricative (that we will call non-fortis /s/, by simplification), but in the way factors of variation do interplay with the phonetic realization of the contrast between /s/ and /s*/. Reported cues to this contrast include properties at the onset of the following vowel (intensity, spectral tilt, voice quality, F1…), but also properties of the fricative per se (longer duration, higher frequency noise, more linguopalatal contact, higher airflow resistance, smaller glottal area for /s*/) [8, for a review]. It seems therefore interesting to question this contrast with a new lens: how is it affected by factors which also could affect these specific acoustic dimensions? Here we will look at prosodic position and following vocalic context effects.

Prosodic position refers to the position of the fricative relative to prominences and boundaries. It is well known by now that the phonetic make-up of segments is affected by the prosodic organization of the utterance they are in, and that segments at strong prosodic positions (under accent or close to strong boundaries) are strengthened spatially and/or temporally. A long-standing idea regarding prosodic strengthening is that it contributes to reinforcing the contrastive attributes of the segments in strong positions [10, for a review]. For instance, studies in Korean have shown that the three-way contrast between fortis, aspirated and non-fortis stops is enhanced in strong prosodic position, with a longer VOT for the non-fortis and aspirated stops but not for the fortis [11, 12, 13]. As far as fricatives are concerned, only a few studies have looked at prosodic position effects, and the results are inconsistent. Kim [14] reported a higher centroid frequency (hereafter, CoG) for the fortis fricative in higher prosodic domains, but results were inconsistent across subjects. On the other hand, Jang [15] found the reverse results: fricatives in stronger prosodic position had a lower CoG value. Since the expected effects of prosodic position, in terms of duration, glottal opening or oral articulation, are targeting phonetic features at play in the phonetic contrast between /s*/ and /s/, further data on this topic are awaited. Regarding our second factor of variation, vowel-dependent variation has also been discussed concerning the aspiration-related cues which can be found at the onset of the post-fricative vowel. Larger differences between /s/ and /s*/ have been found before low vowels than high vowels [8, 16]. Here, we will look at the effect of two anticipatory phenomena which should affect the frequency of the fricative noise: anticipatory labialization in the context of /u/ and the well-known palatalization of /s/ and /s*/ in the context of /i/.

Eight Female native Seoul Koreans aged from 22 to 35 y.o were recorded while producing tests sentences containing /V_1#CV_2/ target sequence with V_1=/a/, C=/s/ or /s*/ and V_2 =/i/ or /ɯ/ or /u/, and # =prosodic boundary. The CV2 target sequence was either positioned at the beginning of an IP (IPi), or phrase-internally at a Word boundary (Wi). Measurements include the durations of V_1, C, V_2, and measurements of the frication noise at three time point (first, mid and last 25.6 window) in the fricative: the four spectral moment, frequency of the peak amplitude in the mid frequency (3-7KHz) and its drop towards the end of the fricative, dynamic amplitude (amplitude differences between the mid frequency band and lower or higher ones) [17].

Preliminary results show the expected longer duration and higher CoG for /s*/ vs. /s/. No difference in the other 3 spectral moments are observed, but interesting results are shown by the mid frequency peak amplitude: higher for the /s*/ at the onset, middle et last portions of the fricative, but with an increased distinction between the two fricatives toward the end of the fricative. This is due to a larger drop of this peak at the fricative offset for /s/, while the friction remains more stable over time for /s*/.

Dynamic amplitude, which intends to capture the degree of sibilance in the fricative, is also higher in /s*/ throughout the three time points.

As expected, prosodic boundary effects were found to increase the duration of both fricatives in IPi position. Interestingly, the CoG of both fricatives was found to lowered in IPi compared to Wi (as found by [15]), with interaction with consonant type in the mid portion of the fricative. Indeed, the lowering of CoG in IP is greater for /s*/, than for /s/, showing an unexpected reduction of the CoG contrast in IPi.

CoG values of the frication noise are also dependent on V2 contexts, with a lowering of CoG in the palatalized (/i/) and rounded (/u/) contexts. Interestingly, this effect of V2 depends on the consonant type, with more V_2-induced noise frequency lowering for the fortis fricative. The measure of peak frequency in the mid frequency band (3-7 Khz) is particularly sensitive to the increase of front cavity length due to the anticipatory rounding of /u/. This lowering of peak frequency from the middle part of the fricative to its end, neutralize this acoustic distinction between /s*/ and /s/.

Taken together, these results show that the contrast between /s*/ and /s/ tend to be more reliably marked by acoustic duration, since both prosodic and contextual influences tend to reduce the spectral distinctions between the fricatives' noise. Speaker specific ways of marking the acoustic contrast between /s/ and /s*/ will also be discussed at the conference.

(1a) IP-initial /V_1 # CV_2/, the target word is in bold.
[mina]$_{IP}$ [hanbʌni aniɾa]$_{IP}$ # [**s*utaɾ**ɯl yʌl pʌn s͈*ɯsejo]$_{IP}$. (7th of the 16 syllables in total)
Translation: Mina, not one time, write s*uta ten times.

(1b) Wd-initial /V_1 CV_2/, the target word is in bold.
[hanbʌni aniɾa]$_{IP}$ [toŋsa **s*utaɾ**ɯl yʌl pʌn s͈*ɯsejo]$_{IP}$. (7th of the 16 syllables in total)
Translation: Not one time, write the verb s*uta ten times.

References

[1] Iverson, G. K. (1983). KOREAN-S. *Journal of Phonetics*, *11*(2), 191-200.
[2] Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of phonetics*, *30*(2), 193-228.
[3] Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of phonetics*, *2*, 161-180.
[4] Park, H. (1999). The phonetic nature of the phonological contrast between the lenis and fortis fricatives in Korean. In *Proceedings of the 14th international congress of phonetic sciences* (Vol. 1, pp. 424-427).
[5] Yoon, K. (2002). A production and perception experiment of Korean alveolar fricatives. *Umseng Kwahak [Speech Sciences]*, *9*(3), 169-184.
[6] Jun, S. A., Beckman, M. E., & Lee, H. J. (1998). Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations for Korean obstruents. *UCLA Working Papers in Phonetics*, 43-68.
[7] Kang, K. S. (2000). On Korean fricatives. *Umseng Kwahak [Speech Sciences]*, *7*(3), 67-82.
[8] Chang, C. B. (2013). The production and perception of coronal fricatives in Seoul Korean: The case for a fourth laryngeal category. *Korean Linguistics*, *15*(1), 7-49.
[9] Kim, H., Maeda, S., & Honda, K. (2011). The laryngeal characterization of Korean fricatives: Stroboscopic cine-MRI data. *Journal of Phonetics*, *39*(4), 626-641.
[10] Cho, T. (2011). Laboratory phonology. *The continuum companion to phonology*, 343-368.
[11] Cho, T., & Jun, S. A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA working papers in phonetics*, 57-70.
[12] Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean.
[13] Georgeton, L., & Fougeron, C. (2014). Domain-initial strengthening on French vowels and phonological contrasts: Evidence from lip articulation and spectral variation. *Journal of Phonetics*, *44*, 83-95.
[14] Kim, S. (2001). *The interaction between prosodic domain and segmental properties: domain initial strengthening of fricatives and Post Obstruent Tensing rule in Korean* (Doctoral dissertation, University of California, Los Angeles).
[15] Jang, M. (2011). Prosodically driven phonetic properties in the production of Korean fricatives. *음성음운형태론연구*, *17*(1), 65-85.
[16] Kang, Y., Kochetov, A., & Go, D. (2009, August). The acoustics of Korean fricatives. In *CRC-Sponsored Summer Phonetics/Phonology Workshop, University of Toronto*.
[17] Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of/s: Results from adolescents. *Journal of Speech, Language, and Hearing Research*.

Predicting Second-language Vowel Perception Based on Acoustic Similarities

Payam Ghaffarvand Mokari[1], Stefan Werner[2] & Daniel Williams[3]

[1,2]*University of Eastern Finland (Finland),* [3]*University of Potsdam (Germany)*
p.ghaffarvand@gmail.com, stefan.werner@uef.fi, daniel.williams@uni-potsdam.de

Over the past few decades, various models have been developed to explain how the native language (L1) phonological system affects second-language (L2) perception and production. The Second Language Linguistic Perception model (L2LP) [1] proposes that acoustics can successfully predict L2 sound learning. L2LP claims that any acoustic variation in native and target vowel production can influence speech perception [2]. L2LP further suggests that listener's initial perception of L2 vowels should closely match the acoustic properties of vowels as they are produced in the listener's L1 [1]. L2LP proposes that L2 categorisation patterns and discrimination difficulties can be predicted through a detailed comparison of the acoustic similarity between the sounds of the L1 and L2.

Some studies have reported that results of cross-language discriminant analyses (LDAs) trained on the acoustic properties of native vowels and tested on those of non-native vowels follow trends seen in listeners' perceptual assimilation patterns [3]. The present study provides new evidence with respect to assessing the effectiveness of LDAs for predicting non-native perceptual assimilation patterns. The eleven Standard Southern British English (SSBE) monophthongs are categorized into the nine Azerbaijani (AZ) vowel categories based on their acoustic properties. Subsequently, the cross-language categorization patterns are compared to AZ listeners' perceptual assimilation of SSBE vowels as previously reported in [4]. The AZ data used in the present study consists vowel productions by 23 female subjects aged 20-35 years old. The SSBE stimuli consisted of the productions of SSBE vowels by two female native speakers. The SSBE data is the same as stimuli used in the perceptual assimilation test in [4].

