ISBN 979-11-983259-0-7

2023 한양·음성언어인지과학 국제심포지엄

HISPhonCog 2023

Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2023

Department of English Language and Literature
Hanyang Institute for Phonetics & Cognitive Sciences of Language (HIPCS)
Hanyang Humanities Enhancement Center (H^2EC)

Linguistic and cognitive functions of fine phonetic detail underlying sound systems and/or sound change

May 26-27, 2023, Hanyang University, Seoul, Korea

Edited by Taehong Cho
Sahyang Kim
Say Young Kim
Sang-Im Lee-Kim

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

HIPCS
Hanyang Institute for Phonetics and Cognitive Sciences of Language

HANYANG UNIVERSITY

ISBN 979-11-983259-0-7

2023 한양·음성언어인지과학 국제심포지엄

HISPhonCog 2023

Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2023

Department of English Language and Literature
Hanyang Institute for Phonetics & Cognitive Sciences of Language (HIPCS)
Hanyang Humanities Enhancement Center (H^{2}EC)

Linguistic and cognitive functions of fine phonetic detail underlying sound systems and/or sound change

May 26-27, 2023, Hanyang University, Seoul, Korea

Edited by Taehong Cho
Sahyang Kim
Say Young Kim
Sang-Im Lee-Kim

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

HIPCS
Hanyang Institute for Phonetics and Cognitive Sciences of Language

HANYANG UNIVERSITY

PROGRAM AT A GLANCE

Day 0 (Thursday, May 25, 2023)	
Special Colloquium: 17:00-18:20	
17:00-18:20	Invited Speaker: Adam Ussishkin (University of Arizona) *Auditory and visual processing of root morphemes in Maltese*

Day 1 (Friday, May 26, 2023)	
08:00-08:50	**Registration** (Morning Coffee & Some Korean Rice Cakes and Donuts)
09:00-09:20	**Opening** (Taehong Cho, Chair) ▪ Hyunchul Chung (Executive Vice President): Welcoming Remarks ▪ Sungho Yoo (Dean, College of Humanities): Opening Remarks
Oral Session 1: 09:20-10:30 **Chair:** Say Young Kim (HIPCS, Hanyang University)	
09:20-10:00	▪ Invited Speaker: John Kingston (University of Massachusetts, Amherst)
10:00-10:30 2 talks	▪ **Holger Mitterer, Eva Reinisch** (U of Malta; Acoustics Research Institute, Austrian Academy of Sciences) ▪ **Chiung-Yu Chang, Alexandra Jesse, Lisa D. Sanders** (U of Massachusetts, Amherst)
10:30-11:00	Coffee Break
Oral Session 2: 11:00-12:25 **Chair:** Sang-Im Lee-Kim (HIPCS, Hanyang University)	
11:00-11:40	▪ Invited Speaker: Lisa Davidson (New York University)
11:40-12:25 3 talks	▪ **Yaqian Huang** (UCLA) ▪ **Marc Garellek, Jianjing Kuang, Osbel López-Francisco, Jonathan D. Amith** (UC, San Diego; U of Pennsylvania; Universidad Autónoma de México, Iztacala; Gettysburg College) ▪ **Rasmus Puggaard-Rode** (Ludwig Maximilian U of Munich)
12:25-13:45	Lunch (1hr 20m)
Oral Session 3: 13:45-15:10 **Chair:** Suyeon Yun (Chungnam University)	
13:45-14:25	▪ Invited Speaker: Adam Albright (MIT)
14:25-15:10 3 talks	▪ **Jennifer Kuo** (UCLA) ▪ **Carmen Kin Man Tang, Regine Lai** (The Chinese U of Hong Kong) ▪ **Bowei Shao, Anne Hermes, Philipp Buech, Maria Giavazzi** (Ecole normale supérieure, Université PSL; LPP (UMR7018, CNRS/Sorbonne Nouvelle))
Poster Session 1: 15:10-17:00 (48 posters: poster#p01-p48) (see below) **Chair:** Jiyoun Choi (Sookmyung Women's University)	
Oral Session 4: 17:00-18:10 **Chair:** Jeff Holliday (Korea University)	
17:00-17:30 2 talks	▪ **Simona Sbranna, Aviad Albert, Martine Grice** (U of Cologne) ▪ **Richard Hatcher, Hyunjung Joo, Sahyang Kim, Taehong Cho** (HIPCS, Hanyang U; Rutgers' U; Hongik U; HIPCS, Hanyang U)

17:30-18:10	▪ Invited Speaker: Alan Yu (University of Chicago)
18:30-21:00	**Reception / Banquet**

Day 2 (Saturday, May 27, 2023)	
08:00-08:50	**Registration** (Morning Coffee & Some Korean Rice Cakes and Donuts)
Oral Session 5: 09:00-10:10 **Chair:** Harim Kwon (Seoul National University)	
09:00-09:40	▪ Invited Speaker: Doug Whalen (CUNY & Haskins Laboratories)
09:40-10:10 2 talks	▪ **Sejin Oh, Alice Yildiz, Cecile Fougeron, Philipp Buech, Anne Hermes** (LPP (UMR7018, CNRS/Sorbonne-Nouvelle)) ▪ **Wendy Sandler** (U of Haifa)
10:10-10:40	Coffee Break
Oral Session 6: 10:40-12:05 **Chair:** Minjung Son (Hannam University)	
10:40-11:20	▪ Invited Speaker: Marianne Pouplier (Ludwig-Maximilians University Munich)
11:20-12:05 3 talks	▪ **Ryan Bennett, Jaye Padgett, Grant Mcguire, Maire Ni Chiosain, Jennifer Bellik (UC, Santa Cruz; UC, Santa Cruz; U College Dublin; UC, Santa Cruz,; UC, Santa Cruz)** ▪ **Jungyun Seo, Sahyang Kim, Taehong Cho** (U of Michigan; Hongik U, HIPCS, Hanyang U) ▪ **Shinichiro Sano, Celeste Guillemot** (Keio U)
12:05-13:30	Lunch (1hr 25m)
Oral Session 7: 13:30-14:40 **Chair:** Holger Mitterer (U. of Malta/HIPCS)	
13:30-14:10	▪ Invited Speaker: Andy Wedel (University of Arizona)
14:10-14:40 2 talks	▪ **Kuniko Nielsen, Rebecca Scarborough** (Oakland U; U of Colorado Boulder) ▪ **Yevgeniy Melguy, Keith Johnson** (UC, Berkeley)
Poster Session 2: 14:40-16:30 (45 posters: poster# p49-p93) (see below) **Chair:** Jonny Jungyun Kim (Pusan National University/HIPCS)	
Oral Session 8: 16:30-17:40 **Chair:** Jongho Jun (Seoul National University)	
16:30-17:00 2 talks	▪ **Canaan Breiss, Donca Steriade** (MIT) ▪ **Suyeon Yun** (Chungnam National U)
17:00-17:40	▪ Invited Speaker: Donca Steriade (MIT)
17:40-18:10	**General Discussion and Closing**

Organized by

Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS)
Convergence Research Center for Articulatory DB and Cognitive Sciences
Hanyang Humanities Enhancement Center (H^2EC)
Department of English Language & Literature, Hanyang University

Sponsored by

Global Research Fund from Hanyang University
National Research Foundation of Korea
Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS)
Hanyang Humanities Enhancement Center (H^2EC)

The Organizing Committee

Taehong Cho (Chair, Director of HIPCS, Hanyang University, Seoul)
Sahyang Kim (Co-Chair, Hongik University & HIPCS)
Sang-Im Lee Kim (Secretary, HIPCS, Hanyang University))
Say Young Kim (HIPCS, Hanyang University)
Suyeon Im (Soongsil University & HIPCS)
Jiyoung Lee (Graduate Student Organizing Manager, HIPCS, Hanyang University)
Eunhwan Lee (Administrative staff, HIPCS, Hanyang University)

Post-doctoral and student staff

Richard Hatcher (Post-doc, HIPCS, Hanyang University)
Dae-yong Lee (Post-doc, HIPCS, Hanyang University)
Jungah Lee (Post-doc, HIPCS, Hanyang University)
Jiyeon Song (Post-doc, HIPCS, Hanyang University)
Ziqian Du (Research Assistant, HIPCS, Hanyang University)
Sungwok Hwang (Research Assistant, HIPCS, Hanyang University)
Hongmei Li (Research Assistant, HIPCS, Hanyang University)
Mingi Park (Research Assistant, HIPCS, Hanyang University)
Jae-Eun Jennifer Shin (Research Assistant, HIPCS, Hanyang University)

TABLE OF CONTENTS

PROGRAM IN DETAIL

<table>
<tr><th colspan="2">Day 0 (Thursday, May 25, 2023)</th></tr>
<tr><td colspan="2"><b>Special Colloquium:</b> 17:00-18:20</td></tr>
<tr><td>17:00-18:20</td><td>Invited Speaker: Adam Ussishkin (University of Arizona)
<i>Auditory and visual processing of root morphemes in Maltese</i></td></tr>
</table>

<table>
<tr><th colspan="2">Day 1 (Friday, May 26, 2023)</th></tr>
<tr><td>08:00-08:50</td><td><b>Registration</b>
(Morning Coffee & Some Korean Rice Cakes and Donuts)</td></tr>
<tr><td>09:00-09:20</td><td><b>Opening</b> (Taehong Cho, Chair)
▪ Hyunchul Chung (Executive Vice President): Welcoming Remarks
▪ Sungho Yoo (Dean, College of Humanities): Opening Remarks</td></tr>
<tr><td colspan="2"><b>Oral Session 1:</b> 09:20-10:30
<b>Chair:</b> Say Young Kim (HIPCS, Hanyang University)</td></tr>
<tr><td>09:20-10:00</td><td>▪ Invited Speaker: John Kingston (University of Massachusetts, Amherst)
<i>Beginning to understand how intensity influences speech perception</i></td></tr>
<tr><td>10:00-10:30
2 talks</td><td>▪ <b>Holger Mitterer, Eva Reinisch</b> (U of Malta; Acoustics Research Institute, Austrian Academy of Sciences)
<i>Selective adaptation between allophones of /r/ in German</i>
▪ <b>Chiung-Yu Chang, Alexandra Jesse, Lisa D. Sanders</b> (U of Massachusetts, Amherst)
<i>Event-related potential responses to morphophonological violations</i></td></tr>
<tr><td>10:30-11:00</td><td>Coffee Break</td></tr>
<tr><td colspan="2"><b>Oral Session 2:</b> 11:00-12:25
<b>Chair:</b> Sang-Im Lee-Kim (HIPCS, Hanyang University)</td></tr>
<tr><td>11:00-11:40</td><td>▪ Invited Speaker: Lisa Davidson (New York University)
<i>The phonetic details of word-level prosodic structure: evidence from Hawaiian</i></td></tr>
<tr><td>11:40-12:25
3 talks</td><td>▪ <b>Yaqian Huang</b> (UCLA)
<i>Tonal and phrasal distributions of sub-phonemic creaky voice in Mandarin</i>
▪ <b>Marc Garellek, Jianjing Kuang, Osbel López-Francisco, Jonathan D. Amith</b> (UC, San Diego; U of Pennsylvania; Universidad Autónoma de México, Iztacala; Gettysburg College)
<i>Tense voice and the role of non-contrastive elements in sound change</i>
▪ <b>Rasmus Puggaard-Rode</b> (Ludwig Maximilian U of Munich)
<i>Covariation between fine phonetic detail and outcomes of sound change in the microtypology of Jutland Danish dialects</i></td></tr>
<tr><td>12:25-13:45</td><td>Lunch (1hr 20m)</td></tr>
<tr><td colspan="2"><b>Oral Session 3:</b> 13:45-15:10
<b>Chair:</b> Suyeon Yun (Chungnam University)</td></tr>
<tr><td>13:45-14:25</td><td>▪ Invited Speaker: Adam Albright (MIT)
<i>Licensing voicelessness in Lakhota</i></td></tr>
<tr><td>14:25-15:10
3 talks</td><td>▪ <b>Jennifer Kuo</b> (UCLA)
<i>Markedness bias in reanalysis: an iterated learning model of Samoan thematic consonant alternations</i>
▪ <b>Carmen Kin Man Tang, Regine Lai</b> (The Chinese U of Hong Kong)
<i>The acquisition of Cantonese phonotactics</i></td></tr>
</table>

	▪ **Bowei Shao, Anne Hermes, Philipp Buech, Maria Giavazzi** (Ecole normale supérieure, Université PSL; LPP (UMR7018, CNRS/Sorbonne Nouvelle)) *Velar palatalization in Italian: Lexical stress induces resistance to sound change*
Poster Session 1: 15:10-17:00 (48 posters: poster#p01-p48) (see below) **Chair:** Jiyoun Choi (Sookmyung Women's University)	
Oral Session 4: 17:00-18:10 **Chair:** Jeff Holliday (Korea University)	
17:00-17:30 2 talks	▪ **Simona Sbranna, Aviad Albert, Martine Grice** (U of Cologne) *Can we use categories when investigating interlanguages?* ▪ **Richard Hatcher, Hyunjung Joo, Sahyang Kim, Taehong Cho** (HIPCS, Hanyang U; Rutgers' U; Hongik U; HIPCS, Hanyang U) *How does focus-induced prominence influence realization of edge tones and segmental anchoring in Seoul Korean – A preliminary report*
17:30-18:10	▪ Invited Speaker: Alan Yu (University of Chicago) *Individual differences in perceptual cue weighting: behavioral, neurophysiological, and social considerations*
18:30-21:00	**Reception / Banquet**

Posters on Day 1 (Friday): 15:10-17:00 (Posters # P01~P48)

(P01) Chun-Hsien Hsu, Wen-Chun. Huang, Tong-Hou Cheong (National Central U)
Investigating MMN Responses to Pitch Contrasts in Monolingual and Bilingual Speakers of Tonal Languages

(P02) Jueyu Lu, Albert Lee (Education U of Hong Kong)
The realization of lexical tones in Sichuan opera

(P03) Hyoju Kim, Jieun Lee (U of Kansas)
English listeners' perceptual adaptation to unfamiliar lexical suprasegmental contrast

(P04) Pamir Gogoi, Luke Horo, Gregory D. S. Anderson (Living Tongues Institute for Endangered Languages)
Acoustic analysis of Glottal Stops in Mundari

(P05) Vahid Sadeghi (Imam Khomeini International U)
Pitch accent alignment in Persian

(P06) Yike Yang, Dong Han (Hong Kong Shue Yan U)
Language dominance influences L1 attrition and L2 acquisition of lexical tones: Data from Mandarin-speaking immigrants in Hong Kong

(P07) Faith Chiu, Laura Bartoševičiūtė, Albert Lee, Yujia Yao (U of Glasgow; U of Essex; Education U of Hong Kong; U of Essex)
Perceiving speech produced with face masks in competing talker environments

(P08) Ok Joo Lee, Kyungmin Lee (Seoul National U)
Dependent pitch cues in tone perception: Evidence from Mandarin Chinese

(P09) Jae-Eun Jennifer Shin, Sahyang Kim, Taehong Cho (HIPCS, Hanyang U; Hongik U; HIPCS, Hanyang U)
Alignment of Prosodic and Syntactic Junctures and Vowel-initial Glottalization in Syntactic Disambiguation: A Preliminary Report

(P10) Soohyun Kwon, Taejin Yoon, Sujin Oh, Jeong-Im Han (Seoul National U; Sungshin Women's U; U of Wisconsin-Milwaukee; Konkuk U)
Variable realization of consonant clusters in Seoul and Gyeongsang Korean

(P11) Zihao Wei (Chinese U of Hong Kong)
Phonological Status of Voiced Fricatives in Fanchang Wu

(P12) Charlize Ma, Effie Kao, Raechel Kitamura, Stephanie Wang, Jahurul Islam, Gillian De Boer, Bryan Gick (U of British Columbia)
Relations between Opinion Convergence, Acoustic Convergence and Movement Convergence in Interlocutors

(P13) Yung-Hsiang Shawn Chang (National Taipei U of Technology)
Effects of consonantal contexts on L2 English tense-lax vowel perception and production

(P14) Cheonkam Jeong, Andrew Wedel (U of Arizona)

The Effect of Cue-specific Lexical Competitors on Hyperarticulation of VOT and F0 Contrasts in Korean stops

(P15) Ai Mizoguchi, Mark K. Tiede, D. H. Whalen (Maebashi Institute of Technology; Haskins Laboratories; Haskins Laboratories)
Effects of fine phonetic detail on speaker identification from Japanese nasal consonants

(P16) Yeong-Joon Kim (MIT)
A generative phonetic approach to the ongoing sound change in Kyengsang Korean

(P17) Maida Percival, Pedro Mateo Pedro, Sonya Bird (U of Toronto; U of Toronto; U of Victoria)
Production and perception of ejective stops in Hul'q'umi'num' and Q'anjob'al

(P18) Shiyu Zhang, Jeffrey Holliday (Korea U)
Phonetic targets in the clear speech vowel productions of native Chinese learners of Korean

(P19) Ernesto Gutierrez Topete (U of California, Berkeley)
Influence of research tasks and linguistic factors on phonetic convergence in language alternation