Figure 1 shows the F1 × F2 vowel space of AZ and SSBE based on means across speakers. It can be seen in Figure 1 that the vowel systems of the two languages are very different. The SSBE vowel space is more dispersed compared to that of AZ. Visual inspection suggests that there are AZ and SSBE vowels that are very similar based on F1 × F2 characteristics. For instance, AZ /y/ and /ɑ/ seem to be very close to SSBE /u/ and /ɑ/, respectively.

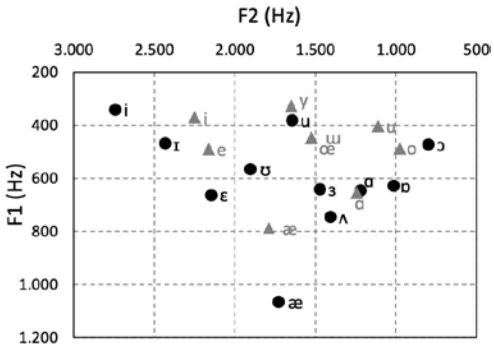


Fig 1. F1 × F2 average values (Hz) for AZ and SSBE vowels based on production of female speakers. Black dots show the SSBE vowels and gray rectangles show AZ vowels.

We first conducted separate LDAs for each language using the F1, F2 (in Bark) tokens. A cross-language LDA was conducted to establish the spectral similarity of SSBE and AZ vowels. The tokens of the SSBE corpus served as the test corpus and were classified with respect to AZ vowel centers of gravity established in the training model. In this way, the LDA classification results (percentage classifications) show how well SSBE vowels fall with respect to the centers of gravity of the AZ corpus tokens. The SSBE vowel category which receives the highest classification percentage for a given AZ vowel indicates that this SSBE vowel is the non-native vowel acoustically closest to the AZ vowel category. Table 1 shows the LDA results for SSBE vowels classified in terms of AZ vowel categories.

Table 1. Average probability scores of predicted group membership for SSBE tokens tested on a AZ model. The AZ vowel category with the highest predicted probability appears in bold and those probabilities below 0.20 appear in grey.

SSBE vowels	AZ vowels								
	æ	ɑ	e	i	ɯ	o	œ	u	y
æ	**1.00**	-	-	-	-	-	-	-	-
ɑ	-	**0.83**	-	-	-	0.16	-	0.01	-
ɪ	-	-	**1.00**	-	-	-	-	-	-
i	-	-	**1.00**	-	-	-	-	-	-
ɔ	-	-	-	-	-	**0.92**	-	0.08	-
ɒ	-	**0.61**	-	-	-	0.38	-	0.01	-
ɛ	0.28	-	**0.72**	-	-	0.00	-	-	-
ɜ	**0.49**	0.27	-	-	0.02	0.14	0.05	0.03	-
ʊ	-	-	**0.73**	-	0.09	-	0.17	-	-
ʌ	0.47	**0.53**	-	-	-	-	-	-	-
u	-	-	0.01	-	0.26	-	0.16	0.01	**0.55**

Based on the results shown in Table 1, SSBE /æ/, /ɑ/ /ɪ/, /i/ and /ɔ/ are each acoustically similar to one AZ vowelÍ with probability scores over 80%. However, other SSBE vowels were acoustically similar to a more than one AZ vowel over different ranges of probability scores. There are cases of similarity overlap where some SSBE vowels are acoustically similar to one AZ category. For instance, /ɪ/ and /i/ vowels both were strongly acoustically similar to AZ /e/. Similarly, SSBE /ʊ/ and /ɛ/ also had the highest probability scores to be categorized as AZ /e/. The most probable categorization of vowels in the three-way contrast of /ɑ-ɒ-ʌ/ is the AZ /ɑ/ category.

To quantify predictions of L2 vowel discrimination difficulty, following [5], we calculated *"cross-language assimilation overlap"* scores. This method gives a score of overlap between each member of L2 contrast and L1 vowels (Table 2).

Table 2. Acoustic overlap scores for AZ listeners.

SSBE contrasts	Overlap	SSBE contrasts	Overlap
/i-ɪ/	100	/ɑ-ɔ/	17
/ʊ-u/	26	/i-ɛ/	72
/ɑ-ʌ/	53	/ɔ-ɒ/	39
/æ-ɛ/	28	/æ-ʌ/	47

The categorization patterns based on LDA were largely in line with the previously found perceptual assimilation patterns [4]. However, our findings show that the LDA cross-language classifications solely based on F1-F2 information of vowels could not accurately predict perceptual assimilation patterns involving all vowels contrasts. Future studies may consider inclusion of other factors like vowel duration, F0, F3 and formant trajectories in the LDAs for more accurate comparisons with perceptual assimilations.

References

[1] Escudero, P. 2005. Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization. PhD thesis, LOT Dissertation Series 113. Utrecht University.

[2] Williams, D., Escudero, P. 2014. Influences of listeners' native and other dialects on cross– language vowel perception. *Frontiers in Psychology 5*, 1065

[3] Gilichinskaya, Y. D., Strange, W. 2010. Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the Acoustical Society of America 128*(2), EL80– EL85.

[4] Ghaffarvand Mokari, P. Werner, S. 2015. Perceptual assimilation of standard southern British English vowels by native Azerbaijanian listeners. In *Proceedings of International Conference on Speech Sciences*, 171–172. Seoul.

[5] Levy, E. S. 2009. Language experience and consonantal context effects on perceptual assimilation of French vowels by American–English learners of French. *The Journal of the Acoustical Society of America 125*(2), 1138–1152.

Testing the Cue-Weighting Transfer Hypothesis with Dutch Listeners' Perception of English Lexical Stress

Annie Tremblay[1], Mirjam Broersma[2], Yuyu Zeng[1], Anna Aumeistere[2],
Jinmyung Lee[1], Hyoju Kim[1], & Seulgi Shin[1]

[1]University of Kansas (USA), [2]Radboud University (Nijmegen, The Netherlands)
atrembla@ku.edu, m.broersma@let.ru.nl, yzengae@ku.edu, aumeistereanna@gmail.com,
jmlee@ku.edu, kimhj@ku.edu, seulgi.shin@ku.edu

This study investigates how listeners' knowledge of acoustic cues to lexical stress in the native language (L1) modulates their perception of lexical stress in a second language (L2), providing a further test of the cue-weighting transfer hypothesis for lexical stress.

Languages that have lexical stress differ in how stress is realized acoustically. To illustrate, whereas stressed syllables (with an intonational pitch accent) have a higher pitch, longer duration, and higher intensity than unstressed syllables (*ceteris paribus*) in both English and Dutch, unstressed vowels show a greater degree of centralization in English than in Dutch [1]. These acoustic differences create perceptual biases in the L1 that may in turn affect listeners' perception of lexical stress in the L2. For example, previous research has shown that English listeners rely more strongly on vowel quality than on pitch, duration, or intensity when perceiving English stress [2,3], whereas Dutch listeners rely more on duration than on vowel quality when perceiving Dutch stress [4]; these perceptual biases have been hypothesized to result in Dutch listeners' greater use of suprasegmental cues to English stress in spoken word recognition compared to English listeners [5], although this has not been tested explicitly.

Research to date, however, does not provide strong evidence in support of the hypothesis that listeners' knowledge of acoustic cues to lexical stress in the L1 determines their perception of lexical stress in the L2. Zhang and Francis [3] tested English and Mandarin listeners' weighting of acoustic cues to English stress. They found that, like English listeners, Mandarin listeners' English stress perception relied most heavily on vowel quality cues, even if lexical stress in Mandarin is signaled primarily by duration cues. Similarly, Chrabaszcz et al. [2], who tested English, Mandarin, and Russian listeners, found that all three groups' stress perception relied primarily on vowel quality cues (the authors had predicted Russian listeners to attend primarily to duration cues). In both studies, L1 effects emerged only in listeners' use of secondary cues to English stress. Crucially, the non-native listeners in these two studies were tested in the United States, raising the possibility that their proficiency in English was too advanced for them to show a strong cue-weighting transfer from the L1 to the L2.

The present study provides another test of the cue-weighting transfer hypothesis for lexical stress, this time focusing on non-native English listeners who lived in an environment where their L1 was spoken: native Dutch listeners in The Netherlands. In addition to shedding light on whether or not listeners can transfer their relative reliance on acoustic cues to lexical stress from the L1 to the L2, this study seeks to determine whether the hypothesis formulated in previous studies with Dutch L2 learners of English (e.g., [5]) can be substantiated.

Native English listeners (*n*=13; data collection ongoing) and native Dutch listeners (*n*=40) completed a cue-weighting stress perception experiment in English. In each trial, listeners heard an auditory stimulus and identified it as *DEsert* (word-initial stress) or *deSSERT* (word-final stress). The auditory stimuli were manipulated in seven acoustically equidistant steps from word-initial stress (Step 1) to word-final stress (Step 7), orthogonally manipulating two dimensions at a time (i.e., pitch [i.e., fundamental frequency] and vowel quality [i.e., first and second formants], duration and vowel quality, pitch and duration) while neutralizing the remaining dimensions at Step 4 (e.g., when pitch and vowel quality were manipulated orthogonally in seven steps,

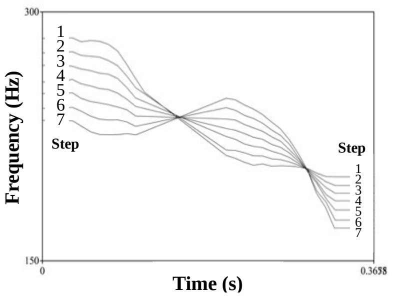


Fig. 1. Pitch Manipulation

duration and intensity were neutralized at Step 4). This was done by resynthesizing both the first and second syllables of the naturally produced disyllabic words, as illustrated in Figure 1 for the manipulation of pitch. The experiment included 147 different auditory tokens, each heard three times in three separate blocks.

Figure 2 shows English and Dutch listeners' proportions of *DEsert* selection for the vowel quality by pitch, vowel quality by duration, and duration by pitch manipulations. Participants' proportions of *DEsert* selection were analyzed with logit mixed-effects models; only significant results are discussed. When the stimuli differed in vowel quality and pitch, both groups used both cues, but English listeners relied more on vowel quality than Dutch listeners and Dutch listeners relied more on pitch than English listeners, with English listeners relying more on vowel quality than on pitch and with Dutch listeners showing similar reliance on both cues. Unlike English listeners, Dutch listeners also showed a greater reliance on pitch at lower (i.e., more *DEsert*-like) steps of the vowel reduction cue and a greater reliance on vowel reduction at lower (i.e., more *DEsert*-like) steps of the pitch cue. When the stimuli differed in vowel quality and duration, both groups used both cues, but English listeners relied more on vowel quality than Dutch listeners, and both groups relied more on vowel quality than on duration. When the stimuli differed in duration and pitch, both groups used only pitch cues, with Dutch listeners relying more on pitch cues than English listeners. Finally, and importantly, the greater the English proficiency (as determined by LexTALE [6] scores), the larger the effect of vowel quality in both sets where this cue was manipulated; by contrast, English proficiency did not modulate the effect of pitch.

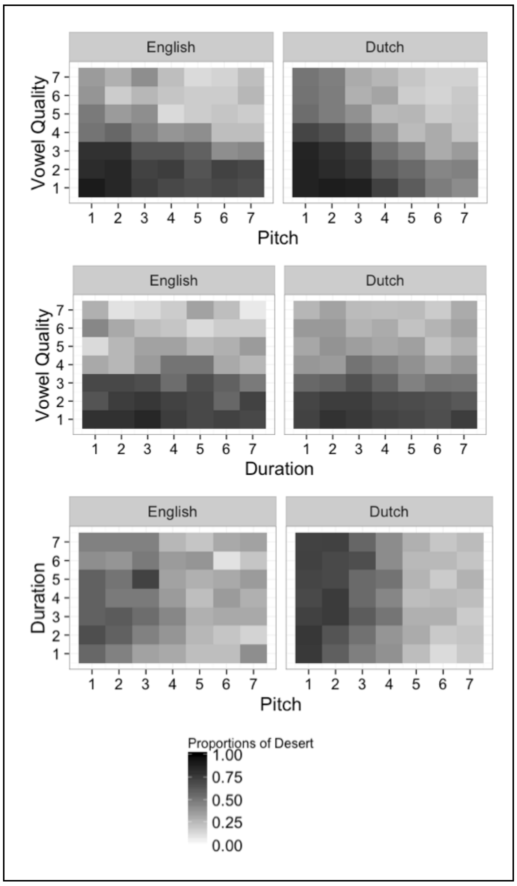


Fig. 2. Proportions of *DEsert* selection

These results indicate that the knowledge of cues to lexical stress in the L1 has an important effect on the perception of stress in the L2, supporting the cue-weighting transfer hypothesis for lexical stress. The results also confirm that Dutch listeners rely more on suprasegmental cues to English stress than English listeners, in line with the hypothesis formulated in previous studies (e.g., [5]). Last but not least, the current findings suggest that listeners' L2 cue-weighting can become more native-like with increased L2 proficiency.

References

[1] Sluijter, A. M. C., & van Heuven, V. J. (1996). 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.

[2] Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research, 57*, 1468-1479.

[3] Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics, 38*, 260-271.

[4] van Heuven, V. J., & de Jonge, M. (2011). Spectral and temporal reduction as stress cues in Dutch. *Phonetica, 68*(3), 120-132.

[5] 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.

[6] Lemhofer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behavioral Research Methods, 44*, 325-343.

Tonal Processing in Mandarin Reduplication: Morphological and Lexical Effects

Feier Gao[1], Siqi Lyu[2] & Chien-Jer Charles Lin[1]

[1]Indiana University Bloomington (USA), [2]Beihang University, Beijing (China)
gaof@iu.edu, siqilv@buaa.edu.cn, chiclin@indiana.edu

Introduction. Prosodic alternations such as tonal changes create a challenge to language processing, as language users need to sort through deviant surface forms to arrive at the correct underlying representation. In Mandarin, Tone 3 (T3) sandhi is a well-known example of grammaticalized tonal alternation where a low tone T3 obligatorily becomes a rising tone (T2) when it is followed by another T3. Therefore, T3 sandhi produces a surface sequence of T2-T3 that is underlyingly T3-T3. Such a process is both grammaticalized (therefore productive) and lexicalized. Recent studies found that Chinese words involving tone sandhi are more effortful to process than non-sandhi sequences and that the underlying T3 of the sandhi syllable is activated even though it surfaces as a T2 (Chien et al. 2016, Zhang et al. 2015).

The current study. Research on sandhi processing has so far focused exclusively on words. The current study investigates the processing and representation of sandhi tones that are derived from morphological processes such as reduplication. Two types of T3-T3 sequences that involve T3 sandhi are compared: verbal reduplications like *jiang(T3)-jiang(T3)* 'speak for a little bit', which add diminished meaning to its base *jiang(T3)* 'speak' (Basciano & Melloni 2017; Packard 1997, 2000) and existing disyllabic words like *li(T3)-jie(T3)* 'to understand'. Note that the morphological reduplication produces an intermediate representation of T2-T3, which further undergoes tone deletion on the second syllable, leading to a surface output of T2-T0 (e.g., *jiang(T2)-jiang(T0)*) while lexicalized sandhi simply surfaces as T2-T3 (e.g., *li(T2)-jie(T3)*). Our goal is to untangle whether and how the underlying tone (T3) in these sandhi syllables is accessible and to look at the interactions between morphological and phonological processes. Table 1 illustrates the three Mandarin constructions being contrasted: (A) reduplication with a T2 stem, which serves as a baseline that involves tonal deletion on the second syllable but not sandhi, (B) reduplication with a T3 stem which involves both sandhi and tone deletion, and (C) lexicalized T3 sequence, which involves sandhi but not tone deletion.

Research method. Thirty native speakers of the northern variety of Standard Mandarin (19-36 years old) participated in a cross-modal syllable-morpheme matching experiment. Target words were embedded in carrier sentences at the sentence-final position, and sentences were visually presented 2 characters at a time (equivalent to the size of disyllabic words in Chinese) on a computer monitor in a rapid serial visual presentation mode (programmed using E-Prime 2.0). A monosyllabic audio was played at the onset of the target word, and participants were asked to judge whether the audio matches the red-coded character of the target word (i.e., the bold-faced red-coded syllables in Table 1) by pressing the yes-no keys on a response box. For each of the three conditions in Table 1, three types of audio stimuli minimally contrasted by tones were used, including T2, which is the underlying and surface tone of the non-sandhi syllable in A and the surface tone of the sandhi syllable in B and C, T3, which is the underlying tone of the sandhi conditions in B and C, and a control tone T1 (the high-level tone in Mandarin), which is not related to the surface or the underlying tones of the stimuli.

Results. Both reaction times (RT) and response types (YES or NO) were analyzed. The percentages of saying yes to the underlying tones (A: T2, B: T3, C: T3) were higher than those of the surface tones (B: T2, C: T2) (Figure 1). Figure 2 showed longer RTs for saying yes to B than A (sandhi vs non-sandhi), which corroborates the cost of accessing the underlying representation of a sandhi syllable. Regarding the responses to T2 (Figure 3), we analyze the Yes and No responses separately: 1) for the participants who responded YES to T2, condition B takes longer than A ($p<0.001$) and C ($p<0.05$); 2) for the no responses to T2, the RT of C was longer than B ($p<0.05$), and both longer than processing the T1s (baseline) within their own group (B: $p<0.01$, C: $p<0.001$).

Both findings suggest that the surface tone of a morphologically complex form like reduplication is less accessible and easier to reject.

Discussion. Matching the underlying tone with the target word is shown to be preferred in all three constructions, including both sandhi conditions B and C (reduplication vs non-reduplication), suggesting equal availability of the underlying representations in morphologically complex sequences and in words. The main difference was found in responses to the surface tone T2. For the participants who matched the target words with the surface tones (Figure 3-left the yes responses), for example, reduplicated sequences were more costly than the lexicalized sequences, suggesting that the reduplicative sandhi sequences take more effort to be parsed as a sandhi word, as opposed to the non-reduplicative sandhi sequences. On the other hand, the underlying phonological representations have a greater influence in reduplicated structure and serve as stronger competitors for the surface tones. Morphological reduplications enhance the underlying tone of the stem and make the surface tones easier to reject.

Conclusion. In our current study, we found a construction-specific effect in processing the Mandarin T3 sandhi words. Verbal reduplications, although being opaque and morphologically complex, retains stronger access to the underlying tone of the sandhi syllables than the non-reduplicative lexicalized constructions. We also show that phonological processes such as tone sandhi are active in both morphologically complex expressions and lexicalized items (i.e., words). This study shows that sandhi is an active process both at the lexical level and at the post-lexical level and that both surface and underlying forms remain in competition in sentence processing.