(P20) Liu Huangmei, Zhang Chunmei (U of Shanghai for Science and Technology)
Sound change reverse via short-term phonetic accommodation: evidence from an in-progress tonal sound change toward the prestigious accent

(P21) Jing Huang, Feng-Fan Hsieh, Yueh-Chin Chang (National Tsing Hua U)
An articulatory study of word-level prominence in two Mandarin dialects

(P22) Qianyutong Zhang, Shanpeng Li, Lei Zhu (Shanghai International Studies U; Nanjing U of Science and Technology; Shanghai International Studies U)
Cross-language perception of parallel encoded emotional and linguistic prosody by Chinese English learners

(P23) Sarah Babinski (U of Zurich)
Structured Suprasegmental Variation: Marking Prominence in Australian Languages

(P24) Christophe d'Alessandro, Gregoire Locqueville (Institut Jean Le Rond D'Alembert - Sorbonne Universite - CNRS; Sorbonne Universite)
Assessment of finger tapping for rhythm control in performative speech synthesis : a preliminary study

(P25) Hohsien Pan, Shaoren Lyu (National Yang Ming Chiao Tung U)
Taiwan Min Nan Checked Tones Sound Changes

(P26) Kuniko Nielsen (Oakland U)
Phonological Categories in perception and production: the link and individual variability

(P27) Michaela Watkins (U of Amsterdam)
An analysis of F0 enhancement in younger speakers of Standard Seoul Korean in medial and initial phrase position

(P28) Bijun Ling (Tongji U)
How does prosody distinguish Wh-statement from wh-question in Shanghai Chinese

(P29) Sungwok Hwang, Sahyang Kim, Taehong Cho (HIPCS, Hanyang U; Hongik U; HIPCS, Hanyang U)
Differential effects of prosodic boundary on glottalization of word-initial vowels in Korean: A preliminary report

(P30) Yusheng Mu (Shanghai International Studies U)
The Effect of L1 Tones and L2 Pitch Accent on Lexical Access

(P31) Zhenting Liu, Regine Lai (Chinese U of Hong Kong)
Development of pitch cues in tone discrimination: evidence from Cantonese

(P32) Sheng-Fu Wang (Academia Sinica)
Final lengthening at Tone Sandhi Group boundaries in Taiwan Southern Min: Boundary strength and surprisal

(P33) Stephen Politzer-Ahles, Yixin Cui (U of Kansas; Hong Kong Polytechnic U)
Allotonic variants do not prime each other: evidence from long-lag priming

(P34) Bonnie J. Fox (U of Hawaiʻi at Mānoa)
Exemplar Effects in the Perception and Production of Advanced and Intermediate L2 Korean Wh-Question Intonation

(P35) James Whang (Seoul National U)
Quantifying Phonetic Informativity: An Information Theoretic Approach

(P36) Shu-Wei Yang, Jung-Yueh Tu (National Chengchi U)
Perception of Korean Coda Consonants /p, t, k/ by Chinese Learners of Korean in Taiwan

(P37) Seung Suk Lee (U of Massachusetts, Amherst)
Detecting the Accentual Phrase boundaries in Seoul Korean using tonal and segmental cues

(P38) May Pik Yu Chan, Meredith Tamminga (U of Pennsylvania)
Is there an optimal window size for a moving window analysis of pitch entrainment?

(P39) Yan Dong (Dalian U of Technology)
Production and Perception of Merging Tones in Dalian Mandarin
(P40) Albert Lee, Yasuaki Shinohara, Faith Chiu, Tsz Ching Mut (Education U of Hong Kong; Waseda U; U of Glasgow; Education U of Hong Kong)
Discrimination and identification of Japanese quantity contrasts by native Cantonese, English, French, and Japanese listeners
(P41) Mei Ying Ki (Chinese U of Hong Kong)
Neutralization of vowel length contrast in Hong Kong Cantonese checked syllables
(P42) Chenhao Chiu, Po-Hsuan Huang (National Taiwan U)
Rounded or unrounded? An examination of high vowels in Taiwan Mandarin
(P43) Jonny Jungyun Kim, Mijung Lee (Pusan National U; Independent researcher)
Vowels in probabilistically easy and hard words produced by L1 and L2 speakers
(P44) Chieh-Ching Chen, Janice Fon (National Taiwan U)
Dialectal Variation of the Effect of Prosodic Prominence on Diphthong Reduction in Taiwan Mandarin – using /aɪ/ as an example
(P45) Hyunjung Joo, Sahyang Kim, Taehong Cho (Rutgers U; Hongik U; HIPCS, Hanyang U)
Tonal alignment with articulatory gestures in South Kyungsang Korean
(P46) Jungah Lee, Taehong Cho, Sahyang Kim (HIPCS, Hanyang U; HIPCS, Hanyang U; Hongik U)
Gender-related variation of nasality and sound change of denasalization driven by prosodic boundaries in Seoul Korean: A preliminary report
(P47) Le Xuan Chan, Rina Furusawa, Rin Tsujita, Seunghun Lee (National U of Singapore; International Christian U; International Christian U; International Christian U)
Prosodic Realizations of Accented and Unaccented Postpositions in Japanese
(P48) Wai-Sum Lee, Eric Zee (City U of Hong Kong)
Sound change and emergence of patterns of the syllable-final consonants in the Chinese dialects

Day 2 (Saturday, May 27, 2023)	
08:00-08:50	**Registration** (Morning Coffee & Some Korean Rice Cakes and Donuts)
Oral Session 5: 09:00-10:10 **Chair:** Harim Kwon (Seoul National University)	
09:00-09:40	▪ Invited Speaker: Doug Whalen (CUNY & Haskins Laboratories) *Challenges of analyzing variability in speech from linguistic and motor control perspectives*
09:40-10:10 2 talks	▪ **Sejin Oh, Alice Yildiz, Cecile Fougeron, Philipp Buech, Anne Hermes** (LPP (UMR7018, CNRS/Sorbonne-Nouvelle)) *Temporal coordination of CV: The case of liaison and enchaînement in French* ▪ **Wendy Sandler** (U of Haifa) *Speech and Sign: The Whole Human Language*
10:10-10:40	Coffee Break
Oral Session 6: 10:40-12:05 **Chair:** Minjung Son (Hannam University)	
10:40-11:20	▪ Invited Speaker: Marianne Pouplier (Ludwig-Maximilians University Munich) *Speakers, listeners, languages: Coarticulatory variability and contrast in spoken language dynamics*

11:20-12:05 3 talks	▪ **Ryan Bennett, Jaye Padgett, Grant Mcguire, Maire Ni Chiosain, Jennifer Bellik** (**UC, Santa Cruz; UC, Santa Cruz; U College Dublin; UC, Santa Cruz,; UC, Santa Cruz**) *Syllable position in secondary dorsal contrasts: an ultrasound study of Irish* ▪ **Jungyun Seo, Sahyang Kim, Taehong Cho** (U of Michigan; Hongik U, HIPCS, Hanyang U) *An acoustic and articulatory study on variation of high vowel devoicing across prosodic contexts and speakers in Korean* ▪ **Shinichiro Sano, Celeste Guillemot** (Keio U) *Contrast enhancement and the distribution of vowel duration in Japanese*
12:05-13:30	Lunch (1hr 25m)
Oral Session 7: 13:30-14:40 **Chair:** Holger Mitterer (U. of Malta/HIPCS)	
13:30-14:10	▪ Invited Speaker: Andy Wedel (University of Arizona) *Structural patterns in the lexicon and grammar evolve to support communication efficiency*
14:10-14:40 2 talks	▪ **Kuniko Nielsen, Rebecca Scarborough** (Oakland U; U of Colorado Boulder) *Acoustic and linguistic influences on f0 imitation* ▪ **Yevgeniy Melguy, Keith Johnson** (UC, Berkeley) *What are you sinking about? Effects of phonetic learning on online lexical processing of accented speech*
Poster Session 2: 14:40-16:30 (45 posters: poster# p49-p93) (see below) **Chair:** Jonny Jungyun Kim (Pusan National University/HIPCS)	
Oral Session 8: 16:30-17:40 **Chair:** Jongho Jun (Seoul National University)	
16:30-17:00 2 talks	▪ **Canaan Breiss, Donca Steriade** (MIT) *Lexical Sources of Phonological Alternation: a role for Voting Bases* ▪ **Suyeon Yun** (Chungnam National U) *Perceptual bases for the compensatory lengthening typology*
17:00-17:40	▪ Invited Speaker: Donca Steriade (MIT) *Resyllabification revisited: arguments for V-to-V intervals as units of rhythm*
17:40-18:10	**General Discussion and Closing**

Posters on Day 2 (Saturday): 14:40-16:30 (Posters # P49~P93)

(P49) Maho Morimoto, Ai Mizoguchi, Takayuki Arai (Sophia U; Maebashi Institute of Technology; Sophia U)
Place Assimilation of the Moraic Nasal to /r/ in Japanese
(P50) Luke Horo, Pamir Gogoi, Gregory D. S. Anderson (Living Tongues Institute for Endangered Languages)
Prominence in Mundari Disyllables and Inflected Polysyllabic Nouns
(P51) Jiwon Hwang, Hyunah Baek, Ellen Broselow (Stony Brook U; Ajou U; Stony Brook U)
Korean speakers' perception of cues to liquid contrasts: an EEG study
(P52) Yufei Niu, Ricky Ka Wai Chan (U of Hong Kong)
Phonetic cue-weighting in the production of Mandarin rising [T2] and low [T3] tones by Japanese learners
(P53) Jiyoung Lee, Sahyang Kim, Taehong Cho (HIPCS, Hanyang U; Hongik U; HIPCS, Hanyang U)
Intergestural CV timing of homophonous words with different morphological structures: A preliminary report on liquid /l/ in Korean
(P54) Chenhao Chiu, Wei-Cheng Hsiao, Huang-Yu Shih, Bao-Yu Hsieh, Yining Weng (National Taiwan U; Chang Gung U; Chang Gung U; Chang Gung U; National Taiwan U)
Estimating tongue stiffness during phonation using ultrasound passive shear wave elastography
(P55) Joo-Kyeong Lee (U of Seoul)
The effects of ultrasound image feedback on Korean L2 learners' production of English /r/ in production traininig

(P56) Tsz Ching Mut, Candide Simard, Apolonia Tamata, Kwing Lok Albert Lee (Education U of Hong Kong; U of South Pacific; U of South Pacific; Education U of Hong Kong)
Focus Prosody in Fijian: a Pilot Study

(P57) Wai Ling Law, Olga Dmitrieva, Lan Li, Jette Hansen Edwards (Hong Kong U of Science and Technology; Purdue U; Chinese U of Hong Kong, Shenzhen, Chinese U of Hong Kong)
Social evaluations of speech vary as a function of perceived speaker nativeness

(P58) Xinyue Liu, Peggy Mok (Chinese U of Hong Kong)
Dialect levelling across generations: A socio-phonetic study of the medial [i] and vowel shift in the Jin dialect spoken in Baotou, China

(P59) Rina Furusawa, Shigeto Kawahara (International Christian U; Keio U)
Exploring the impact of phonological restrictions on phonetic implementation patterns using singing voice: The case of kobushi singing in Japanese

(P60) Changhe Chen, Jonathan Havenhill (U of Hong Kong)
Vowel Nasality and Nasal Excrescence in Shanghai Chinese

(P61) Zhen Qin, Sang-Im Lee-Kim (Hong Kong U of Science and Technology; Hanyang U)
The effect of second-language learning experience on Korean listeners' use of pitch cues in the perception of Cantonese tones

(P62) Renata Kochančikaitė, Mikael Roll (Lund U)
Individual Differences in Phonological Proficiency and Correlation with Pitch Sensitivity: Three Types of Perceivers

(P63) Gwanhi Yun, Jae-Hyun Sung (Daegu U; Kongju National U)
Non-native articulatory variability in English phonological rule application: Evidence from Korean and Indian learners

(P64) Sophia Burnett (CY Cergy Paris U)
In my humble opinion: The prosodic portrayal of the non-standard 1sg

(P65) Hongmei Li, Sahyang Kim, Taehong Cho (Yanbian U; Hongik U; HIPCS, Hanyang U)
Effects of focus and lexical tones on preboundary lengthening and its kinematic characteristics in Mandarin Chinese: A preliminary report

(P66) Yubin Zhang, Annie Rialland, Louis Goldstein (U of Southern California; CNRS/Sorbonne-Nouvelle; U of Southern California)
A dynamical systems approach to F0 hard-landing downtrends in Embosi

(P67) Hye-Sook Park, Sunhee Kim (Seoul National U)
Acoustic Correlates of Emphatic Accent in French Vowels /i, a, u/

(P68) Jiyoung Jang, Argyro Katsika (U of California, Santa Barbara)
Edgy articulation: the kinematic profile of Accentual Phrase boundaries in Seoul Korean

(P69) Suyeon Im, Sahyang Kim, Taehong Cho (Soongsil U; Hongik U; HIPCS, Hanyang U)
Some asymmetrical pre- versus post-focal effects on articulatory realization of prominence distribution in Korean: A preliminary report

(P70) Frank Lihui Tan, Youngah Do (U of Hong Kong)
Bottom-up Learning of Phonetic System using Autoencoder

(P71) Lindy Comstock (UCLA)
Russian prosody as a special case of the mobile stress system in Indo-European languages

(P72) Dong Han, Mingyang Yu (Hong Kong Shue Yan U; City U of Hong Kong)
A Perceptual Study on the Distinctive Feature of Entering Tones in Guangzhou Cantonese

(P73) Donghyun Kim, Andrew Lee, Ron Thomson (Kumoh National Institute of Technology; Brock U; Brock U)
The roles of talker variability, lexical frequency, and listener characteristics in second language speech perception

(P74) Daniel Pape (McMaster U)
Perceptual stability of sibilants undergoing acoustic variation: Interplay between acoustic processing versus influences of articulatory and/or motor patterns

(P75) Jake Aziz (UCLA)
Acoustic Evidence for Gestural Alignment: Vowel Devoicing in Malagasy

(P76) Jiyeon Song, Sahyang Kim, Taehong Cho (HIPCS, Hanyang U; Hongik U; HIPCS, Hanyang U)
A Preliminary Study about Disappearing Laryngeal and Supralaryngeal Articulatory Distinction of the Three-way Contrast of Korean Velar Stops

(P77) Hyun-ju Kim (State U of New York, Korea)
Pitch accent in North Kyungsang Korean spoken word recognition

(P78) Dae-yong Lee, Melissa Baese-Berk (HIPCS, Hanyang U; U of Oregon)
Effect of listeners' linguistic experience on generalization of adaptation

(P79) Yishan Huang (U of Sydney)
Is Southern Min Tone Circle a Real Thing?
(P80) Ching-Hung Lai, Chenhao Chiu (National Cheng Kung U; National Taiwan U)
The effects of ultrasound biofeedback on vowel acoustics
(P81) Hsueh Chu Chen, Jing Xuan Tian (Education U of Hong Kong)
Cross-linguistic Influences among L1, L2, and L3 Monophthongs by Cantonese Speakers in the Multilingual Context

(P82) Eunkyung Sung, Sunhee Lee, Sehoon Jung (Cyber Hankuk U of Foreign Studies; Cyber Hankuk U of Foreign Studies; Kyungsung U)
The Effects of VOT on Lexical Access by L1 and L2 Listeners: An Eye-Tracking Study

(P83) Drew Crosby (U of South Carolina)
SSANOVA as a Method of Examining Nasality in Korean Aegyo

(P84) Chi-Wei Wang, Bo-Wei Chen, Bo-Xuan Huang, Ching-Hung Lai, Chenhao Chiu (National Taiwan U; National Taiwan U; National Taiwan U; National Cheng Kung U; National Taiwan U)
Evaluating forced alignment for under-resourced languages: A test on Squliq Atayal
(P85) Ted Kye (U of Washington)
Denasalization and the phonological representation of voiced stops in Lushootseed
(P86) Hijo Kang, Hyun-ju Kim (Chosun U; State U of New York, Korea)
L2 phonetic development in English stress acquisition by EFL and ESL Korean speakers
(P87) Jieun Lee, Hanyong Park (U of Kansas; U of Wisconsin-Milwaukee)
Within-category cue sensitivity in native language perception and its relation to non-native phonological contrast learning
(P88) Jingxuan Tian (Education U of Hong Kong)
Cross-linguistic Influences on Speech Prosody by Cantonese Multilingual Speakers
(P89) Yichen Wang, Benjamin Kramer, Noah Macey, Michael Stern, Yuyang Liu, Jason Shaw (Michigan State U; Yale U; Yale U; Yale U; Yale U; Yale U)
Crosslinguistic Influence on the Gestural Dynamics of Focus Prosody in Native Mandarin Learners of L2 English
(P90) Po-Hsuan Huang, Chenhao Chiu (National Taiwan U)
Tonal Coarticulation in Taiwan Mandarin and Taiwan Southern Min
(P91) Minkyoung Hong, Jeffrey Holliday (Korea U)
The role of L2 English in the perception of L3 Korean sibilants by L1 speakers of French and Vietnamese
(P92) Junkai Li, Chen Lu (Tianjin U; Shaanxi Normal U)
Syllabic perception of vowels: evidence from interlanguage
(P93) Jonathan Havenhill, Ming Liu, Shuang Zheng, Jonah Lack (U of Hong Kong)
Audiovisual enhancement in clear speech production of English laterals

Oral Presentations

Day 1

(Friday, May 26, 2023)

Beginning to understand how intensity influences speech perception

John Kingston[1]

[1]*University of Massachusetts, Amherst (USA)*

jkingstn@umass.edu

We might expect a more intense sound would simply provide more energy, and thereby convey greater prominence, and possibly that the sound is or belongs to a stronger prosodic constituent. Or greater intensity might trade off perceptually with a longer duration, as Repp (1979) showed for aspiration; for a fuller account of this trade off among the correlates of [voice] judgments, come to our ICPhS poster in Prague. The results to be reported there, like those that are the focus of this talk, show how the relative intensity of one acoustic interval does and does *not* influence the perception of a neighboring interval. While the specific purpose of the experiments reported here is to provide evidence relevant to the debate about whether the objects of speech perception are articulatory gestures or auditory qualities, their more general purpose is to remedy the neglect of a ubiquitous acoustic property.