Table 1 Three Mandarin constructions and examples

Construction	Tonal representation	Sandhi	Deletion	Example
A	/T2-T2/→[T2-T0]	×	✓	***tán***-*tan* 'talk (for a little bit)'
B	/T3-T3/→[T2-T0]	✓	✓	***xiáng***-*xiang* 'think (for a little bit)'
C	/T3-T3/→[T2-T3]	✓	×	***lí***-*jiě* 'understand'

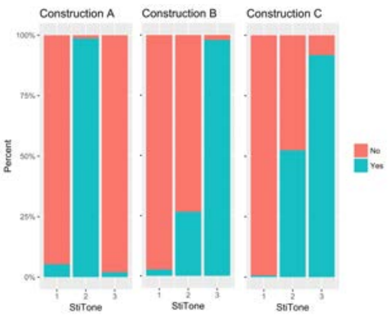


Fig. 1 Percentage of response types

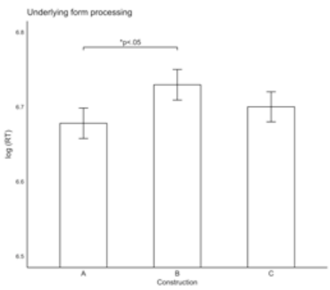


Fig. 2 RTs for saying yes to the underlying tone

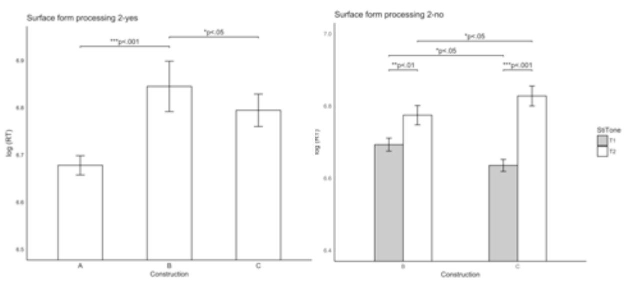


Fig. 3 RTs of responses to the surface T2-YES (left), NO (right)

References

[1] Chien, Y.-C., Sereno, J. A., & Zhang, J. (2016). Priming the representation of Mandarin tone 3 sandhi words. *Language, Cognition, and Neuroscience, 31*(2), 179–189.

[2] Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics*, 33 (2015), 149-162.

[3] Basciano, B., & Melloni, C. (2017). Event delimitation in Mandarin: The case of diminishing reduplication. *Rivista Di Linguistica*, *(Articolo su rivista)* vol. 29, 143-166.

[4] Packard, J. (1997). New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese. *Trends in Linguistics Studies and Monographs* 105. Berlin and New York: Mouton de Gruyter.

[5] Packard, J. (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge: Cambridge University Press.

Same parenthetical, different prosodic realization: evidence from German

Heather Weston

Leibniz-Centre General Linguistics (Germany)
weston@leibniz-zas.de

Parentheticals refer to material that is inserted in the middle of an utterance, such as *He is I think at home*. Their frequent use in spontaneous speech poses challenges to linguistic theory in different areas. An open question for syntacticians is how parentheticals are inserted into the surrounding material. It has long been assumed that parentheticals form their own phrase and thus intonational domain. However, for German, few studies examine prosodic data, and those that do report conflicting results: some studies conclude that parentheticals are always set apart prosodically in some way [1,2], usually with pitch changes, and others have found that parentheticals are often prosodically integrated, and many prosodic features are optional [3].

One challenge in previous work is the diverse nature of parentheticals, which can range from full sentences to subject–verb units such as *I believe*; many studies investigated different types of parentheticals. However, [3] found that shorter parentheticals are more likely to be prosodically integrated into surrounding material than longer ones. At the same time, [4] found that even the same short parenthetical can be realized with different prosodic phrasing in Icelandic. To address these contradictions, the current study investigates the prosodic nature of one short parenthetical, *glaube/denke ich* 'I think'. If the length of the parenthetical is decisive for prosodic behavior, characteristics should be similar across the dataset. However, if the same parenthetical is produced with different prosodic characteristics, this would suggest that speakers can choose whether or not to set apart the parenthetical and thus that they may make use of prosodic means to encode certain meanings. To control for possible speaker differences, the dataset includes multiple instances of the same parenthetical produced by the same speaker.

Secondary analysis was performed on a dataset of 180 instances of the parenthetical *glaube/denke ich* 'I think' taken from acoustic and respiratory recordings of 12 adult German speakers (7 female) in a laboratory setting. Spontaneous monologues were elicited via an item-choice task: participants were asked to choose 5 out of 10 items to take to a deserted island and motivate their choice (≈ 2 min./trial; 9 trials/participant). During transcription, the repeated use of the parenthetical 'I think' was noted and inspired the current data-driven analysis due to both the large number of instances and their formal homogeneity, which controls for the length and meaning of the parenthetical and makes it possible to compare the prosodic realization of the same parenthetical between and within speakers. The acoustic analysis software PRAAT [5] was used to visually inspect the pitch contours for pitch jumps and to measure intensity, speech rate and pauses immediately before, during and after the parenthetical, following [1].

The results to date show that the parenthetical *glaube/denke ich* is indeed realized with different prosodic characteristics across the dataset, also within speakers. For example, figure 1 shows the parenthetical *denke ich* seemingly integrated into the utterance 'and then I would I think still'. However, in another instance in the same trial (fig. 2; 'a book on plants I would I think also take with me'), the speaker prosodically sets off the parenthetical from the surrounding material by realizing it with lower pitch, lower intensity and creaky voice.

The fact that the same parenthetical 'I think' can have different prosodic profiles raises questions about whether a different meaning is being encoded. One striking and perhaps relevant result so far is that 'I think' often follows another subject–verb pair, like 'I would' or 'one can'. Here, the parenthetical could be removed without changing the grammaticality or truth-conditionality of the utterance. It is thus plausible that the parenthetical serves an optional function, such as weakening speaker commitment to the proposition expressed by the verb, similarly to a sentence adverbial, as proposed by [6]. In this case, the use of different prosodic characteristics could be evidence of cognitive encoding to distinguish shades of meaning.

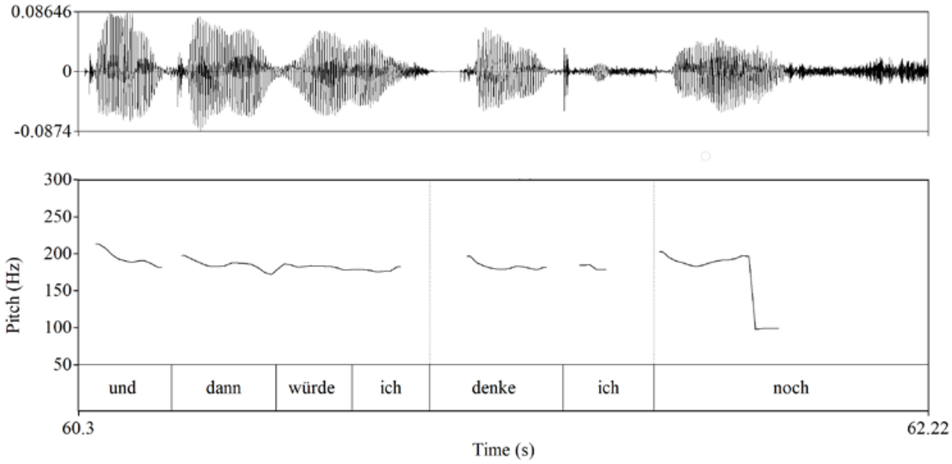


Fig.1 Pitch extract with denke ich 'I think' integrated in the surrounding material

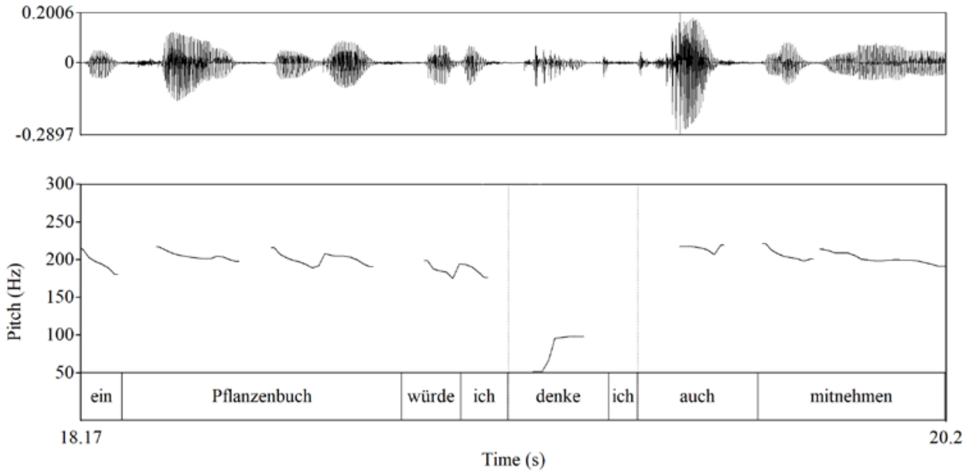


Fig.2 Pitch extract with denke ich 'I think' set off by lower pitch and intensity

References

[1] Bergmann, P. (2012). The prosodic design of parentheses in spontaneous speech. In P. Bergmann, J. Benning, M. Pfeiffer, & E. Reber (eds.), *Prosody and embodiment in interactional grammar* (pp. 103-141). Berlin/Boston: Walter de Gruyter.

[2] Döring, S. (2007). Quieter, faster, lower, and set off by pauses? Reflections on prosodic aspects of parenthetical constructions in modern German. In N. Dehé, & Y. Kavalova (eds.), *Parentheticals* (pp. 285-307). Amsterdam: John Benjamins**.**

[3] Peters, J. (2006). Syntactic and prosodic parenthesis. In R. Hoffmann, & H. Mixdorff (eds.), Proceedings from *Speech Prosody 2006, Third International Conference, Dresden, Germany, May 2-5, 2006* (paper 245). Available from https://www.isca-speech.org/archive/sp2006/sp06_245.html

[4] Dehé, N. (2018). The prosodic phrasing of parenthetical comment clauses in spontaneous spoken language: evidence from Icelandic *held ég*. *Studia Linguistica, 72*(1), 128-164.