All the experiments examine how the intensity of a preceding /al/ or /ar/ context influences the categorization of a following /da–ga/ target continuum. Since Mann (1980) first reported that listeners respond “ga” more often after /al/ than /ar/, manipulations of such stimuli have been used to argue that the objects of speech perception are the articulatory gestures that produce the speech signal’s acoustic properties, or alternatively, that they are the auditory qualities evoked by those acoustic properties (for reviews, see Fowler, 2006; Lotto & Holt, 2006). Listeners also respond “ga” more often after a non-speech analogue of /al/, which suggests that they perceive the target as contrasting spectrally with that spectrally high context, rather than compensating for coarticulation with that more anterior context. Three experiments test and reject the alternative mechanism, informational masking, which Viswanathan, Fowler, & Magnuson (2009); Viswanathan, Magnuson, & Fowler (2013) have proposed as responsible for the shift in categorization produced by the non-speech contexts. The first does so by showing that more intense speech contexts do not increase the size of the shift, the second by showing that more intense non-speech contexts don’t do so either, and that the shift gets smaller as the spectral distance between the contexts and targets increases, and third by showing that non-speech contexts with energy in the same auditory bands as the target shift categorization less than those with energy in complementary auditory bands. If they are available in time, the results will be discussed of a fourth experiment that tests an alternative account of Viswanathan et al.’s (2009) finding that band-passed speech contexts which preserved the spectral difference between /al/ and /ar/ but which were thereby rendered non-speech failed to shift categorization.

Taken together, these results show that the target’s perception is not simply determined by how intense its context is, but by how energy is distributed across the target’s and context’s spectra; that is, by how intensity is greater at some frequencies than others, and how differences between target and context in which of their frequencies are more intense interact perceptually within and between target and context. Just as we have learned that some milliseconds influence perception more than others, so, too, do some deciBels. (These studies have all been carried out in collaboration with Amanda Rysling, of the University of California, Santa Cruz.)

References

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. Perception and Psychophysics, 68 , 161-177.

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). Perception and Psychophysics, 68 , 178-183.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. Perception & Psychophysics, 28 , 407-412.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. Language and Speech, 22 , 173-189.

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. Psychonomic Bulletin and Review , 16 , 74-79.

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2013). Similar response patterns do not imply identical origins: An energetic masking account of nonspeech effects in compensation for coarticulation. Journal of Experimental Psychology: Human Perception and Performance, 39 (4), 1181-1192.

Selective adaptation between allophones of /r/ in German

Holger Mitterer[1,2] and Eva Reinisch[2]
[1]*University of Malta,* [2]*Hanyang University,* [3]*Austrian Academy of Sciences*

Selective adaption (SA) is an experimental paradigm that is often used to investigate pre-lexical representations of the speech signal (Harnad, 1987). In this paradigm, participants are exposed to a series of adaptor stimuli before categorizing a test stimulus from a continuum between two speech sounds. SA seems to reflect general perceptual principles, with counterparts in visual perception, such as the waterfall illusion (Goldstein, 1958), in which observers watch a waterfall for around 30s and then perceive stationary objects moving upwards. Such a contrastive effect is also observed in speech perception, with listeners perceiving the test stimuli as contrasting with the adaptors. That is, after hearing a serious of /b/-initial words, a stimulus from a [ba]-[da] continuum is more likely to be perceived as /da/, contrasting with the /b/-initial adaptors (Kleinschmidt & Jaeger, 2015).

The paradigm has had a waxing and waning popularity (Kleinschmidt & Jaeger, 2015) and critics pointed out potential post-perceptual influences on SA (Harnad, 1987). More recently, SA has been used to investigate pre-lexical representations in spoken-word recognition with the rationale that SA between adaptors in coda position and test stimuli in onset position would reflect position-invariant phonemic representations at a pre-lexical level. While one study found such an effect (Bowers et al., 2016), others (Mitterer et al., 2018; Samuel, 2020)) did not and pointed out phonetic confounds in (Bowers et al., 2016). One critical finding was that there was no adaptation between word-initial trilled /r/ and a sonorant /r/ in the coda position in Dutch (Samuel, 2020). Here we present three SA experiments, conducted online with about 30 participants each (advertised via *prolific.co*), that aim to replicate and extend this finding in German. Similar to Dutch, German has a large variety of allophones of /r/, including uvular and alveolar trills and a vocalized /r/ in the coda position (e.g., *Fischer*, Engl., 'fisherman', [fiʃɐ]).

The first experiment used an alveolar trill-lateral continuum ([rozə]-[lozə], Engl. 'rose'-'lottery tickets') as target stimuli, and different adaptor series containing either alveolar trills [r] (the maximal overlap with the test stimuli), uvular fricatives [ʁ], or vocalized versions of /r/ ([ɐ]). A control-adaptor condition was generated from words that did not contain any variant of /r/ or /l/. Results replicated (Mitterer et al., 2018) for Dutch, such that the most sonorant adaptor, the vocalized [ɐ], did not trigger any SA for the test stimuli containing an alveolar trill (with a Bayes Factor supporting the null over an alternative hypothesis, see Table 1 for a summary of the results). However, the uvular fricative, though phonetically different from the trill, led to SA. Experiments 2 and 3 focused on the uvular fricative and used an [ʁ]-[h] continuum ([ʁozə]-[hozə], Engl. 'rose'-'trousers', note that in German /h/ is the phoneme closest to [ʁ]) as test stimuli. The adaptor series contained either alveolar trills, uvular trills, or uvular fricatives. In Experiment 2, /r/ in the adaptors was word-initial (e.g., [ra:t], Engl. 'advice'), while in Experiment [3], it was word-medial but still in the syllable onset (e.g., *Barock,* Engl., 'baroque', [barɔk]). In both experiments, the surprising result was that the [r] adaptors caused stronger adaptation effects on the uvular-fricative stimuli from the test continuum than [ʁ] adaptors. In fact, in Experiment 3, the [ʁ] adaptors, with the same allophone as the test stimuli, even failed to produce any selective adaptation at all, while the trill adaptors did.

Overall, the results show that phonemic overlap is not sufficient to generate SA, which questions the assumption of phonemic representations at a pre-lexical level. However, in some cases SA is observed between different allophones, with the surprising result that alveolar trilled /r/ leads to stronger adaptation for both alveolar-trill and uvular-fricative test stimuli. This indicates that SA, as early critics already suggested (Harnad, 1987), may also arise at later, post-perceptual (rather than prelexical) levels of processing. For [r], this may be due to saliency of the amplitude modulation in trills (Delgutte & Kiang, 1984). With such post-perceptual influences, selective adaptation may not be the ideal paradigm to reveal prelexical representations in spoken-word recognition.

Table 1: Overview of selective-adaptation effects in the current study

	/r/ target stimulus	Adaption effects			
		strongest	→	weakest	
Exp1	[rozə] (alveolar trill)	[#rV…] >	[#ʁV…] >	[…Vɐ(C)#]	= ∅
Exp2	[ʁozə] (uvular fricative)	[#rV…] =	[#ʀV…] >	[#ʁV…]	> ∅
Exp3	[ʁozə] (uvular fricative)	[#...rV…] =	[#...ʀV…] >	[#...ʁV…]	= ∅

Note: "= ∅" means that an adaptor condition is similar to the control condition according to a Bayes Factor. "#" indicates a word boundary.

References

Bowers, J. S., Kazanina, N., & Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, *87*, 71–83. https://doi.org/10.1016/j.jml.2015.11.002

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *The Journal of the Acoustical Society of America*, *75*(3), 897–907. https://doi.org/10.1121/1.390599

Goldstein, A. G. (1958). On the after-effects of the "waterfall" and "spiral" illusions. *The American Journal of Psychology*, *71*, 608–609. https://doi.org/10.2307/1420264

Harnad, S. (1987). *Categorical perception: The groundwork of cognition*. Cambrigde University Press.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning? *Psychonomic Bulletin & Review*, *23*(3), 678–691. https://doi.org/10.3758/s13423-015-0943-z

Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language*, *98*(Supplement C), 77–92. https://doi.org/10.1016/j.jml.2017.09.005

Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language*, *111*, 104070. https://doi.org/10.1016/j.jml.2019.104070

Event-related potential responses to morphophonological violations

Chiung-Yu Chang[1], Alexandra Jesse[1] & Lisa D. Sanders[1]

[1]*University of Massachusetts Amherst (USA)*
chiungyuchan@umass.edu, ajesse@umass.edu, lsanders@umass.edu

This study focuses on the morphophonological processing of the English plural suffix -*s*, which is conditioned by the stem ending. The plural marker is typically /z/ (e.g., *shoe-s* /ʃu-z/), but it surfaces as /s/ after a voiceless sound (e.g., *cat-s* /kʰæt-s/) and /ɪz/ after sibilants (e.g., *wish-es* /wɪʃ-ɪz/). One possibility for representations and processing of phonological variations is that the complex word is stored and retrieved as a whole. Alternatively, the suffix could have a separate abstract representation, possibly along with all the potential phonological forms. To test these hypotheses, we compared event-related potentials (ERPs) evoked by nouns that take the canonical /z/ form of the plural suffix (e.g., *shoe-s* /ʃu-z/) to those evoked when the canonical /z/ is replaced by the wrong allomorph /s/ (e.g., /ʃu-s/) or by an unrelated phoneme /v/ (e.g., /ʃu-v/). Differences in the ERPs between the wrong allomorph /s/ and the unrelated phoneme /v/ would indicate that the abstract representation of the suffix, presumably with all its allomorphs, is accessed during speech perception rather than a representation of the whole word.

The second goal of this study is to determine the effects of phonotactic status and congruency of phonetic cues. The auditory stimuli included naturally spoken and cross-spliced items. For one set of spliced stimuli, a stem ending with a voiced obstruent was spliced to form phonotactically legal (e.g., bag-/z/) and illegal coda consonant clusters (bag-/s/ and bag-/v/; /gs/ and /gv/ are unattested). For another set, stems ending with a vowel or an approximant were used (e.g., *shoe* /ʃu/). The resulting suffixed forms are all phonotactically legal, but the segment preceding /s/ becomes unnaturally long compared to the naturally spoken items. Because the duration of the preceding vowel is a cue to obstruent voicing [1], we hypothesized that splicing causes incongruent phonetic cues. Both phonotactic status and congruency of phonetic cues may lead to different ERP response patterns.

Method: Eight native English listeners (of a planned 30) listened to noun phrases and rated the acceptability of the pronunciation in an ERP experiment. Each noun phrase began with *the*, *one*, or *two*. The noun was either singular or plural after *the*, always singular after *one*, and always plural after *two*. As summarized in Table 1, we manipulated the phonotactic legality and phonetic cue congruency in Spoken, Spliced & Legal, Spliced & Illegal conditions. We also presented nouns that end with /z, s, v/ to control for the acoustic differences between these sounds in a Control condition. Electroencephalography was time-locked to the onset of the fricatives /z, s, v/ and extracted from -100 to 500 ms. At least 23 artifact-free epochs contributed to the averaged ERP for each condition and participant. Mean amplitudes in 140-240 ms from left central electrodes (FC1, FC3, C1, C3) and 360-500 ms from central posterior electrodes (CPz, Pz, P1, P2, POz) were analyzed. These time windows and regions of interest were selected according to a previous study on violations of French voicing assimilation [2].

Results: Figure 1 illustrates the grand average ERPs when the determiner is *the*. Negative deflections appeared in the early time window for /s/ and /v/ violations in the Spoken condition. Also, a late positivity was evident at posterior sites in all conditions. We subtracted the mean amplitudes in the Control condition from the other conditions in subsequent comparisons.

Although the mean amplitude in the early time window for the /s/ violation was numerically more negative than for the /v/ violation or the canonical /z/, these differences were not statistically significant ($F(2, 14) = 3.09$, $p = .08$). The numerical differences in the late positivity were also not significant ($F(2, 14) = 3.05$, $p = .08$). In the Spliced & Legal condition, significant differences in both the early negativity ($F(2, 14) = 5.58$, $p = .02$) and late positivity ($F(2, 14) = 4.72$, $p = .03$) were observed. Post hoc single-step tests using the R package multcomp showed that the early negativity was smaller for /s/ than /z/ ($p < .01$) and /v/ ($p < .01$). The late positivity for /v/ was larger than for /z/ ($p = .29$). No other pairwise comparison was significant. As for the Spliced &

Illegal condition, the mean amplitudes were different in the late ($F(2, 14) = 5.61$, $p = .02$) but not the early time window ($F(2, 14) = 1.87$, $p = .19$). Post hoc tests showed a larger late positivity for /v/ than /z/ ($p < .01$) and /s/ ($p = .03$), but ERPs elicited by allomorphs /z/ and /s/ did not differ ($p = .77$).

Discussion and conclusion: In the Spoken condition, the ERP elicited by the wrong allomorph /s/ appeared to differ from the unrelated phoneme /v/ and the canonical /z/. The ERPs to /s/ also differed from /v/ in the Spliced & Illegal condition. Taken together, these findings provide some tentative evidence for the abstract representation of the plural suffix being accessed during speech perception. Our data also suggest that phonotactic status and phonetic cue congruency modulate the ERPs elicited by morphophonological violations. The early negativity to the phonetically incongruent /s/ was reduced in the Spliced & Legal condition, suggesting that providing a primary cue consistent with /z/ was sufficient to eliminate the differences in ERPs elicited by the /z/ and /s/ forms of the suffix. Meanwhile, the difference between conditions in the late positivity was enhanced in the Spliced & Illegal condition. Thus, our preliminary data suggest that abstract morphological representations of the English plural suffix, the congruency of acoustic cues with morphophonological forms, and phonotactic constraints all influence the processing of clear speech.

Table 1 Experimental conditions and examples of stimuli.

Condition	Phonotactic legality	Phonetic cue congruency	Example
Spoken	legal	congruent	*shoes* /ʃu-z/, /ʃu-s/, /ʃu-v/
Spliced & Legal	legal	incongruent	*shoes* /ʃu-z/, /ʃu-s/, /ʃu-v/
Spliced & Illegal	illegal	N/A	*bags* /bag-z/, /bag-s/, /bag-v/
Control	legal	congruent	*maze* /meɪz/, *bus* /bʌs/, dive /daɪv/

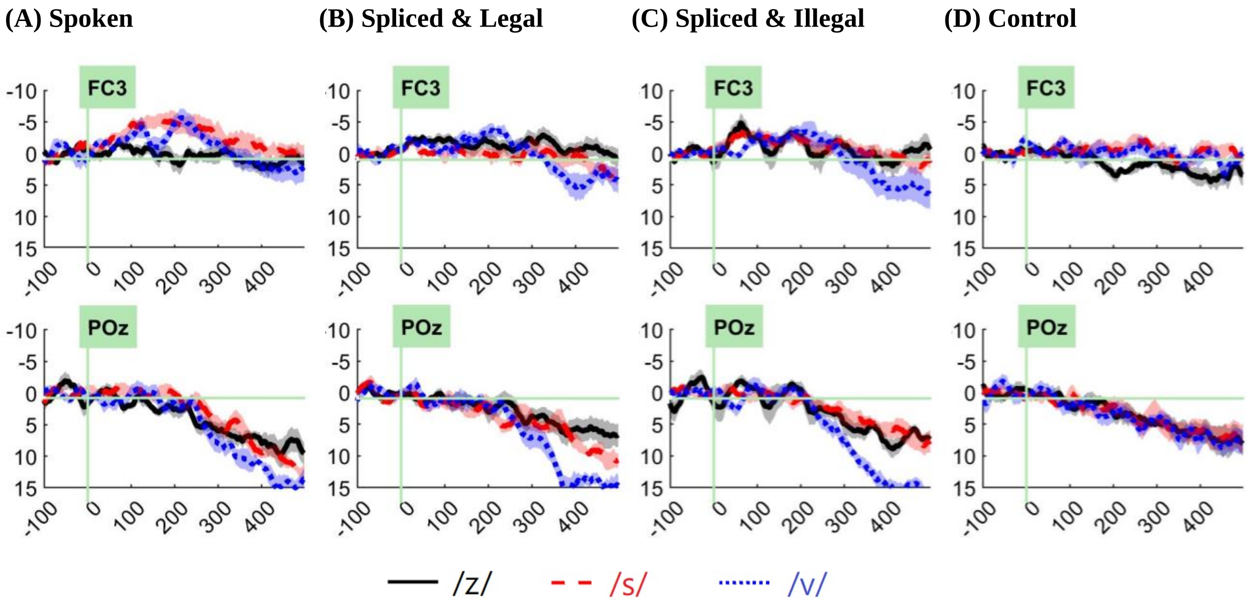


Fig.1 Grand average ERPs evoked by /z/ (solid black), /s/ (dashed red), and /v/ (dotted blue) from two representative electrodes FC3 and POz. Shaded areas indicate one standard error.