[5] Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

[6] Thompson, S., & Mulac, A. (1991). A quantitative perspective on the grammaticization of epistemic parentheticals in English. In E. Traugott, & B. Heine (eds.), *Approaches to grammaticalization* (pp. 313-327). Amsterdam: John Benjamins.

The prosodic structure and the underlying tonal pattern of AP in Chungnam Korean

Sejin Oh[1,2] & Sun-Ah Jun[3]
[1]*The Graduate Center, CUNY,* [2]*Haskins Laboratories,* [3]*UCLA*
soh@gradcenter.cuny.edu, jun@humnet.ucla.edu

There are six major dialect groups spoken in South Korea: Seoul (including Gyeonggi), Gangwon, Chungcheong, Jeolla, Gyeongsang, and Jeju. Among these, Seoul Korean, the standard dialect of Korean, is the most studied dialect, including the prosodic structure and intonational phonology in the Autosegmental-Metrical (AM) framework [1, 2, 3, 4]. The intonational phonology of the Gyeongsang, Jeolla, and Jeju dialects has also been studied in the same theoretical framework (cf. Gyeongsang [5, 6]; Jeolla [1, 2, 7]; and Jeju [8, 9]). However, the intonation system of the Chungcheong and Gangwon dialects have rarely been described and have never been analyzed in the AM framework.

The goal of the current study is to propose a prosodic structure of Chungnam (South of Chungcheong) Korean defined by intonation in the AM framework, focusing on the tonal pattern of a small phrase, corresponding to the Accentual Phrase (AP) in other dialects of Korean. Ten native speakers of the Chungnam dialect from Seosan city participated in the experiment (4F, 6M; Mean Age: 55.3). The materials consisted of two sets of dialogs varying in word-initial segment types, 30 sentences with the second word varying in length from 3 to 7 moras, and natural conversation. To investigate the underlying tonal pattern of AP, F0 values were measured in the middle of each vowel, and the falling slopes from the F0 peak on the second mora to the following F0 minimum in the second word were calculated. In addition, the pitch tracks of all utterances were analyzed to determine the tonal categories and the prosodic structure defined by a tone, following the ToBI conventions [10] as well as the Korean-ToBI convention [11].

The results show that the prosodic structure of Chungnam dialect is very similar to that of the revised model of Seoul dialect [3, 4, 12] and that of South Jeolla (henceforth, Chonnam) dialect [7] by having three prosodic units higher than a word: an Intonational Phrase (IP) > an Intermediate Phrase (ip) > an Accentual Phrase (AP) (see Fig.1). An ip is marked by a phrase final boundary tone (L- as in Fig.2(a) or H- as in Fig.2(b)) on its last mora, which is slightly lengthened. In both figures in Fig.2, the peak of the AP after the ip boundary is higher than that before the ip boundary, indicating that an ip is the domain of pitch reset. An IP is marked by a final boundary tone realized on an IP-final mora, which is substantially lengthened. We found five types of IP-final boundary tones: (from the most frequent to the least: L% > LHL% > LH% > HL% > H%). As in Seoul Korean, the boundary tone of a higher prosodic unit overrides that of a lower prosodic unit. Fig. 3a shows an example pitch track of a three-word sentence produced in three APs forming one ip/IP.

The results also show that, as in Seoul and Chonnam dialects, the AP-initial tone of Chungnam dialect is sensitive to the laryngeal feature of the AP-initial segment (i.e., H when the AP-initial segment is either aspirated or tense, but L otherwise), but the overall tonal pattern of AP in Chungnam is the same as that of Chonnam, i.e., T-H-L (T=H or L), with a mora being a tone bearing unit (see Fig.4). However, the falling slope from the H to the following L is shallower as the number of moras increases in an AP (see Fig.5), suggesting that the L tone is not associated with the third mora of the AP, as observed for Chonnam speakers in [2]. Instead, the location of min F0 is found either on the AP-final (Fig.6a) or penult mora (and stayed low until the end of the AP, see Fig.6b). Therefore, we posit that the underlying tonal pattern of AP is THLL, where the initial two tones are associated with the AP-initial two moras and the final two L tones are associated with the AP-final two moras. But the penult L is often phonetically undershot when the AP is short or produced fast, generating variability across speakers. However, as shown in Fig. 3b, the penult L is clearly realized when the AP is longer than 3 moras and when the AP-final mora carries a boundary tone of a higher prosodic unit that begins with an H (e.g., H- or H%, HL%).

In sum, the prosodic structure of Chungnam Korean is similar to that of Seoul or Chonnam Korean: IP > ip > AP > Word. However, the underlying tonal pattern of AP differs across these three dialects. It is THLH in Seoul, THL in Chonnam, but THLL in Chungnam Korean.

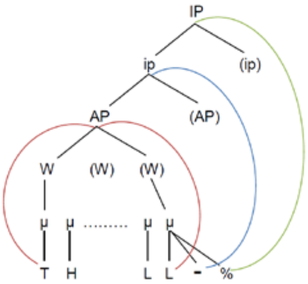


Fig. 1 Prosodic hierarchy and tonal affiliations of the Chungnam dialect. (T=H or L)

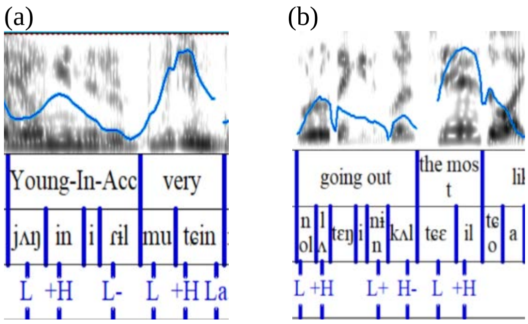


Fig. 2 Ex. of ip boundary tone: (a) L-, (b) H-

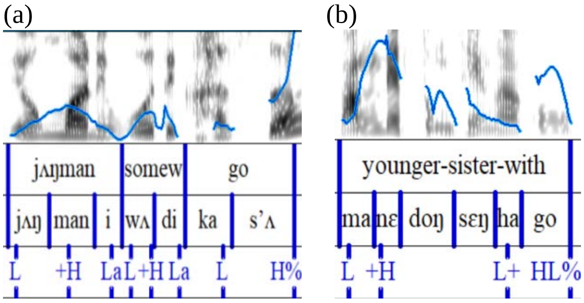


Fig. 3 Overridden by a higher prosodic unit

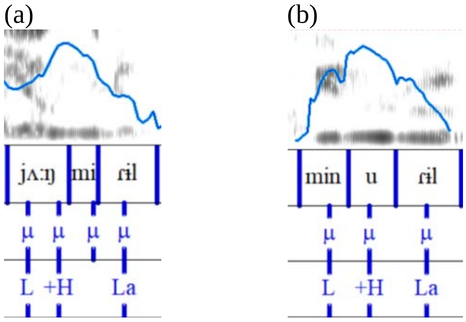


Fig. 4 Association of two AP-initial tones: (a) 1st syllable with 2 moras; (b) with 1 mora.

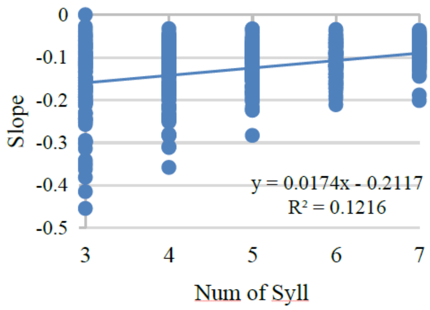


Fig. 5: A falling slope (from F0 peak to min)

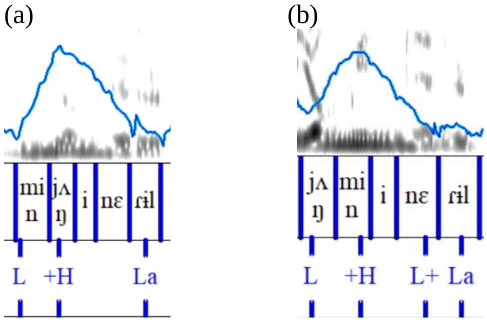


Fig. 6: AP patterns: (a)[LHL] vs. (b)[LHLL]

References

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

[2] Jun, S.-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy, *Phonology. 15*(2):189-226.

[3] Jun, S.-A. (2011). Prosodic markings of complex NP focus, syntax, and the pre-/post-focus string. In *Proceedings of the 28th WCCFL* (pp. 214-230). Somerville, MA: Cascadilla Press.

[4] Jun, S.-A. & Cha., J. (2015). High-toned [il] in Korean: Phonetics, Intonational Phonology, and Sound Change. *Journal of Phonetics 51*: 93-108.

[5] Jun, J., Kim J., Lee H., & Jun, S.-A. (2006). The Prosodic Structure and Pitch Accent of Northern Kyungsang Korean. *Journal of East Asian Linguistics 15*(4): 289-317

[6] Kim, J. & Jun, S.-A. (2009). Prosodic Structure and Focus Prosody of South Kyungsang Korean. *Language Research 45*(1): 43-66.

[7] Jun, S.-A. (1989). The Accentual Pattern and Prosody of Chonnam Dialect of Korean. in S. Kuno et al. (eds.) *Harvard Studies in Korean Linguistics. III.* pp. 89-100.

[8] Lee, S.-H. (1999). An acoustic study on the intonation pattern of Cheju dialect in Korean. *Proceedings of the Acoustical Society of Korea Conference*. *18*(2): 369-372.

[9] Lee, S.-H. (2014). The intonation patterns of accentual phrase in Jeju dialect. *Phonetics and Speech Sciences 6*(4): 117-123.