References

[1] Denes, P. (1955). Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America*, *27*(4), 761-764.

[2] Sun, Y., Giavazzi, M., Adda-Decker, M., Barbosa, L. S., Kouider, S., Bachoud-Lévi, A. C., ... & Peperkamp, S. (2015). Complex linguistic rules modulate early auditory brain responses. *Brain and Language*, *149*, 55-65.

The phonetic details of word-level prosodic structure: evidence from Hawaiian

Lisa Davidson

New York University (USA)
lisa.davidson@nyu.edu

Previous research has shown that the segmental and phonetic realization of consonants can be sensitive to word-internal prosodic and metrical boundaries [1-3]. At the same time, other work has shown that prosodic prominence, such as stressed or accented syllables, has a separate effect on phonetic implementation [4-6]. This talk focuses on the word-level factors affecting the phonetic implementation of glottal and oral stops in Hawaiian. We first investigate whether prosodic or metrical structure, or prosodic prominence such as stressed syllables account for the realization of glottal stop. We then extend the same analysis to the realization of voice onset time (VOT) in oral stops to determine whether both of these phonetic correlates have a similar or different relationship with the prosodic and/or metrical structure of Hawaiian.

Data come from 8 speakers (4M, 4F) from Ka Leo Hawaiʻi, a 1970s Hawaiian language radio show featuring interviews with bilingual Hawaiian/English speakers. The targets of the first analysis were phrase-medial phonemic glottal stops (N=758, e.g. [ˈʔa.ka] 'to laugh', [ˈpu.ʔu] 'hill'). Three broad categories for classifying the realization of /ʔ/ were identified: no glottalization or only a dip in F0 (27%), a period of glottalization (66%), or a full glottal closure (7%). Using the computational prosodic grammar developed in Parker Jones [7], the words were automatically categorized as to the stress of the syllable (primary stress, secondary stress, unstressed), the position of the glottal stop in the word (initial vs. medial), and the position of the glottal stop relative to prosodic word boundaries (prosodic word initial vs. medial). A multinomial regression for phonemic glottal stops indicated that glottalization was less likely and full glottal closures were more likely in prosodic word-initial position (see Figure 1) (e.g. {(ki:)}{(ʔa.ha)} 'cup'). Neither stressed nor heavy syllables were significant factors. One interpretation of these results is that a full closure may help indicate prosodic word boundaries, which could resolve cases where stress assignment does not disambiguate possible parses, e.g., {(ho:)}{ʔo.(a.ka)} or {(ho:.ʔo)}{(a.ka)}, 'to open'.

The second analysis examined the VOT duration of the oral stops /p/ and /k/ (N=5692, e.g., [papa] 'class', [kokoke] 'near'. Hawaiian only has /t/ in restricted contexts). Like for the glottal stops, the oral stops were also categorized for syllable stress, (lexical) word position, and prosodic word position. An additional aspect of the analysis was to determine whether stops in Hawaiian are aspirated or unaspirated, which had not previously been conclusively established [8, 9]. First, results indicate that oral stops in the Hawaiian speakers of this generation are unaspirated (avg /p/=24ms, /k/=39ms). The effects of the prosodic factors show that there are no main effects of stress or prosodic word on VOT duration, but there is an effect of lexical word position and an interaction between word position and stress (see Figure 2). Whereas VOT is longer in word-initial position for secondary and unstressed syllables, there is no difference for word position when a syllable has primary stress. If this difference is perceptible in secondary and unstressed syllables, perhaps it could serve as a cue that a less prominent syllable belongs to the beginning of a word, instead of being the final syllable of a preceding word.

Taken together, results show that word-internal metrical structures do condition phonetic realization, but prosodic prominence does not for either kind of stop. In contrast, in studies of languages like English, German or Polish, the insertion of glottalization and full stop realizations is more likely before stressed syllables or at higher prosodic boundaries [10, 11], and are taken to be the articulatory reflection of prosodic strengthening. Relatedly, languages with unaspirated stops like Dutch and Sahaptin Yakima demonstrate shortening of the stop release in prominent syllables, a fortition effect [12, 13], whereas a lengthening effect possibly attributable to word initial position occurs in secondary and unstressed syllables in Hawaiian. These results for both types of stops may reflect the recruitment of phonetic correlates to disambiguate or protect weaker elements, which is complemented by recent findings on the use of non-phonemic glottalization in Hawaiian to demarcate single-vowel function words (e.g. [a] or [i]) where they might otherwise be imperceptible [14].

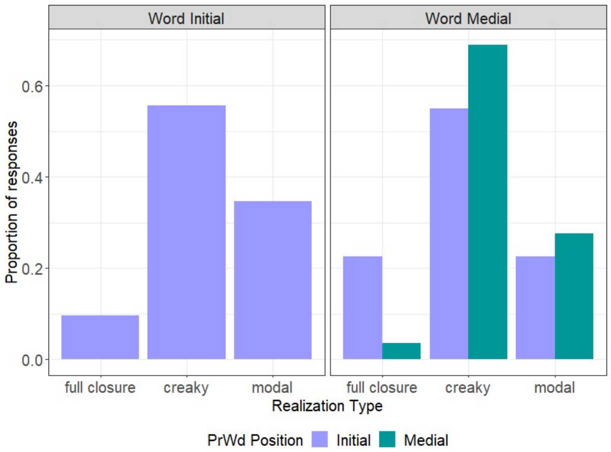


Figure 1. Proportions of full closure, creaky and modal implementations by prosodic and lexical word position.

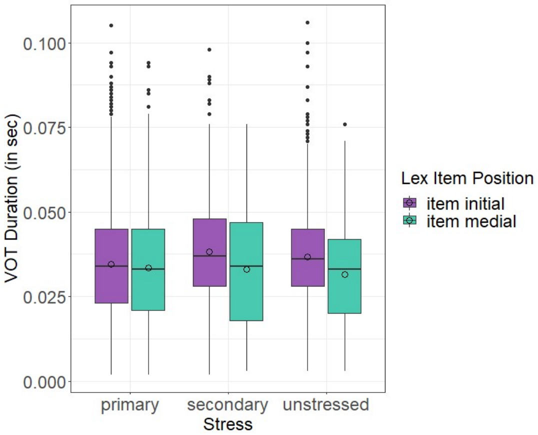


Figure 2. VOT durations by stress level and word position

References

[1] Vaysman, O. 2009. Segmental Alternations and Metrical Theory. PhD dissertation, MIT.
[2] Shaw, J. 2007. /ti/~/tʃi/ contrast preservation in Japanese loans parasitic on segmental cues to prosodic structure. In: Trouvain, J. and Barry, W. (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1365-1368). Saabrücken, Germany.
[3] Bennett, R. 2018. Recursive prosodic words in Kaqchikel (Mayan), *Glossa*. 3: 1, 67.
[4] Cho, T., and Keating, P. 2009. Effects of initial position versus prominence in English, *Journal of Phonetics*. 37: 4, 466-485.
[5] Garellek, M. 2014. Voice quality strengthening and glottalization, *Journal of Phonetics*. 45, 106-113.
[6] Katsika, A., and Tsai, K. 2021. The supralaryngeal articulation of stress and accent in Greek, *Journal of Phonetics*. 88, 101085.
[7] Parker Jones, O. 2010. A computational phonology and morphology of Hawaiian. DPhil. thesis, Oxford
[8] Parker Jones, O. 2018. Hawaiian, *Journal of the International Phonetic Association*. 48: 1, 103-115.
[9] Elbert, S., and Pukui, M. K., *Hawaiian Grammar*, Honolulu: University of Hawaii Press, 1979.
[10] Malisz, Z., Zygis, M., and Pompino-Marschall, B. 2013. Rhythmic structure effects on glottalisation: A study of different speech styles in Polish and German, *Laboratory Phonology*. 4: 1, 119-158.
[11] Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. 1996. Glottalization of word-initial vowels as a function of prosodic structure, *Journal of Phonetics*. 24, 423-444.
[12] Cho, T., and 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.
[13] Hargus, S. 2005. Prosody in two Athabaskan languages of Northern British Columbia. In: Hargus, S. and Rice, K. (eds.), *Athabaskan Prosody* (pp. 393-423). Amsterdam: John Benjamins Publishing Company.
[14] Davidson, L., and Parker Jones, O. submitted. Non-phonemic glottalization in a language with a glottal stop phoneme: a case study of Hawaiian.

Tonal and phrasal distributions of sub-phonemic creaky voice in Mandarin

Yaqian Huang

University of California Los Angeles
huangyq@g.ucla.edu

Lexical tones are primarily characterized by contrastive pitch patterns. Recent studies on the interaction between tone and phonation have enriched our understanding of non-contrastive phonetic dimensions that additionally capture tonal space [1,2]. In Mandarin Chinese, a tone language with contrastive pitch and without contrastive phonation, creaky voice is often found in tones that contain a low pitch target, most frequently with the lowest dipping Tone 3 (21[4]) and occasionally with rising Tone 2 (35) and falling Tone 4 (51) [3,4,5]. Sub-phonemic creak includes period doubling and vocal fry, which were found to have different gender distributions [6] and whose phonatory features have been detailed in [7]. It still remains unclear whether period doubling and vocal fry carry different linguistic functions in tone production and perception. This paper thus investigates how the production of period doubling and vocal fry may vary by prosodic factors such as tone and phrasing as they exist as different realizations of sub-phonemic creaky voice in Mandarin.

Methods. The materials are from a Mandarin corpus of simultaneous audio and electroglottography (EGG) recordings collected to document contextual tonal variation. The sentences consist of a fixed carrier phrase with varying stimuli of trisyllabic compounds, and the words are coded by phrasal positions:

‘I	teach	you		STIMULUS		how-to	say’	(gloss)
wo3	*tɕau1*	*nʲi3*	*Syll1*	*Syll2*	*Syll3*	*tsən3-mɤ0*	*ʂʷo1*	(sentence)
UI	PS2	PS1	PI	PM	PF	AS	UF	(phrasal positions)

Trisyllabic compounds have 64 (4 tones x 4 tones x 4 tones) varying tonal combinations. Three sets of 64 sentences with two repetitions (384 sentences in total) were elicited per recording. 20 native Mandarin speakers (10 F) in college participated in the production experiment. The corpus has 7680 sentences. We labeled the lexical tones and excluded neutral tones based on their phonetic realization. Fig. 1 shows examples of creak from the corpus: period doubling, characterized by alternating pulses between amplitudes and/or periods [8], and vocal fry, characterized by low f0 and high glottal constriction [9]. Both creak subtypes were identified anywhere in the phrase – during the trisyllabic stimuli or the carrier phrase – and coded with tone and phrasal positions, using the EGG signal, to avoid possible formant-induced interferences with the voicing signal. A total of 5848 tokens of period doubling and 1574 tokens of vocal fry have been identified and used in the prosodic analysis.

Results. The distribution of period doubling and vocal fry by tone, shown in Fig. 2, was analyzed for the compound stimuli only because tones do not vary elsewhere in the corpus. Overall, Tones 2 and 3 have more tokens of creak (either period doubling or vocal fry) than Tones 1 and 4. Vocal fry is rarely observed for Tone 1. For Tones 2 and 3, period-doubled tokens gradually increase as a function of the phrasal positions from the left edge (PI) to the right edge (PF) in trisyllabic sequences. This pattern is not attested in Tones 1 and 4 with an even distribution, probably due to fewer occurrences of creak for those tones. In contrast, the distribution of vocal fry does not seem to be conditioned by sentence-medial phrasal positions.

The majority of the creaky tokens are associated with the after-stimulus (AS) and utterance-final (UF) positions, which is consistent with the findings in [3] that creaky voice in Mandarin frequently occurs in the final and penultimate positions. Interestingly, for both women and men, a linear increase is only observed in period-doubled voice starting from the second word (PS2) to UF. While period doubling is mostly found in UF, vocal fry mostly occurs immediately after the stimulus (AS), which is a post-focal position. Tonal influences are also observed in these environments; e.g. fewer occurrences of vocal fry than period doubling are found in the UF position associated with a high-pitched Tone 1, and PS1 with the low Tone 3 shows more instances of vocal fry at least for women. These are consistent with the findings of tonal distribution in Fig. 2. However, UI also with Tone 3 has substantially fewer occurrences of vocal fry than PS1, possibly due to the phrasal effect. Thus, prosodic position seems to be a stronger driving factor than tone in favoring the occurrences of the two creaky voice subtypes.

Discussion. We hypothesize that period doubling is more strongly driven by utterance edges than is vocal fry, because the former reflects unstable voicing: towards the end of the utterance, voicing becomes progressively less stable [10]. Consequently, the occurrences of period doubling in utterance-final position could be a byproduct of downdrift (declination) as f0 progressively lowers towards the utterance edge without necessary constriction (see [7] on non-constricted quality during period doubling). The increasing trend of period doubling towards the end of the utterance is comparable to a similar phonation-ending gesture that [10] has found for both irregular and regular phonation at the end of an utterance. The utterance-final irregular phonation is usually produced with short intervals of adduction followed by longer intervals of abduction and/or with incomplete closure of the vocal folds, rather than adducted like vocal fry [10]. This hypothesis, if proved correct, has ramifications for speech production studies, such that data elicitation is recommended in non-final contexts to avoid conflation of modal and non-modal phonation in both segmental and suprasegmental units. For example, the realization of period doubling in Mandarin Tone 3 or other tones will result in similar kinds of articulation and voicing instability. In contrast, vocal fry is typically triggered by a low and compressed f0 range with more probable constriction, associated with the post-focal position [11]. The fact that vocal fry occurs more restrictively in the penultimate position suggests that it could signal a stronger linguistic role such as marking a weak prosodic element post-focally. These results help clarify subcategories within creaky voice and have implications for the taxonomy of creaky voice as a refined phonetic category with potentially different linguistic functions.

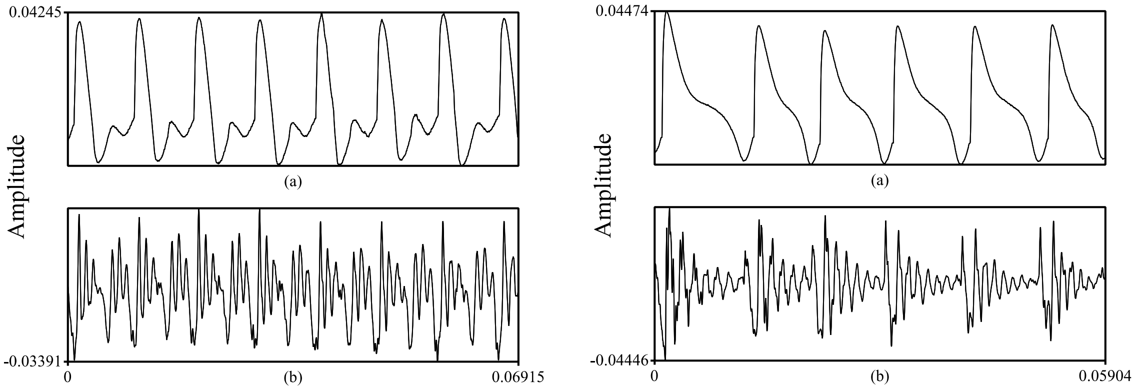


Figure 2. EGG (a) and audio (b) waveforms of period doubling (left) and vocal fry (right).

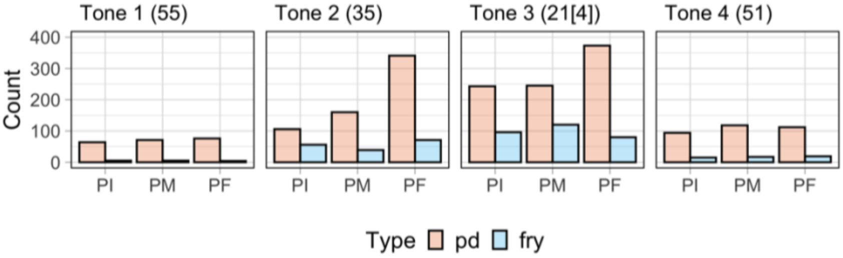


Figure 1. Raw count of period doubling and vocal fry across tones in sentence-medial positions. PI = phrase initial, PM = phrase medial, PF = phrase final.

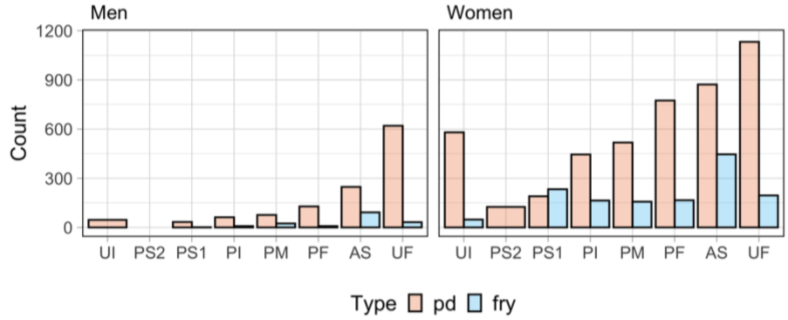


Figure 3. Raw count of period doubling and vocal fry across different phrasal positions throughout the sentence by gender.

[1] Kuang, J. (2013). *Phonation in tonal contrasts* (Doctoral dissertation, UCLA). [2] Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America*, *133*(2), 1078-1089. [3] Belotel-Grenié, A., & Grenié, M. (2004). The creaky voice phonation and the organisation of Chinese discourse. In *International symposium on tonal aspects of languages: With emphasis on tone languages*. [4] Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, *142*(3), 1693-1706. [5] Huang, Y., Athanasopoulou, A., & Vogel, I. (2018). The effect of focus on creaky phonation in Mandarin Chinese tones. *University of Pennsylvania Working Papers in Linguistics*, *24*(1), 12. [6] Yu, K. M. (2010). Laryngealization and features for Chinese tonal recognition. In *Eleventh Annual Conference of the International Speech Communication Association*. [7] Huang, Y. (2022). Articulatory properties of period-doubled voice in Mandarin. *Proc. Speech Prosody 2022*, 545-549. [8] Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1993). Perception of supraperiodic voices. *The Journal of the Acoustical Society of America*, *93*(4):2337–2337. [9] Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In *ICPhS*, volume 2015, pp. 2–7. [10] Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of voice*, *20*(2), 171-186. [11] Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. In *ICPhS*, volume 2011, pp. 152–155.

Tense voice and the role of non-contrastive elements in sound change

Marc Garellek[1], Jianjing Kuang[2], Osbel López-Francisco[3] & Jonathan D. Amith[4]

[1]*University of California, San Diego (USA),* [2]*University of Pennsylvania (USA),* [3]*Universidad Autónoma de México, Iztacala (Mexico),* [4]*Gettysburg College (USA)*

mgarellek@ucsd.edu, kuangj@ling.upenn.edu, osbel9@gmail.com, jonamith@gmail.com

Sound changes involving non-modal (usually breathy or creaky) vowels typically describe their development; e.g., many languages develop breathy or creaky vowels from what is analyzed as a former glottal consonant. Thus Proto-Mazatec *VhV and *VʔV > V̤ and V̰ in Jalapa Mazatec [1]. We address the opposite direction: sound *changes away from* non-modal vowels. This question has been under-explored in research on phonation change (cf. [2] for loss of breathy vowels in Khmer).

We explore possible changes in non-modal vowels by focusing on tense vowels, which in comparison to prototypically 'creaky' vowels are characterized by a weaker form of glottal constriction, higher periodicity, and a higher f0 [3]. We focus on tense vowels because their overall weaker creak can lead to distinct changes from other kinds of creaky voice. The main empirical question is, *What are the secondary (non-contrastive) acoustic correlates (as indicators of different articulations) to tense vowels in three languages:* Zongozotla Totonac (Tepehua-Totonac), spoken in Mexico, and Bo and Southern Yi, two Yi (Sino-Tibetan) languages from China. Totonac is toneless and contrasts modal vs. 'glottalized' vowels, which recent work has shown to be tense. Bo and Southern Yi contrast tense vs. lax (weakly breathy) vowels, and tone is orthogonal to phonation.

Audio recordings for the three languages were made in the field: 8 Zongozotla Totonac (ZT) speakers, 9 Bo speakers, and 12 Southern Yi (SY) speakers. Target words are minimal pairs contrasting in phonation; only 8 pairs were included for ZT, because minimal pairs involving phonation are generally rare and of low functional load across Totonac languages [4], a point we will return to below. About 40 pairs were recorded for Bo and SY. Recordings were segmented for the tense vowel vs. non-tense (modal or lax) counterpart; target vowels were analyzed for measures of voice quality using VoiceSauce [5]. We first investigate measures primarily associated with changes in phonation, such as H1*-H2* (lower with increased constriction), CPP (lower with increased irregularity). Then we measure secondary correlates to phonation: f0 (primarily associated with tone, sometimes higher with tense voice) and F1 (primarily associated with vowel quality, sometimes higher with tense voice). For more details on the relationship between phonation types and these measures, see [3]. For each measure/language, linear mixed-effects models (with maximal random-effects structure) were run to test whether phonation type (tense vs. non-tense) significantly (at $p < 0.05$) predicts a change in the mean. (For concision model outputs are omitted here.)

For ZT, H1*-H2* distinguishes tense vs. modal vowels, with tense vowels showing slightly lower values along this measure. For all other measures, no significant differences were found. Thus tense vowels in Totonac are weakly glottalized, with no secondary correlates such as changes in pitch or vowel quality. For Bo, tense vowels have lower H1*-H2*, higher CPP (suggesting they are less noisy), and higher in f0 than lax ones. Some statistically significant differences in phonation on F1 and F2 do occur but these vary unsystematically by vowel type. For SY, tense vowels in comparison to lax vowels have lower H1*-H2* and higher CPP (suggesting they are less noisy), but no difference in f0. Additionally, F1 for tense vowels is consistently higher. Results for f0, F1, and H1*-H2* are shown in Figure 1.

The results suggest several paths of sound change *away from* tense voice, depending on whether articulatory configurations involve tongue root retraction (SY) or greater longitudinal tension in the vocal folds (Bo). Because tense voice is a weaker form of glottalization, well within the range of comfortable modal phonation (cf. [3]), it is possible to hold all other configurations constant and minimally differ from modal voice in terms of vocal fold contact [6]. This appears to be the case for ZT, whose tense vowels are only indexed by (slight) glottalization. In turn, this suggests that in ZT *tense voice is likely to merge with modal voice*, rather than change into another type of contrast, because there are no clear secondary correlates to the contrast that could undergo enhancement [7]. In contrast, the main secondary correlate to the phonation contrast in Bo is higher f0. This implies a path towards sound change whereby the *phonation contrast merges with and complexifies the preexisting tone system*. Something similar has arguably occurred in the Vietic branch of Austroasiatic [8]. In Southern Yi, however, the main secondary correlate is F1, whereby tense vowels have higher F1 than lax ones. The implication for sound change is that *tense voice can be transphonologized into a more complex set of vowel contrasts* than was present in the historical form of the language. Something similar has been argued to have occurred across the Khmer branch of Austroasiatic, due to transphonologization of the 'breathy' register to a more complex vowel system [2].

The differences across languages also suggest a role that the lexicon plays in the realization of tense voice. The weak phonetic nature of tense vowels in ZT might be related to the fact that the contrast in Totonac has few minimal pairs and an overall low functional load. Indeed, the phonation contrast in other varieties of Totonac is described as very weak or as having disappeared [4,9]. On the other hand, in the Yi languages the tense-lax contrast has a higher functional load, with many minimal pairs and with higher phonological significance. Overall, we see that even for a specific subtype of non-modal phonation – tense voice – there can exist language-specific non-contrastive elements, and together with the lexicon these may play a role in predicting distinct paths of sound change.

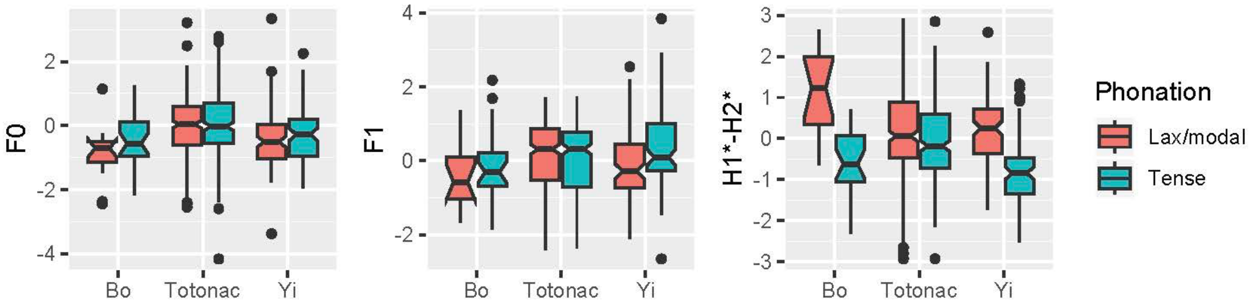


Fig.1 Boxplots of mean f0 (left), F1 (middle) and H1*-H2* (right), z-scored by speaker.

References

[1] Kirk, P. L. (1966). *Proto-Mazatec phonology*. Ph.D. thesis, University of Washington.

[2] Wayland, R. P. & Jongman, A. (2002). Registrogenesis in Khmer: A phonetic account. *Mon-Khmer Studies*, *32*, 101-115.

[3] Keating, P., Kuang, J., Garellek, M., Esposito, C., & Khan, S. (2023, in press). A cross-language acoustic space for vocalic phonation distinction. To appear in *Language.* URL : https://linguistics.ucla.edu/people/keating/Keating_etal_Language_accepted_Dec2022.pdf (accessed March 22, 2023).

[4] McFarland, T. A. (2009). *The phonology and morphology of Filomeno Mata Totonac*. Ph.D. thesis, University of California, Berkeley.

[5] Shue, Y. L., Keating, P., Vicenik, C., & Yu, K. (2011). Voicesauce: a program for voice analysis. Proceedings of the 17th international congress of phonetic sciences, 1846–1849.

[6] Kuang, J., & Keating, P. (2014). Vocal fold vibratory patterns in tense versus lax phonation contrasts. *The Journal of the Acoustical Society of America*, *136*(5), 2784-2797.

[7] Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization* (pp. 51-97). Oxford, UK: Oxford University Press.

[8] Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica*, *19*, 333-363.

[9] MacKay, C. J., & Trechsel, F. R. (2018). An alternative reconstruction of Proto-Totonac-Tepehua. *International Journal of American Linguistics*, *84*(1), 51-92.

Covariation between fine phonetic detail and outcomes of sound change in the microtypology of Jutland Danish dialects

Rasmus Puggaard-Rode[1]

[1]*Ludwig Maximilian University of Munich (Germany)*
r.puggaard@phonetik.uni-muenchen.de

In this paper, we present a case of regional covariation between fine phonetic detail in one prosodic context and sound change in a different prosodic context. The case in question is the process of *stop gradation* in varieties of Danish spoken on the Jutland peninsula. Simply put, stop gradation resulted in stop phonemes acquiring radically different allophones in different prosodic contexts. Dialectal variation in stop gradation is well-described, but the mechanisms that caused this variation are not well-understood. Through acoustic-phonetic exploration of a legacy corpus of dialect speech, we show that the different regional outcomes of stop gradation correspond very well to variation in fine phonetic detail in stop realization throughout the peninsula.

In Modern Standard Danish, stop gradation is usually analyzed as a phonological process whereby /p t k/ are realized as voiceless aspirated [pʰ tʰ kʰ] in 'strong' position, and voiceless unaspirated [p t k] in 'weak' position, while /b d ɡ/ are realized as voiceless unaspirated stops [p t k] in strong position and semivowels [ʊ̯ ɣ̯ ɪ̯] in weak position [1]. In this context, 'strong' position (SP) refers to the syllable-initial position before full vowels, and 'weak' position (WP) refers to the syllable-final position *or* the syllable-initial position before neutral vowels. See the following examples of WP–SP alternation with the proposed phonemes /t d/:

		/t/	**/d/**	
WP:	[ʋæ**t**]	*vat*, 'cotton wool'	[so.ˈli**ɣ̯**ˀ]	*solid*, 'solid'
SP:	[ʋæ.ˈ**tʰ**+eːˀɐ]	*vattere*, 'apply cotton wool'	[so.li.**t**i.ˈtʰ+eːˀt]	*soliditet*, 'solidity'

Dialectological research from the early 20th century has shown that stop gradation affected almost all parts of the Danish speaking area, but with various WP outcomes. In some areas, /b d ɡ/ surface as semivowels in WP; in some areas, they surface as voiced fricatives in WP; in yet others, they surface as voiceless fricatives in WP. This is illustrated in Fig 1 (adapted from [2]), which maps WP /ɡ/ after front vowels. SP variation has never before been considered, even though there are strong reasons to expect covariation between WP and SP. If we assume that the WP and SP allophones belong to a single phonological category, it follows that WP and SP allophones might covary in phonologically well-motivated ways. More specifically, Fig 1 shows a clear geographical pattern of highly sonorous WP allophones in the northern part of the peninsula, and less sonorous WP allophones in the southern part of the peninsula, with a seemingly gradual cline in between. We hypothesize that the precise phonetic implementation of the laryngeal contrast in SP stops should show a similar cline; in particular, we would expect that voicing is a relatively strong SP cue in areas with highly sonorous WP allophones, and that aspiration is a relatively strong SP cue in areas with less sonorous WP allophones.

Since much of the regional variation in Danish was leveled over the course of the past century [3], we test this hypothesis using a large legacy corpus of sociolinguistic interviews that was collected between 1971–1976. Using these recordings, we carried out an exploratory study of variation in SP stop acoustics in the traditional Jutland Danish varieties. Voice onset time (VOT) was measured from 10,650 tokens of SP /p t k/, and the presence or absence of continuous closure voicing was recorded for 6,854 tokens of SP /b d ɡ/. The speakers come from 213 different locations in the peninsula.

The resulting measures were analyzed with spatial generalized additive mixed models. These models include two-dimensional smooth variables modeling geographical coordinates, allowing us to model a non-linear geographical effect. The models also include a host of other dependent variables that are known to influence VOT and the relative likelihood of closure voicing, including e.g. place of articulation, speaker gender, and stress. The results of these models are plotted in Figs 2 and 3, respectively. In both cases, the models perform significantly better than corresponding non-spatial models.

Figs 1, 2, and 3 show striking similarities. Perhaps unsurprisingly, there is obvious covariation between VOT and closure voicing; when SP /p t k/ are cued with high VOT, SP /b d g/ are less likely to be voiced, and *vice versa*. These differences are gradient rather than categorical; it is not simply the case that varieties implement the SP laryngeal distinction with either aspiration or closure voicing. As predicted, we also see correspondences between VOT and voicing in SP, and regional outcomes of stop gradation in WP: where VOT is low and the likelihood of voicing is high in SP, stop gradation resulted in voiced fricatives or semivowels in WP; where VOT is high and the likelihood of voicing is low in SP, stop gradation resulted in voiceless fricatives in WP. There is congruence between SP fine phonetic detail and WP allophone selection: more voicing-prone areas lenite in a more distinctly sonorous direction, while more aspiration-prone areas lenite in a direction that is more likely to maintain voicelessness. In this respect, Jutland Danish provides an illuminating microtypology of how gradient variation in fine phonetic detail can feed directly into sound change and categorical phonology.

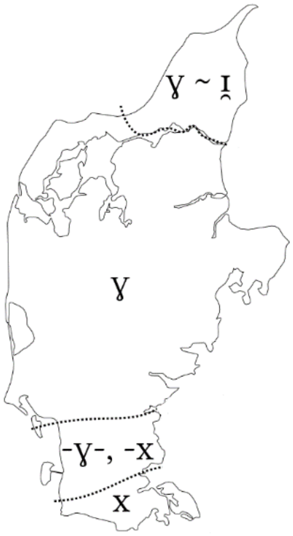


Fig 1

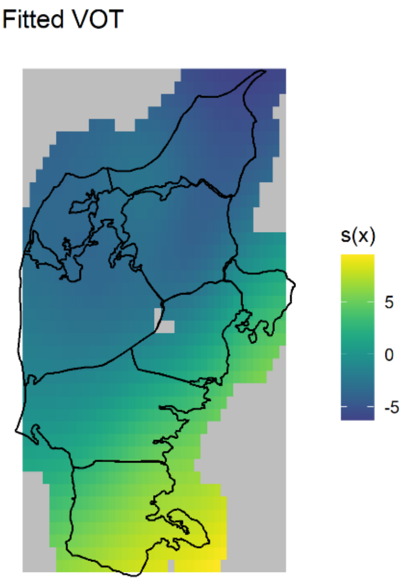


Fig 2

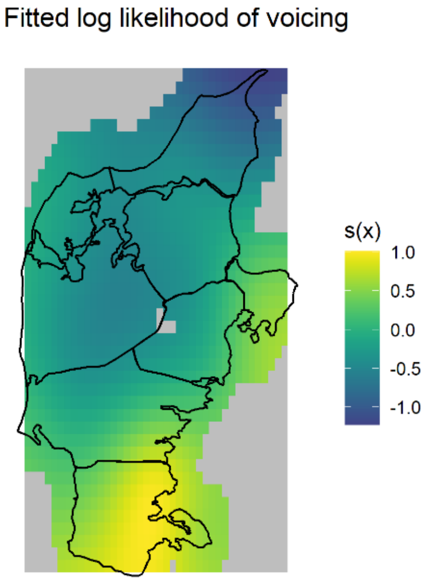


Fig 3

References

[1] Horslund, C.S., Puggaard-Rode, R. & Jørgensen, H. (2022). A phonetically-based phoneme analysis of the Danish consonant system. *Acta Linguistica Hafniensia* 54(1), 73–105.