[10] Beckman, M., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press. pp. 9-54.

[11] Jun, S.-A. (2000) K-ToBI (Korean ToBI) labelling conventions: Version 3, *Speech Sciences*, 7, 143-169.

[12] Jun, S.-A. (2006). Intonational Phonology of Seoul Korean Revisited, *Japanese-Korean Linguistics 14*, Stanford: CSLI. pp.15-26.

L2 tone processing as revealed by the incidental learning of tone-segment mappings

Ricky Chan & Bruce Wang

Speech, Language and Cognition Laboratory, School of English, University of Hong Kong (Hong Kong)
rickykwc@hku.hk, brucexw@hku.hk

Background: Tone languages (e.g. Cantonese, Thai, Mandarin) employ contrastive pitch patterns (i.e. lexical tones) to distinguish word meaning. Most previous research on L2 tone acquisition has focused on explicit processes such as L2 tone discrimination/identification [e.g. 1,2,3], and whether learners' tonal background and prior musical training may facilitate L2 tone perception [e.g. 4,5]. However, the ability to identify/discriminate L2 tones does not entail the ability to form abstract tone categories at the syllable level, and using tone categories as lexical cues [cf. 6]. This contribution is motivated by the hypothesis that, for learners whose native language is non-tonal, a long-term difficulty in tone learning concerns repurposing pitch patterns from intonation cues to the formation of abstract tone categories at the syllable level. This hypothesis was tested with an experiment on the incidental learning of tone-segment mappings (constraints by segments on the tone a given word can carry). The learning of these mappings hinges on the encoding of pitch patterns as abstract tone categories at the syllable level.

Subjects: 80 subjects participated (20 Cantonese musicians, 20 Cantonese non-musicians, 20 English musicians, and 20 English non-musicians)[1].

Learning targets: i) Words beginning with an aspirated stop (e.g. /p^h/, /t^h/ or /k^h/) carry a rising tone; ii) words beginning with an approximant (e.g. /l/, /w/ or /j/) carry a falling tone.

Procedure: All stimuli were monosyllabic nonce words generated by the Salika speech synthesizer. Subjects first completed an AX discrimination task, which tested if they could distinguish the two target tones. The AB pairs differ only in the tone the words carry. Subjects then completed a word learning task adapted from [7,8]. In the training phase, they listened to a nonce word and repeated it aloud in each trial. This incidentally trained the subjects on the learning targets (i.e. they attended to the segments and tones but were not told their connections). In the testing phase, subjects were presented with two possible words and asked to decide which one sounded better. Subjects' learning of the target tone-segments connections was assessed by determining whether they could transfer the knowledge they acquired from the training to novel items in "critical trials" and "extension trials". Items in the critical trials had the same consonants used in the training along with new vowels, whereas those in the extension trials had a new aspirated stop/approximant onset. Sound pairs for the critical and extension items (differed only in terms of the tone they carried (e.g. /p^hɯːmR/ vs. / p^hɯːmF/). As such, it was when participants possessed knowledge of the target tone-segment mappings that they would show a preference for words which follow the target rules (e.g. /p^hɯːmR/ in the case above).

Results and Discussion: Figure 1 shows the average accuracy and log-reaction time (logRT) of the four groups for the AB pairs in the AX discrimination task. The four subject groups performed similarly well on distinguishing the two tones perceptually. Figure 2 shows the average accuracy of the four groups for the critical items and extension items. 95% confidence intervals were generated to determine whether individual group performed significantly above chance level (i.e. showing a learning effect). Results are presented in tables 1 and 2. For both critical items and extension items, the 95% confidence intervals of both Cantonese musicians and non-musicians do not contain the chance level value (50% accuracy), revealing that they performed significantly above chance. However, the 95% confidence intervals for English musicians and non-musicians contain the chance level value, meaning that they showed no learning effect. In sum, despite similar ability to distinguish tones perceptually, Cantonese learners could learn the target tone-segment connections but English learners could not, potentially because English learners failed to repurpose

[1] Subjects with six years or more formal musical training and have played music/sang regularly in the past two years were classified as musicians; those with less than 2 years of casual musical experience and have not played music/sang regularly in the past 2 years were classified as non-musicians.

pitch from intonational cues to forming tone categories at the syllable level [2]. Also, the fact that musical training did not facilitate the target incidental learning provides evidence for the separation of music and speech. More details on the statistical analyses will be presented in the conference.

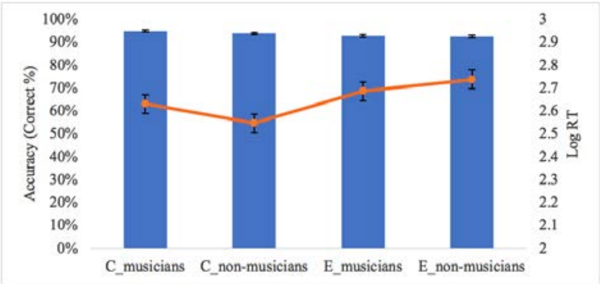


Fig.1 Accuracy and log-transformed reaction time of the four groups for the AB pairs in the AX discrimination task. (C=Cantonese, E=English; error bar = 1 SE)

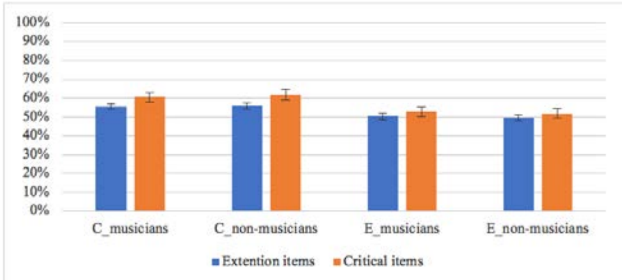


Fig. 2 Accuracy of the four groups for the critical items and extension items in the pronunciation judgment task. (C=Cantonese, E=English; error bar = 1 SE)

Group	Accuracy on critical items (%)	Upper Bound	Lower Bound
Cantonese musicians	60.80	65.17	56.26
Cantonese non-musicians	62.22	66.53	57.70
English musicians	52.99	57.56	48.37
English non-musicians	51.85	56.44	47.24

Group	Accuracy on extension items (%)	Upper Bound	Lower Bound
Cantonese musicians	55.65	59.89	51.33
Cantonese non-musicians	55.94	60.17	51.62
English musicians	50.28	54.60	45.96
English non-musicians	49.58	53.89	45.26

Tables 1 and 2. Accuracy (% correct) and 95% confidence intervals (from upper bound to lower bound) for the prediction of each participant group based on generalised linear mixed-effects models of the critical items (top) and extension items (bottom) (chance level = 50%).

References

[1] Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268-294.

[2] So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and speech*, *53*(2), 273-293.

[3] Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the acoustical society of America*, *106*(6), 3649-3658.

[4] Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*, 395–421.

[5] Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681-712.

[6] Sebastián-Gallés, N., & Díaz, B. (2012). First and second language speech perception: Graded learning. *Language Learning*, *62*(s2), 131–147.

[7] Chan, R. & Leung, J. (2014). Implicit learning of L2 word stress regularities. *Second Language Research*, *30*(4), 463-484.

[8] Chan, R. & Leung, J (2018). Implicit Knowledge of L2 Lexical Stress Rules: Evidence from the Combined Use of Subjective and Objective Awareness Measures. *Applied Psycholinguistics*, *39*(1), 37-66.

Prosodic boundary and prominence effects on vowel nasalization in Australian English versus American English

Hyunjung Joo[1], Jiyoung Jang[2], Sahyang Kim[3], Taehong Cho[1], & Anne Cutler[4]

[1]Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS), Hanyang University (Korea),
[2]University of California, Santa Barbara (USA), [3]Hongik University (Korea),
[4]MARCS Institute, ARC Centre of Excellence for the Dynamics of Language, Western Sydney University (Australia)
hyunjung.joo.ling@gmail.com, jiyoungljang@gmail.com, sahyang@gmail.com,
tcho@hanyang.ac.kr, a.cutler@westernsydney.edu.au

Coarticulation is an inevitable low-level phonetic process that underlies connected speech across languages, entailing cross-linguistic similarities in phonetic implementation [1]. It is, however, also known to be conditioned by higher-order linguistic structures, such as prosodic structure, in a language-specific way [2,3]. Recent studies [4,5] have shown that coarticulatory vowel nasalization is fine-tuned by different sources of prosodic strengthening – i.e., boundary and prominence, enhancing syntagmatic vs. paradigmatic linguistic contrasts respectively. The studies have also revealed cross-linguistic differences in the phonetics-prosody interface as reflected in coarticulatory V-nasalization, suggesting that the seemingly low-level phenomenon of V-nasalization is internalized in a language-specific phonetic grammar [8,9]. This further opens up a possibility of cross-dialectal variations in this process, which has not yet been investigated. The present study explores cross-dialectal similarities and differences in prosodically-conditioned coarticulatory V-nasalization by examining how prosodic boundary and prominence influence V-nasalization in carryover (#NVC) and anticipatory (#CVN) contexts in Australian English (AusE) and comparing the results with those found in American English (AmE) [4].

A recent study on AmE [4] collected the coarticulatory V-nasalization data from 15 native AmE speakers [4], and measured N duration and V-nasalization in #NVC and #CVN contexts in different prosodic boundary (phrase-initial/phrase-final vs. phrase-medial) and prominence (Phonological Focus, Lexical Focus, No Focus) conditions. Results showed that both N duration and V-nasalization decreased in the phrase-initial #NVC whereas they both increased in the phrase-final #CVN as compared to the phrase-medial #NVC/CVN# contexts. Under prominence (i.e., phonological and lexical focus), in both #CVN and #NVC, N duration increased but V-nasalization decreased, both of which contribute to the paradigmatic contrast enhancement by enhancing N's nasality and V's orality. Importantly, the coarticulatory resistance effect was pervasive throughout the vowel.