[2] Bennike, V. & Kristensen, M. (1898–1912). *Kort over de danske folkemål med forklaringer*. Copenhagen: Gyldendal.

[3] Pedersen, I.L. (2003). Traditional dialects of Danish and the de-dialectalization 1900–2000. *International Journal of the Sociology of Language* 159, 9–28.

Licensing voicelessness in Lakhota

Adam Albright (MIT)

Obstruent voicing in Lakhota (Siouan) exhibits several typologically remarkable properties. Stops and fricatives exhibit different distributions of voicing, each interesting in its own right, and which raise further questions when combined in the same language. Phonological analyses of the typology of voiced obstruents have generally focused on constraints against voicing in various contexts. In this paper, I show that the Lakhota distributions run counter to several commonly observed restrictions on obstruent voicing. In particular, stops are required to be voiced in contexts that cross-linguistically frequently favor devoicing (in consonant clusters, and morpheme-finally), while fricatives are required to be voiceless in these same contexts. I argue that these discrepancies are due to phonetic conditions on *devoicing* segments in Lakhota–that is, on implementing voicelessness. In particular, I show that obstruents are required to be voiced in contexts where they have short duration, and where a rapid glottal abduction gesture would be required to produce voicelessness.

Stops in Lakhota contrast robustly for aspiration and ejection (*p*, p^h, *p'*), but they are contextually neutralized to voiceless unaspirated stops in certain contexts, such as in clusters with other obstruents (*pt, kp, sp, xt, ks, pʃ*, etc.). However, stops are also contextually neutralized to *voiced* in certain contexts [1,2]. One such context is in clusters with sonorants: *pl, kl, km, kn, pj, kw* → *bl, gl, gm, gn. bj, gw* (etc.). When the sonorant is a liquid or nasal, these voiced stops are also followed by short epenthetic vocalic periods ($b^{ə}l$, $g^{ə}l$, $g^{ə}m$, $g^{ə}n$) [3]. Voicing before sonorants is phonetically complete, with voicing through the entire stop closure. It applies productively whenever such sequences arise through morphological concatenation: e.g., /ʃ'ak(A)+ja/ → [ʃ'agja] 'strongly'. Why do stops voice in clusters with sonorants? At first blush, this resembles regressive voicing assimilation, since it occurs before voiced consonants. However, there are several obstacles to such an analysis. First, cross-linguistically, sonorants typically do not trigger regressive voicing assimilation, but voicing is triggered by all (and only) sonorants in Lakhota. Second, voicing assimilation rarely crosses an epenthetic vowel cross-linguistically. Finally, voicing assimilation typically affects both stops and fricatives, but in Lakhota, fricatives pattern differently from stops. In what follows, I will argue that voicing in this context is not due to assimilation, but rather, to a ban on voiceless stops when they are the sole obstruent in a cluster.

Unlike stops, fricatives in Lakhota contrast for voicing and ejection (*s, z, s'*). However, in clusters, they are systematically neutralized to *voiceless*, even before sonorants (*sn, ʃn, xm, xn*, *xl*, *ʃw* etc.). This is part of a broader restriction: all fricatives in clusters must be voiceless, both in C1 (*sp, sk, ʃp , sn, xn, xl, ʃw*, etc.) and C2 position (*ps, ks, pʃ, kʃ*). Why are fricatives in clusters always voiceless, while stops vary depending on the sonority of the other member? This is not due to some general dispreference against voiced stops in the language; in fact, in certain positions of the root, fricatives must actually be voiced. Parallel to stops, this appears to be due to a special ban on voiced fricatives in clusters.

The key property that distinguishes singleton obstruents from those in clusters is duration. This can be seen by comparing the closure duration of stops in singleton CV contexts with those in CCV contexts. Singleton and cluster tokens were extracted from naturally occurring radio broadcast speech from two fluent native speakers. The ideal comparison is stop and fricative duration in controlled comparisons: *a*[t]*a* vs. *a*[t]*ka* and *ak*[t]*a*, and *a*[s]*a* vs. *a*[s]*ka* and *a*[k]*sa*. Due to limitations of recording quality, the analysis focused on stridents, which also occur more frequently. The results show that singleton fricatives are indeed longer in duration than those in clusters (mean 137ms vs. 96ms). For stops, the analysis focused on intervocalic vs. fricative-stop clusters, to avoid ambiguities of segmentation in stop-stop clusters. Here, too, the results show that singleton stops have longer duration than those in clusters (96ms vs. 64ms). These singleton durations are in line with, but overall shorter, than those reported for VCV in careful citation forms by [4]. Finally, in order to ensure that these differences reflect durational patterns of obstruents in general, and not just voiceless stops and fricatives, a small number of aspirated stops in Vk^{h}V vs. Vtk^{h}V were compared; these, too, show longer closure duration in singletons than clusters (88ms vs. 61ms).

This durational difference supports an analysis of neutralization in which the requirement to voice stops and devoice fricatives in clusters is tied to their short duration in this position. Specifically, this can be modeled with a pair of MinDist conditions [5] demanding that laryngeal contrasts in stops be supported by adequate differences in closure or VOT duration, while contrasts in fricatives must be supported by adequate differences in frication duration. With an appropriate choice of threshold, contrasts are tolerated among singletons, but prohibited within clusters. This analysis achieves broader coverage than analyses that focus on licensing laryngeal contrasts with release cues [6], because it is able to explain why the second members of clusters neutralize, even though they are prevocalic. Finally, I hypothesize that stops undergo voicing in stop+sonorant clusters in order to avoid the rapid glottal abduction gesture needed to produce a short voiceless stop (indicated here simply as *Rapid). Voicelessness can be achieved in singleton stops, which are longer, through passive devoicing, yielding a closure that is voiceless for much of its closure duration. This is not possible for short duration stops.

/apla, ap^{h}la, ap'la/	MinDist	*Rapid	Ident	*Gesture
a. a{b,p^h,p'}la	3!		1	4
b. apla		1!	2	1
→ c. abla			3	

/asla, azla, as'la/	MinDist	*Rapid	Ident	*Gesture
a. a{s,z,s'}la	3!			5
→ b. asla			2	1
c. azla			2	2

A final benefit of this duration-based analysis is that it extends easily to another mystery of Lakhota laryngeal phonology: in morpheme-final position, stops voice, while fricatives devoice. As [4] demonstrate, morpheme-final voiced stops are very short, relative to intervocalic stops. I show that by using duration as a licensing factor, a similar neutralization can be derived as in stop+sonorant clusters.

References

[1] Rood, David and Allan Taylor (1996). Sketch of Lakhota, a Siouan language, Pt. I. *Handbook of North American Indians, vol. 17: Languages,* 440–482. Washington, DC: Smithsonian.

[2] Rood, David (2016). The phonology of Lakota voiced stops. In *Advances in the study of Siouan Languages and Linguistics* (Catherine Rudin and Bryan J. Gordon, eds), 233–255. Berlin: Language Science.

[3] Boas, F. and Deloria, E. (1941). *Dakota Grammar*. Vol. 23 of Memoirs of the National Academy of Sciences. United States Government Printing Office, Washington.

[4] Blevins, Juliette and Ander Egurtzegi and Jan Ullrich (2020). Final Obstruent Voicing in Lakota: Phonetic Evidence and Phonological Implications. *Language* 96:294–337.

[5] Flemming, Edward (2017). Dispersion Theory and Phonology. Oxford Research Encyclopedia of Linguistics. https://doi.org/10.1093/acrefore/9780199384655.013.110

[6] Steriade, Donca (1997). Phonetics in phonology: The case of laryngeal neutralization. UCLA ms.

Markedness bias in reanalysis: an iterated learning model of Samoan thematic consonant alternations

Jennifer Kuo[1]

[1]*University of California, Los Angeles*
Jenniferkuo2018@ucla.edu

Paradigms with conflicting data patterns can be difficult to learn, resulting in child errors (e.g. *go/goed* instead of *go/went* in English). Such errors can in turn be adopted into speech communities, resulting in a type of change over time I refer to as *reanalysis*. Existing models of morphophonology, such as Albright's [1, 2] Minimal Generalization Learner, predict reanalysis to be frequency-matching, occurring in a way that matches probabilistic distributions within a paradigm. I propose that in fact, reanalysis responds to two factors: both frequency matching and the reduction of markedness.

In this study, I use iterated learning models to investigate this issue in a set of Samoan alternations. Two models are compared: one that is frequency-matching, and one which has a markedness learning bias. I find that the latter model performs better. I further propose that the markedness effects allowed to influence reanalysis are restricted to those already active in the language (e.g. in root phonotactics), and show that Samoan is consistent with this proposal.

In some Samoan suffixes, a consonant of unpredictable quality surfaces, as exemplified in (1) for the ergative suffix [3]. This pattern arose due to a historic process of final consonant loss. As a result, all consonants were deleted at the end of unsuffixed stems, but maintained in suffixed forms (e.g. *inum/*inum-ia 'to drink' →inu/inu-mia).

In general, the allomorph that surfaces should be traceable back to the historic stem-final consonant in Proto-Oceanic (POc) [4]. For example, [inu]/[inu-**m**ia] 'to drink' comes from POc *inu**m**, and [pulu]/[pulu-**t**ia] 'to plug up' comes from *bulu**t**. However, in modern Samoan, the observed alternant often does *not* match the historical POc one; for example, [ŋuŋu] (<POc *ŋuŋul) 'arthritis' should have the suffixed form [ŋuŋu-lia], but instead [ŋuŋu-a] is observed. These mismatches suggest that language learners have carried out reanalysis in multiple instances. To investigate the direction of reanalyses, I collected 358 POc forms with known Samoan reflexes, taken from the Austronesian Comparative Dictionary [5]. POc stems were compared against 584 Samoan stem/ergative pairs collected from Milner's Samoan Dictionary [6].

I find that reanalysis is sensitive to transvocalic consonant OCP effects. In particular, suffixed forms are more likely to be reanalyzed if they violate an OCP constraint against coronal sonorants (*[+COR,+son]...[+COR,+son]), which assigns violations to stems such as [**la**n**u**] 'color'. In fact, in modern Samoan, there are almost no suffixed forms of the type [pu**l**i-**n**a] or [pu**n**i-**l**ia] (n=2/584). Using a Monte Carlo simulation [7], visualized in Figure 1, forms that violate coronal sonorant OCP are shown to be underrepresented in modern Samoan, given the distribution of final consonants in POc.

Moreover, I find that OCP[+COR,+son] is also active in Samoan root phonotactics. Specifically, a probabilistic constraint-based phonotactic model [UCLA Phonotactic Learner; 8] was trained on 1600 Samoan roots from Milner's Samoan Dictionary [6]. The resulting model assigns significant weight to the constraint OCP[+COR,+son]. This finding is compatible with the proposal that markedness effects in reanalysis are restricted to those already active in the language.

These results are confirmed using a model of reanalysis implemented in Maximum Entropy Harmonic Grammar [Maxent; 9]. The model is iterated to simulate the cumulative effects of reanalyses over time. In other words, at each "generation", the learner induces a grammar based on input data and then uses this grammar to generate data that is passed down to the next generation. Two models are compared: 1) a baseline model that is purely frequency-matching, and 2) a markedness-biased model in which the constraint OCP[+COR, +son] is biased to have high weight using the method laid out by Wilson [10]. I find that the markedness-biased model performs significantly better than the purely distributional baseline model. In sum, the Samoan data supports the view that reanalysis (and more generally morphophonology) is guided both by the statistical patterns that learners encounter and by principles of markedness.

(1) *Ergative suffix allomorphy in Samoan*

ERG	STEM	SUFFIXED	GLOSS
a	rere	rere-a	'to take'
ina	iloa	iloa-ina	'to see, perceive'
tia	pulu	pulu-**t**ia	'to plug up'
sia	laka	laka-**s**ia	'to step over'
ŋia	tutu	tutu-**ŋ**ia	'to light a fire'
fia	utu	utu-**f**ia	'to draw water'
mia	inu	inu-**m**ia	'to drink'
lia	tatau	tatau-**l**ia	'to hang up'
na[1]	ʔai	ʔai-**n**a	'to eat'
ʔia	momo	momo-**ʔ**ia	'to break in pieces'

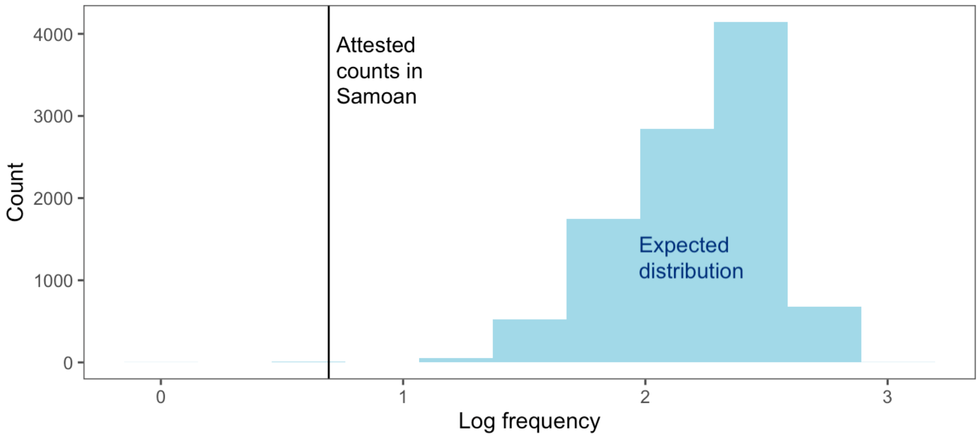


Figure 1: Attested [puli-na]/[puni-lia] words vs. expected distribution from POc

References

[1] Albright, A. (2002). *The identification of bases in morphological paradigms* [PhD thesis, UCLA].

[2] Albright, A. & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/ experimental study. *Cognition,* 90(2), 119–161.

[3] Mosel, U. & Hovdhaugen, E. (1992). *Samoan reference grammar*. Scandinavian Univ. Press.

[4] Pawley, A. (2001). Proto Polynesian *-CIA". In J. Bradshaw & K. L. Rehg (eds.), *Issues in Austronesian Morphology: A festschrift for Byron W. Bender*. Pacific Linguistics.

[5] Blust, R. and Trussel, S. (2020) *Austronesian Comparative Dictionary*. https://www.trussel2.com/ACD

[6] Milner, G. B. (1966). *Samoan Dictionary; Samoan-English, English-Samoan*. ERIC, 1966

[7] Mooney, C. Z. (1997). *Monte Carlo Simulation*. Thousand Oaks, CA: Sage Publications.

[8] Hayes, B. & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry,* 39(3), 379–440.

[9] Goldwater, S. & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In J. Spenader, A. Eriksson, & Ö. Dahl (eds.), *Proceedings of the Stockholm workshop on variation within Optimality Theory* (pp. 111-120). Stockholm University, Sweden.

[10] Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science, (30)*5, 945–982.

[1] Note that when the ergative suffix starts with /n/ the allomorph is /na/ rather than /nia/

The acquisition of Cantonese phonotactics

Tang Kin Man Carmen[1] & Regine Lai[1]
[1]*The Chinese University of Hong Kong (Hong Kong SAR)*
tangkm.carmen@link.cuhk.edu.hk, ryklai@cuhk.edu.hk

This work investigates infants' learning mechanisms for Cantonese phonotactics, and whether their learning trajectory of phonological structures can be predicted by structural complexity and naturalness of phonotactic constraints. Structural complexity can be defined by the factors of distance (i.e. local vs long-distance) and the window size of restricted sequences of sounds. Window size can be thought of as *n*-gram, the larger the window size, the longer the sequence, and the more complex it is. For example, English phonotactics allows the sequences of birgrams such as *st*- and trigrams *str*- which has window size of 2 and 3 sounds respectively; but not **sr*- or **ftr*-. Cantonese syllables are simpler than those in English, only (C)V(C) structures are legal, and a number of phonotactic gaps were identified [1] (shown in (1) – (4) below). These gaps allow us to compare the acquisition of syllables that are of different structural complexity.

<u>Low structural complexity</u>:

1) [bigram, local] Labial onsets cannot be followed by a front rounded vowel (e.g. */byu/)
2) [bigram, local] Coronal onsets cannot be followed by /uː/ (e.g. */tu/)

<u>Higher structural complexity</u>:

3) [trigram, non-local] Labial onset cannot co-occur with labial coda (e.g. */bab/)
4) [trigram, local] /ɔː uː/ cannot occur between coronal onsets and codas (e.g. */tot/)

The first 2 phonotactic rules are considered to be less complex because they are local dependencies with the window size of 2. (4) is considered more complex even though it is a local dependency but its window size is 3. (3) is regarded more complex because it is a non-adjacent dependency that restricts the cooccurrence of the onset and coda of the syllable while the vowel in the medial position is variable.