The present study employed the same experimental design used in [4] and collected data from 14 native speakers of AusE. Examples of test sentences are laid out in Table 1. N duration and A1-P0 values (V-nasality) were extracted using a Praat script [6]. A1-P0 values were obtained at two absolute timepoints (25ms, 50ms from N into the V) and at three relative timepoints (25%, 50%, 75% of the V) in order to see whether the coarticulatory process is a time-locked phenomenon (absolute measure) or it is under speakers' control (relative measure).

The comparison between AusE and AmE revealed interesting cross-dialectal differences and similarities. N duration in #NVC was generally longer in AusE than in AmE, but no such dialectal difference was observed for #CVN. Furthermore, compared to AmE, AusE showed more V-nasalization in the #NVC context (Fig.1g) but less V-nasalization in the #CVN context (Fig.1j); but this dialectal difference was observed only when the vowel's physical distance from N was exactly the same (i.e., absolute timepoints). The results also revealed remarkable cross-dialectal similarities in terms of the patterns of prosodic strengthening. Both dialects showed a boundary-induced shortening of N, accompanied by less V-nasalization in phrase-initial #NVC (Fig.1c and i), enhancing the syntagmatic CV contrast with the increased consonantality of N and orality of V. But the reverse was true for the phrase-final CVN#: N duration and V-nasalization both increased (Fig.1f and l), indicating the loosening of articulatory linkage between the oral constriction and the velum lowering gesture phrase-finally. The two dialects also showed a comparable prominence-induced effects, with lengthening of N and less V-nasalization under prominence in both #NVC and CVN# (Fig.1b,e,h k). This pattern can be interpretable as coarticulatory resistance with the paradigmatic contrast enhancement. It was also found that in both dialects, the boundary and prominence effects on V-nasalization were not the physiological/biomechanical time-locked effects, but pervasive throughout the vowel, indicating that the coarticulatory process is controlled by the speaker with reference to higher-order prosodic structure.

The cross-dialectal differences reflected in a non-contrastive phonetic process strongly suggest that the low-level phonetic coarticulation is regulated differently even across dialects of the same language [7,8,9].

The cross-dialectal similarities are grounded on linguistic contrasts that may underlie the phonetics-prosody interface. The results imply that although AusE and AmE differ in the magnitude of coarticulatory nasalization in carryover vs. anticipatory contexts, such seemingly different coarticulatory effects across dialects operate in much the same way by making reference to linguistic contrast enhancement in universally applicable ways.

Table 1. Examples of test sentences with a CVN word for each condition of boundary and focus. Target segments are underlined and focused words are in bold. # indicates different prosodic boundary.

	IP condition	Wd condition
Phon FOC	A: Were you supposed to write **BOB**? B: No. I was supposed to write **BOMB** #, wasn't I?	A: Did you write 'say **BOB** fast again'? B: No. I write 'say **BOMB** # fast again.'
Lex FOC	A: Were you supposed to write **WAR**? B: No. I was supposed to write **BOMB** #, wasn't I?	A: Did you write 'say **WAR** fast again'? B: No. I write 'say **BOMB** # fast again.'
No FOC	A: Were YOU supposed to write **bomb**? B: No. JOHN was supposed to write **bomb #**, wasn't he?	A: Did you write 'say **bomb** FAST again'? B: No. I write 'say **bomb** # SLOWLY again.'

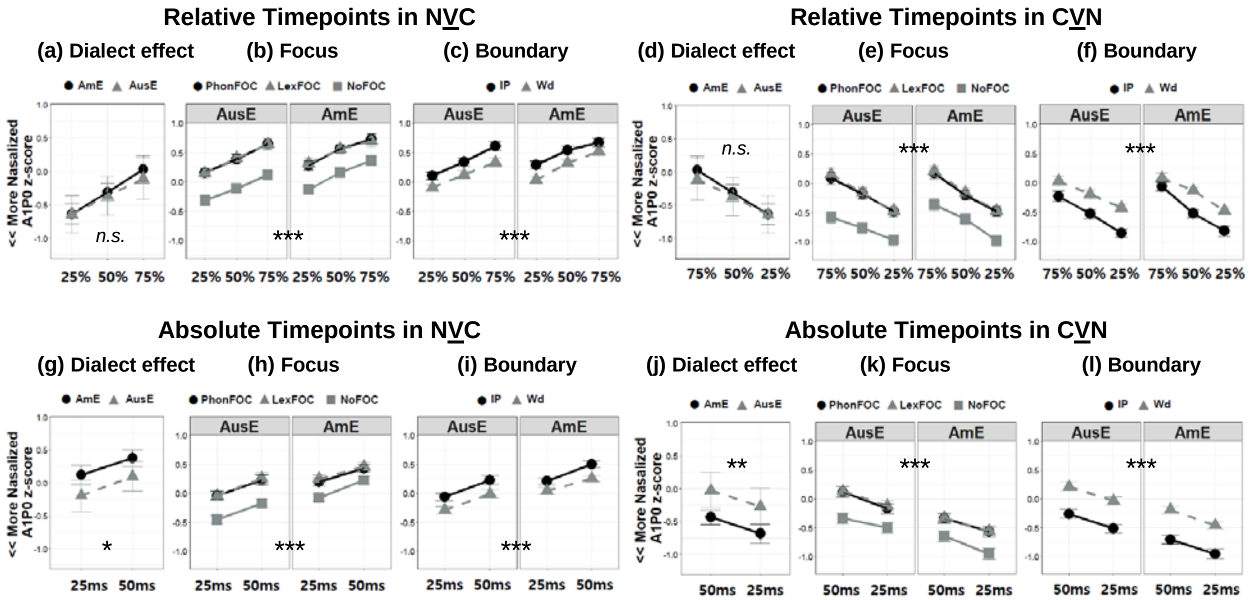


Fig 1. V-nasalization: Dialect, Focus and Boundary effects in #NVC and CVN# at relative and absolute timepoints in AusE and AmE. (AmE data from [4].) '***', '**' and '*' refer to <.001, <.01 and <.05, respectively.

References

[1] Kühnert, B., Nolan, F. (1999). *The origin of coarticulation*. In: Hardcastle, W., Hewlett, N. (eds), Coarticulation: Theory, Data and Techniques. Cambridge University Press, Cambridge, 7-30.

[2] Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass, 10*(3), 120-141.

[3] Kim, S., Kim, J., Cho, T. (2018). Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics*, *71*, 65-80.

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

[5] Jang, J., Kim, S., Cho, T. (2018). Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences. *The Journal of the Acoustical Society of America*, *144*(1), EL33-EL39.

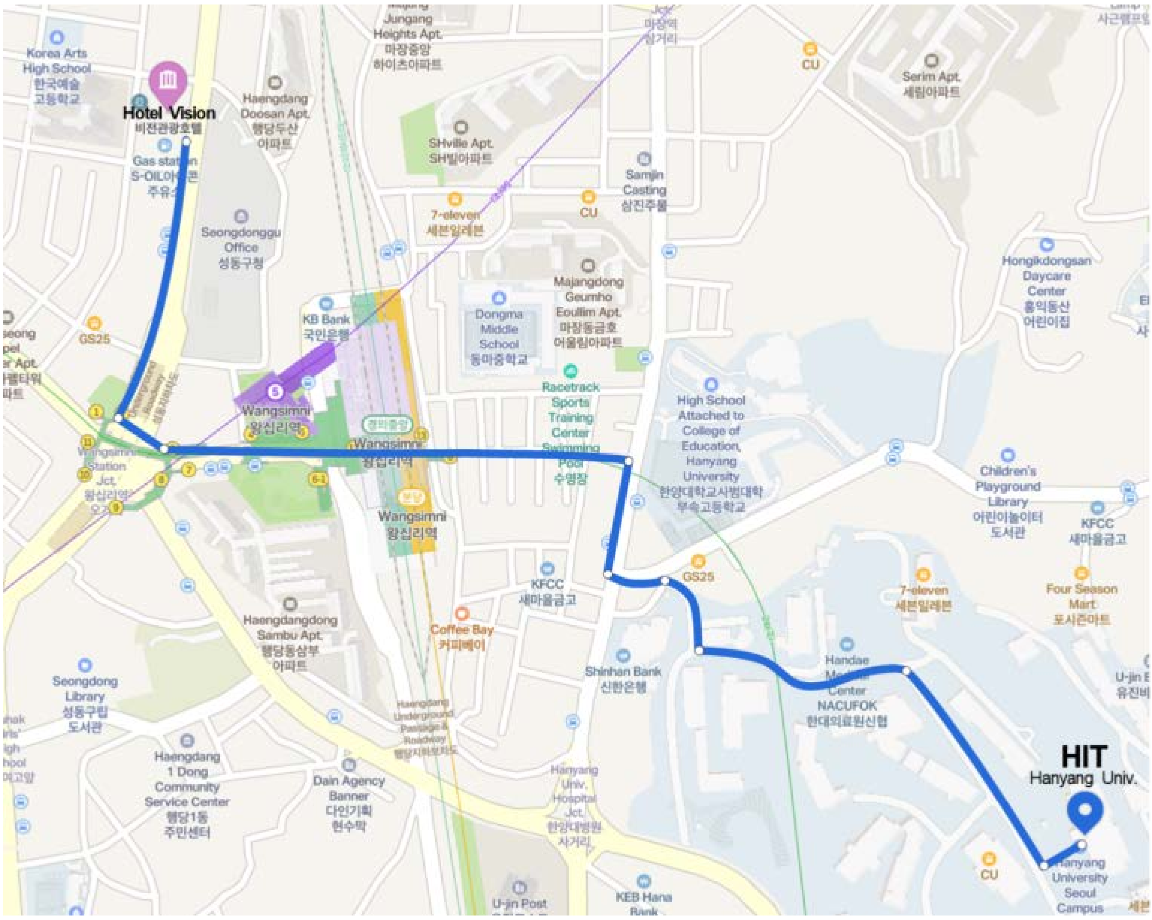
[6] Styler, W. (2015). *On the acoustical and perceptual features of vowel nasalization.* Ph.D. dissertation. University of Colorado.