Obligatory Contour Principle (OCP) is a commonly found restriction for natural languages, which prevents homorganic sounds from occurring in sequence. For example, word-likeness results show that the violation of root consonant cluster in Arabic led to lower word likeness ratings [2]. (1), (3) and (4) violate OCP. (1) can be viewed as a violation because of sharing a [+labial] feature, while (3) and (4) violate OCP at the consonant tier, because the initial and final consonant share the same place of articulation. (2), on the other hand, violates a less natural phonotactic constraint, because it is more natural for coronal onsets to be followed by front vowels; violation of (2) would therefore lead to a more marked structure [3, 4].

The coronal-labial contrast in the Cantonese phonotactic gaps may also contribute to the results. This is because the backness of the vowel is influenced by the coronal onset, whereas they are not affected by labial onsets [3]. Moreover, in an examination of coronal and labial onset frequencies using Cifu [5], a Cantonese frequency lexicon, among all the monosyllabic word frequencies, coronal onsets have 328,132 occurrences, while labial onsets only have 104,041. This shows that Cantonese-learning infants have more exposure to the legal structure involving coronal onsets.

As for the learning trajectory of structural patterns, infants as young as 5 months old show evidence for rule learning when there are multimodal cues [6]. 7-month-olds can track patterns equivalent to the complexity of trigrams: for example, they can distinguish ABB patterns from ABA [7]. Headturn experiment results show that both 7- and 9-month-old age groups can segment words based on local transitional probability (TP) and stress cues respectively [8]. Infants' ability to detect non-adjacent dependency emerges at around 12 months old [9]. Yet, in another study the 12-month-old group failed to learn non-adjacent dependencies like *aXc*, while 15-month-olds succeeded [10]. Thus, to examine the learning trajectory of Cantonese phonotactic rules, which involve both local and non-local dependencies at different window sizes, we tested Cantonese-learning infants in 5 cross-sectional age groups (5, 7, 9, 12, 14 months) ($N = 150$).

We tested the infants on a head-turn preference paradigm. 32 pseudowords corresponding to the four Cantonese phonotactic gaps were created, four legal-illegal word pairs for each phonotactic gap. The factors of *locality* and *window size* were tested within-subject. Each trial contains four exemplars of the same type (e.g. *Labial bigram – legal*), making there 8 trials in total. Each trial was set to have a maximum looking time at 80s. The order of the trials were counterbalanced. The infants' looking time were recorded for analysis.

A linear model was built with the fixed factors *Age* (5 levels: 5, 7, 9, 12 &14m), *Legality* (2 levels: legal *vs* illegal), *Window size* (2 levels: bigram *vs* trigram), and *Place* (2 levels: coronal and labial), and logged looking time as dependent variable. There is a significant effect for *Place* F(1, 1015) = 3.974, $p = .046$) and *Age* (F(4, 145) = 2.649, $p = .036$), also an interaction between *Legality* and *Place* (F(1, 1015) =

12.279, *p* = .0005). Post-hoc pairwise comparisons reveal significant looking time differences for coronal gaps at 9-month-olds (*bigram*: est = .463, se = .224, *t* = 2.063, *p* = .039) and 7-month-olds (*trigram*: est = .45, se = .224, *t* = 2.008, *p* = .045). There were also looking time differences for the trigram labial gap at 5- (est = -0.433, se = .224, *t* = -1.93, *p* = .054) and 9-month-olds (est = -0.423, se = .224, *t* = .1.89, *p* = .060).

Since looking time results show different preferences for legality for coronal and labial gaps, separate models were built for the two types of gaps. For both models, *Legality*, *Age*, and *Window size* were used as fixed factors. In the coronal model, significant effect for Legality was found (F(1, 435) = 4.954; *p* = .027). Infants show novelty preference for the coronal gaps, and according to the familiarization model by Hunter & Ames (1988), novelty preference signifies a more mature stage of learning than familiarity preferences (as cited in [11]). For labial gaps items, significant *Legality* (F(1, 435) = 6.961; *p* = .009) and *Window size* (F(1, 435) = 6.301; *p* = .012) effects were found. There were also a marginal effect for *Age* (F(1, 145) = 2.30, *p* = .062) and a three-way interaction for *Legality x Age x Window size* (F(4, 435) = 2.032, *p* = .089). Post-hoc pairwise comparisons reveal marginal looking time differences for trigram at 5- (est = -.432, SE = .225, *t* = 1.921, *p* = .055) and 9-month-olds (est = -.428, SE = .225, *t* = 1877, *p* = .061). Infants showed a familiarity preference towards legal stimuli for the labial trigram gap. With reference to Hunter & Ames' (1988) familiarization model, this may reflect that the structure learned is still considered complex relative to age.

The opposite direction of preferences for coronal and labial gaps may represent the different learning trajectory for coronal and labial rules, while these differences in learning trajectory can possibly be attributed to the complexity differences. The successful acquisition of the gaps with trigrams labial gap may indicate that window size is outweighed by other factors, as bigram restrictions are not necessarily learned earlier than trigrams. Overall, no looking time difference was found for the bigram labial gap. Even though this pattern violates OCP, infants' looking time data does not show legality contrasts. On the other hand, infants successfully distinguish patterns that violate OCP at the consonant tier level (i.e. trigram labial, trigram coronal). This may indicate the gradient nature of OCP violations. The opposite direction of looking preferences between the two trigram gaps may reflect the differential encoding for true local trigram patterns and non-local patterns (i.e. aXb, where X is variable). Finally, the data does not rule out frequency effect for coronal gaps. It is possible that the accumulative effect of local dependency and OCP violation makes infant notice the trigram coronal gaps at 7 months old, while violation at local dependency alone (bigram coronal gap) is more recognizable than solely tier-based OCP violation (trigram labial gap).

References

[1] Kirby, J. P., & Yu, A. C. (2007, August). Lexical and phonotactic effects on wordlikeness judgments in Cantonese. In *Proceedings of the International Congress of the Phonetic Sciences XVI* (Vol. 13891392).

[2] Frisch, S. A., & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place in Arabic. *Language*, 91-106.

[3] Flemming, E. (2003). The relationship between coronal place and vowel backness. *Phonology*, *20*(3), 335-373.

[4] Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language learning and development*, *1*(3-4), 289-316.

[5] Lai, R., & Winterstein, G. (2020, May). Cifu: a frequency lexicon of Hong Kong Cantonese. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 3069-3077).

[6] Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Developmental science*, *12*(4), 504-509.

[7] Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77-80.

[8] Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, *39*(4), 706.

[9] Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive psychology*, *67*(3), 130-150.

[10] Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183-206.

[11] Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice*, *13*(4), 341-348..

Velar palatalization in Italian: Lexical stress induces resistance to sound change

Bowei Shao[1,2], Anne Hermes[2], Philipp Buech[2] & Maria Giavazzi[1]

[1] *Département d'Etudes Cognitives, École Normale Supérieure, Université PSL (France),*
[2]*Laboratoire de Phonétique et Phonologie (UMR 7018), CNRS/Sorbonne Nouvelle (France)*
bowei.shao@ens.psl.eu, anne.hermes@sorbonne-nouvelle.fr, philipp.buech@sorbonne-nouvelle.fr, maria.giavazzi@ens.psl.eu

Introduction: Palatalization is the process through which a velar stop, /k, g/, is fronted to a palatal/palato-alveolar affricate or fricative. It applies more frequently before high front vowels than before other vowels. In Romance languages, the origins of this phonological process are to be found in Late Latin, since Latin did not have any palatal consonants in its earlier stages: this process is known as the 2nd Romance palatalization, and it occurred after the 5th century CE **[1,2]**. It occurred before all front vowels, both root internally and at the morpheme boundary, and independently of morpheme boundaries and stress position **[3]**. Palatalization of velars in contemporary Italian, instead, takes place at the boundary between the root and inflectional (or derivational) suffixes in /-i/. In masculine nouns and adjectives, palatalization before the plural ending /-i/ is predominantly stress-conditioned: it occurs in words with antepenultimate stress such as [ˈko.mi.ko] – [ˈko.mi.tʃi] 'comedian'-'comedians', while it is much rarer in words with penultimate stress such as [ka.ˈdu.ko] – [ka.ˈdu.ki] 'caducous'-'caducous *pl.*' **[4,5]**. The former case presents a stressed syllable FAR from the target of palatalization /k/, while in the latter the /k/ is directly POST-tonic. Whereas the perceptual and articulatory underpinnings of palatalization before high front vowels have been studied extensively **[6, 7]**, the phonetic bases of the stress-conditioned process are not well understood (cf. **[5]**). The aim of this study is to explore the articulatory (Basis1) and acoustic (Basis2) bases of this stress-conditioned process in Italian. In Articulatory Phonology, lexical stress is modulated by a tempo-spatial μ-gesture **[8]**. We postulate that the resistance of POST /k, g/ to palatalize is related to the μ-conditioned articulation of the stressed vowel directly preceding them. In POST, the μ-conditioned stressed vowel introduces coarticulatory resistance between the following consonant and the final vowel, which in turn favors a plosive articulation (Basis1). The large opening gesture of the stressed vowel leads to a later target achievement of the following consonant, causing a longer closure duration. The resulting longer closure duration decreases the perceptual salience of the release with respect to the total duration of the consonant, which in turn favors a plosive categorization (Basis2).

Method: First, an ***acoustic study*** was conducted on 18 speakers. Target words were trisyllabic nonce words, structured /C$_1$V$_1$.C$_2$V$_2$.C$_3$V$_3$/, differing solely by the position of stress on the first or the second syllable (e.g., /ˈpi.ta.ki/, /pi.ˈta.ki/). The nonce words were designed to compare how the target consonants /k, g, tʃ, dʒ/ (in C$_3$ position) were produced in both FAR and POST contexts. The V$_3$ position is occupied by /i/. C$_1$V$_1$.C$_2$V$_2$ sequences were /pita/, /fesa/, /pufa/ /tipa/ and /suta/. For example, in the nonce words /ˈpi.ta.ki/ and /pi.ˈta.ki/, C$_3$ /k/ was in FAR and POST contexts respectively. The nonce words were written in their orthographic forms with lexical stress marked by an accent (e.g., pítachi, pitáchi). They were embedded in a carrier phrase "*Dimmi __ di nuovo*" and randomized. The 18 speakers repeated the word list three times, yielding 3328 tokens. Second, an ***articulatory study*** (EMA, AG501) is currently on-going. So far, we have recorded four speakers and we present one speaker here. The structure of the nonce words is the same as in the acoustic study, except that the V$_2$ position is occupied by both /a/ and /e/. They were embedded in a carrier phrase "*Pimpa parte da __ la mattina presto*" and randomized. More data will be available at the conference (aim: 15 speakers).

Results: The ***acoustic analysis*** of stressed vowels shows that the μ-conditioned stressed vowels are more than twice as long ($\bar{x}$ = 185 ms, σ = 45) as the unstressed ones ($\bar{x}$ =82ms, σ=29). GAMM analysis predicts that the shorter the unstressed vowel [a], the lower its F1 and the higher its F2 (Fig. 1). The analyses of C$_3$V$_3$ (e.g. [ki, gi, t͡ʃi, d͡ʒi]) in both conditions show that V$_3$ could not be responsible for the blocking of velar palatalization as it has virtually the same acoustic characteristics in both FAR and POST conditions. However, C$_3$ has longer closure duration in

POST compared to in FAR. The preliminary ***articulatory analysis*** based on one speaker shows that the μ-conditioned vowel has longer and larger tongue body movement (Fig. 2). The trajectory from V_2 to V_3 (i.e. [a-i], [e-i]) is significantly different in POST and FAR. Even when normalizing the duration from V_2 to V_3, target achievement of V_3 is observably earlier in FAR than in POST.

Discussion: The articulatory results suggest that the stressed vowel causes resistance to the coarticulation between the following consonant and the final vowel, supporting Basis1. As shown in Fig 3, the μ-conditioned vowel has longer and larger gestural activation intervals (compare *a* → *e,* to *a* → *c*). More importantly, the μ-conditioned vowel has spillover effects on the following consonant and vowel gestures **[9]**: the upcoming gestures are expected to have larger displacements, which in turn results in a delayed maximum constriction of [i], as can be seen in Fig. 2. The acoustic results confirm that the closure duration is modulated by the position of stress, which may serve as the acoustic grounding for Basis2.

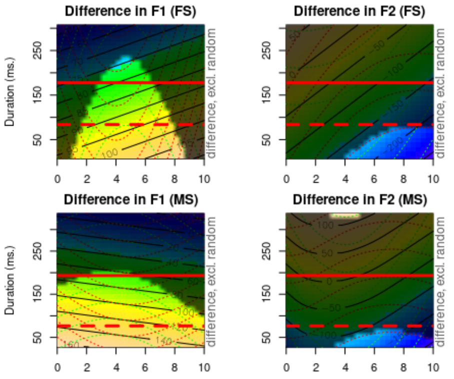


Figure 1. Difference between stressed [a] and unstressed [a] over time (x-axis) modulated by duration (y-axis). The shaded region indicates the area where the difference is non-significant. The red solid and dashed lines indicate the mean duration of stressed and unstressed [a] respectively. FS and MS are for female and male speakers respectively.

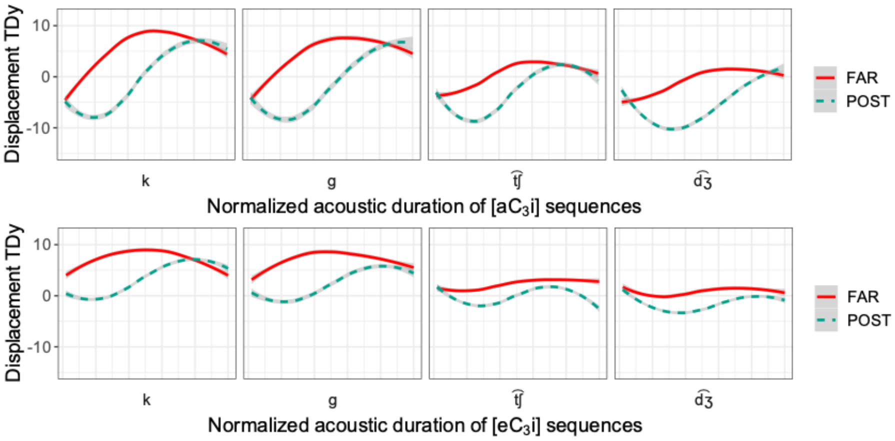


Figure 2. Tongue dorsum displacement trajectory of $V_2C_3V_3$ in y-axis according to stress conditions. Red solid lines represent trajectory when stress is FAR from C_3, green dotted lines represent POST stress C_3. X-axis represents normalized duration from the acoustic onset of V_2 ([a] and [e]) to the acoustic offset of V_3 [i].

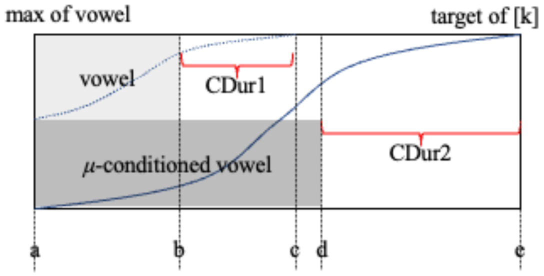


Figure 3. Schematic representation of μ-conditioned/normal vowel and following [k]. The height/width of the boxes represent tongue dorsum displacement/release duration; the blue lines represent the tongue dorsum trajectory in y-axis; the gray boxes represent the second half of the acoustic duration of concerned vowels.

References

[1] Petrosino, R. & Calabrese, A. (2022). Palatalization in Romance. In C. Gabriel, R. Gess & T. Meisenburg (Eds.), *Manual of Romance Phonetics and Phonology*, Berlin, Boston: De Gruyter, pp. 173-214.

[2] Rohlfs, G. (1966). *Grammatica storica della lingua italiana e dei suoi dialetti*, vol. 1: *Fonetica* (trad. Salvatore Persichina), Torino: Einaudi.

[3] Celata, C. & Bertinetto, P.M. (2005). Lexical Access in Italian: Words with and without Palatalization, in *Lingue e linguaggio*, 2: 293-318.

[4] Giavazzi, M. (2009). On the application of velar palatalization in Italian. Poster presented at the 17th Manchester Phonology Meeting.

[5] Giavazzi, M. (2010). *The phonetics of metrical prominence and its consequences on segmental phonology*. Ph.D. dissertation; MIT Department of Linguistics and Philosophy.

[6] Ohala, J. (1992). *What's cognitive, what's not, in sound change; in Kellermann, Morrissey, Diachrony within synchrony: language history and cognition*. Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 14, pp. 309–355, Frankfurt: Lang.

[7] Guion, S.G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica*, 55 (1-2):18–52.

[8] Saltzman E, Nam H, Krivokapic J, Goldstein L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. Proceedings of the 4th International Conference on Speech Prosody (Speech Prosody 2008), Campinas, Brazil, p. 175-84.