[7] Cho, T. (2015). Language effects on timing at the segmental and suprasegmental levels. In: Redford, M. A. (ed), *The Handbook of Speech Production*. Hoboken, NJ: Wiley-Blackwell, 505-529.

[8] Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of phonetics, 18*(3), 321-334.

[9] Cho, T., Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics, 27*(2), 207-229.

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).

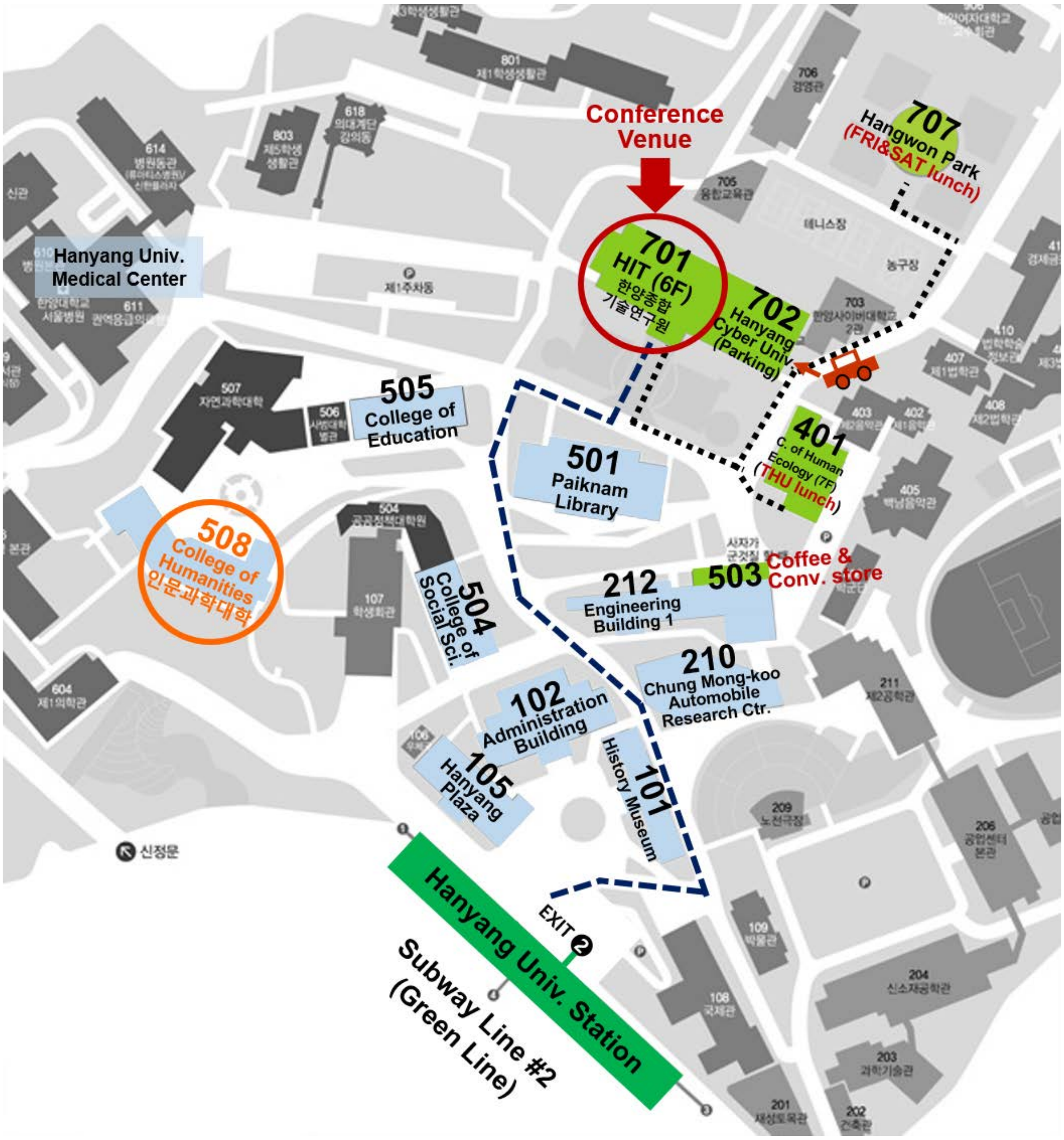
(2) Get off at the next station, Hanyang University Station (한양대역). Exit from Exit 2. Walk to HIT.

- **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.
- PARKING: You can purchase parking tickets at the registration desk (up to 12 hours/day). You may park at any 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 (workshop day, THU): College of Human Ecology (**Building #401**, 7^{th} level).
- LUNCH (main conference days, FRI&SAT): Hangwon Park (**Building #707**, B1 level)
- COFFEE & CONVENIENCE STORE: **Building #503**
- 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.

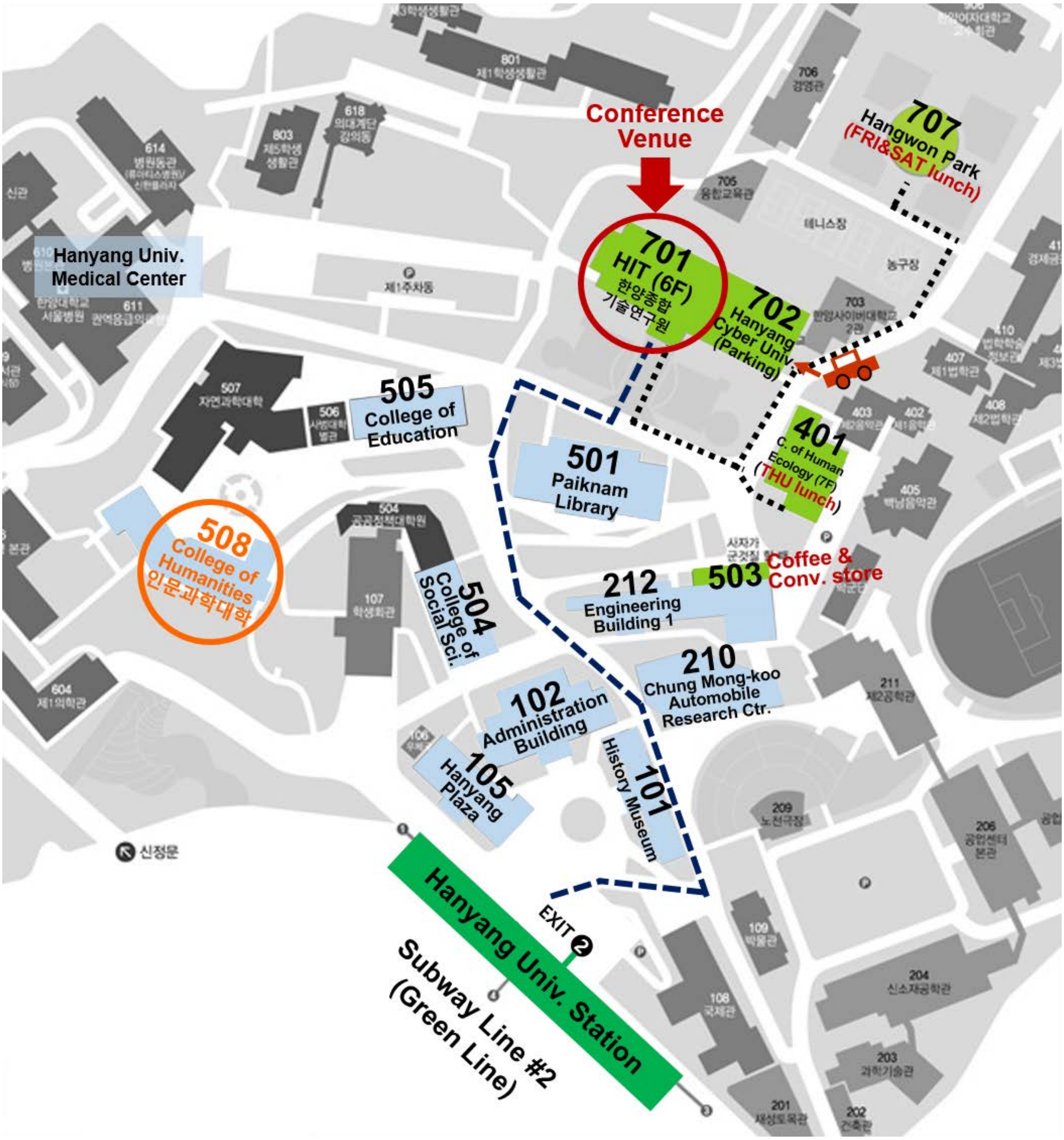
HISPhonCog 2019: Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2019

●

April 30, 2019.
Hanyang University, Seoul, Korea

	:	2019 4 30
	:	2019 4 30
	:	
	:	•

Printing Date	:	April 30, 2019
Publishing Date	:	April 30, 2019
Publisher	:	Cho, Taehong
Publishing Place	:	Hanyang Institute for Phonetics and Cognitive Sciences of Language



ISSN 2586-7997