[9] Katsika, A., & Tsai, K. (2021). The supralaryngeal articulation of stress and accent in Greek. *Journal of Phonetics*, *88*, 101085.

Can we use categories when investigating interlanguages?

Simona Sbranna[1], Aviad Albert[2] & Martine Grice[3]
[1,2,3]IfL Phonetics – University of Cologne (Germany)
{s.sbranna, a.albert, martine.grice@}uni-koeln.de

***Keywords**: L2 prosody, continuous phonetic parameters, periodic energy, categorical analysis.*

Introduction. The prosodic marking of information status within noun phrases (NPs) reportedly differs in German and Italian. According to [1, 2], German speakers deaccent post-focal given information, whereas Italian speakers accent the second word of the noun phrase, regardless of information status. This also appears to be the case for Italian learners of German when speaking their L2. However, a categorical analysis of accentuation might not be appropriate, or even possible, for interlanguages, in which categories are constantly updated.

Methods. We elicited two different information structures – given-new (GN) and new-given (NG) – in NPs composed of a disyllabic noun and a disyllabic adjective in L1 German, L1 Italian, and L2 German. We performed a continuous analysis to explore speakers' modulation of F0, focussing on the alignment of the peak and following fall in pitch, and periodic energy mass as a measure for prosodic strength [3]. Mass is the sum integral of the duration and intensity of the periodic component within syllables. Its values are computed with respect to other syllables in the NP, such that weak mass (indicating prosodic attenuation) corresponds to values below one, while strong mass (indicating prosodic enhancement) corresponds to values above one. We complemented this analysis with a ToBI annotation of accentuation and pitch accent types following [4] for Italian and [5] for German and attempt to interpret the F0 and mass values in the interlanguage in categorical terms, both in relation to accentuation (presence or absence of accent) and to accent types.

Results. Averaged F0 contours for the two conditions across language groups are presented in Fig. 1. Aggregated mass values are presented for the third syllable of the NPs (the stressed syllable of the new word in GN and of the post-focal given word in NG) in Fig. 2. Results for L1 Italian show that prosodic marking of information status is not realised through deaccentuation on the second word as in L1 German, as demonstrated by strong mass on post-focal given material (Fig. 2, NG). However, contrary to previous results, Italian speakers do in fact mark information status prosodically. This is achieved by modulating F0 on the first word, with an F0 peak aligned earlier when this word is new (NG condition, Fig. 1). This difference can be interpreted as two different pitch accent types: the first word of the NP can be described as bearing (L+)H* for GN and H*+L for NG, with the trailing L tone added in the latter case to highlight that H is very close to the syllable onset and that the F0 fall occurs mainly on the first syllable. The accent on the second word is similar across conditions and can be analysed as L*. In German L1, results of the acoustic analysis are in line with the literature and provide evidence for the deaccentuation of post-focal given material (with weak mass on the second word of NG, Fig. 2) and the tendency to align an F0 peak with new or focussed elements (Fig. 1, with a "hat pattern" in GN, i.e. a high F0 plateau extending to the stressed syllable of the second word). This peak alignment is generally interpreted as a H* or L+H* pitch accent on the first word of NG, with the post-focal given element being deaccented, while the hat pattern found in GN can be described as H* for the first word and H* or (H+)!H* for the second word. In L2 German, learners prosodically mark information status by modulating the alignment of F0 on the first word analogous to the pattern in their L1 (Fig. 1). Regardless of the information structure, the second word in L2 German NPs was produced with flat F0, which could be interpreted as L*, as can be found in postfocal accents in longer constituents in Italian. This second word was also produced with weak mass values, which could be interpreted as deaccentuation, as in L1 German, but only in the NG condition (Fig. 2). In the L2, the more salient marking of the first word in NG, with an earlier F0 peak, can lead to the perception of lower prominence on the following given word although in isolation these final words in the NP are very similarly produced across conditions.

Conclusion. The current study of continuous parameters revealed patterns for L1 Italian learners of L2 German which did not emerge in previous categorical analyses. Moreover, we found that a categorical interpretation could describe the two L1s well, but was difficult to apply to the L2. Labelling phonological categories entails some degree of subjectivity due to the annotator-specific perception of meaning and expectations based on their native language. As a result, different annotators can make different choices, and the individual-specific bias is even more problematic when labelling an L2. Thus, an investigation of the modulations of continuous acoustic parameters can offer a deeper understanding of linguistic phenomena by providing acoustic evidence for a categorical description. Therefore, the two approaches, continuous and categorical, should complement each other especially when analysing complex and dynamic systems like interlanguages, where categories undergo a continuous process of restructuring based on the input and feedback that learners receive.

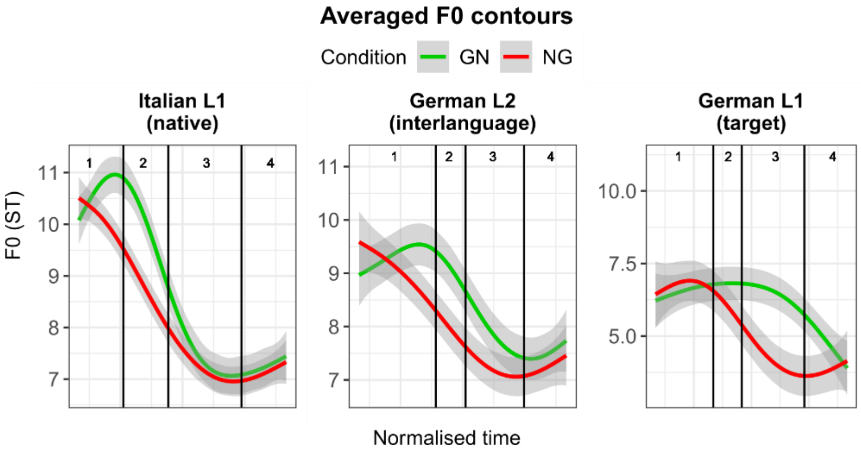


Figure 1. Averaged F0 contours pooled across speakers for each language group. The y-axis shows F0 in semitones, while the x-axis shows normalised time aligned at the boundary between the two words of the noun phrase. Syllables of the noun phrase are numbered from one to four and syllable boundaries are marked by vertical black lines. The grey area around the contours represents the standard error and contours are colour-coded according to their information structure condition: green for given-new (GN) and red for new-given (NG).

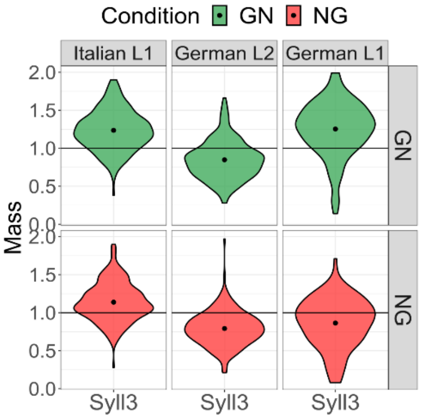


Figure 2. Aggregated values of mass for syllable three across language groups. Mean values are represented by black dots. Information structure conditions are colour-coded and positioned on two separate rows: green for given-new (GN, upper row with Syll3 being a new item) and red for new-given (NG, bottom row with Syll3 being a given item).

References

[1] Swerts, M., Krahmer, E. & Avesani, C. (2002), Prosodic marking of information status in Dutch and Italian: a comparative analysis, *Journal of Phonetics*, 30, 4, 629-65.

[2] Avesani, C., Bocci, G., Vayra, M. & Zappoli, A. (2015), Prosody and information status in Italian and German L2 intonation, *Il parlato in [italiano] L2: aspetti pragmatici e prosodici /[Italian] L2 Spoken Discourse: Pragmatic and Prosodic Aspect*, Milano, Franco Angeli, 93-116.

[3] Albert, A., Cangemi F., Ellison T. M. & Grice M. (2020). *ProPer: PROsodic analysis with PERiodic energy*. OSF. March 4. doi:10.17605/OSF.IO/28EA5.

[4] Grice, M., D'imperio, M., Savino, M., & Avesani, C. (2005). Towards a strategy for ToBI labelling varieties of Italian. In S.-A. Jun (Ed.), *Prosodic Typology and Transcription: A Unified Approach*. Oxford, Oxford University Press, 362–389.

[5] Grice, M., Baumann, S., & Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford, Oxford University Press, 55-83.

How does focus-induced prominence influence realization of edge tones and segmental anchoring in Seoul Korean – A preliminary report

Richard Hatcher[1], Hyunjung Joo[2,1], Sahyang Kim[3,1], Taehong Cho[1]
[1]Hanyang Institute for Phonetics & Cognitive Sciences of Language, Hanyang University (Korea),
[2]Rutgers University (USA), [3]Hongik University (Korea)
richard.j.hatcher.jr@gmail.com, hyunjung.joo@rutgers.edu, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

In Seoul Korean, the intonational tune comprises edge tones assigned to Accentual Phrases (#LH...LH#) and a boundary tone associated with the final syllable of the Intonational Phrase [1, 2]. Accentual Phrases often have two rises (#LH…LH#) at their edges. Unlike many European languages, Korean does not have an accentuation system that involves a placement of pitch accent on a stressed syllable; instead, phrasing is used, with focus triggering the insertion of a prosodic boundary and optionally dephrasing post-focal elements [1, 5, 6]. While this has been proposed as the phonology of focus in Korean [1, 3, 5], however, relatively little is known about the phonetic effects of focus on tunes and tone-segment alignments in Seoul Korean, particularly in shorter sentences with fewer Accentual Phrases. This study therefore investigates how focus affects the tune of short intonational phrases in Seoul Korean.

Production experiment: The production experiment involved 14 young adult native speakers of Seoul Korean, 7 females and 7 males. Participants read sentences containing two monosyllabic target words with codas of differing sonority: ***pam*** ('chestnut; night') and ***pap*** ('cooked rice'). The phrasal position of the target words was varied to observe the interaction between focus effects and position, with the target words appearing in one of three positions: IP-initial, IP-medial, and IP-final. The target word was followed by ***twiɛ(ta)*** ('behind') in the IP-initial and medial contexts, so that the focus occurred either on the target word (focal condition) or on ***twiɛ(ta)*** (prefocal condition), as shown in the example sentence. In IP-final position, the focus occurred either on the target word (focal condition) or on the preceding word ***ʌnni*** ('sister') (post-focal condition). The sonorant portions of all words in the sentence were segmented, and f0 was measured at nine equidistant time points within each word using the Straight algorithm. Three generalized additive mixed models (GAMM) were fitted to the data, one for each of the phrasal contexts, to investigate how the tune and tone-segment alignment patterns may differ between the focused and unfocused conditions in the three prosodic positions.

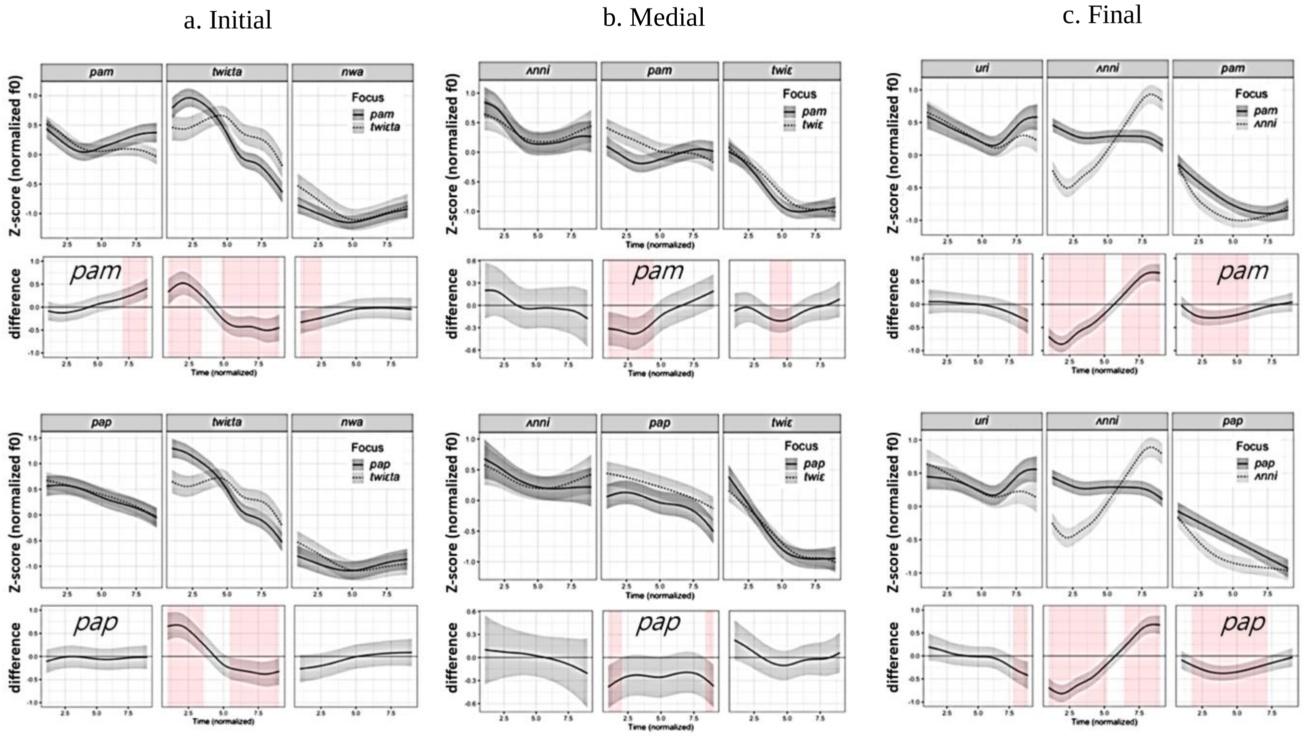
Table 1: An example set of target words in carrier phrases with varying focus and boundary conditions. Target words are underlined and focused elements are in bold.

Phrase	Prefocal	Focal	Postfocal
Initial	*ani. # pam* ***twi****ɛta nwa. #* No. Put it **behind** the chestnut.	*ani. #* ***pam*** *twiɛta nwa. #* No. Put it behind the **chestnut**.	--- ---
Medial	*ani. # ʌnni pam* ***twi****ɛ. #* No. **Behind** sister's chestnut.	*ani. # ʌnni* ***pam*** *twiɛ. #* No. Behind sister's **chestnut**.	*ani. #* ***ʌnni*** *pam twiɛ. #* No. Behind **sister**'s chestnut.
Final	--- ---	*ani. # uri ʌnni* ***pam****. # twiɛta nwa. #* No. My sister's **chestnut**. Put it behind.	*ani. # uri* ***ʌnni*** *pam. # twiɛta nwa. #* No. My sister's **chestnut**. Put it behind.

Results: The resulting plot smooths and difference plots are shown in Fig. 1 below. In the IP-initial context (Fig.1a), the difference in f0 predicted by the GAMM between the focal *vs.* prefocal conditions on the target words was mainly observed during the following ***twiɛta*** 'behind' and not during the target words themselves. The f0 peak during the following syllable occurred earlier with a larger magnitude in association with a H tone for phrase-initial focus. In contrast, in the IP-medial context (Fig.1b), focus effects on the target words were phonetically evident primarily during the target words with a lowered f0 trough in the focused condition in association with a L tone (although marginal for *pap).* In the IP-final context (Fig.1c), the focus effect on target words was evident during the target words in association with a

L% (boundary) tone being realized later when focused. The study also observed that the preceding word ***ʌnni*** ('sister') showed the clearest effects of focus in the experiment, with both the initial trough and following peak undergoing substantial scaling. Finally, although the coda's sonorancy showed some microscopic difference (see upper vs. lower panels of Fig.1a), the general tonal targets were realized in a similar fashion under focus.

Discussion: Overall, these findings provide valuable insights into the phonetic realization of focus in Seoul Korean and emphasize the importance of understanding the complex interplay between focus and prosodic structure in shaping the language's intonational contours. Furthermore, the observed differences in tune and tone-segment alignment patterns across various prosodic positions and tonal and segmental contexts have significant implications for developing models of intonational phonology in Seoul Korean. Future research can expand on these results by investigating how these patterns generalize across target words and sentences of varying lengths.


a. Initial

b. Medial

c. Final

Figure 1: Visualization of non-linear smooths (above) and difference plots (below) in (a) phrase-initial, (b) phrase-medial and (c) phrase-final positions. Grey ribbons represent pointwise 95%-confidence intervals of f_0. Pink vertical bars in the difference plots signify which portions of the two smooths significantly differ from one another.

References

[1] Jun, S.-A. (1996). *The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure*. New York ; London: Garland Pub.

[2] Jun, S.-A. (1998). "The Accentual Phrase in the Korean prosodic hierarchy," *Phonology*, 15(2), 189–226. doi: 10.1017/S0952675798003571.

[3] Jun, S.-A. (1996). "Influence of microprosody on macroprosody: A case of phrase initial strengthening," *UCLA Work. Pap. Phon.* (92), 97–116.

[4] Jun, S.-A. (2005). "Korean Intonational Phonology and Prosodic Transcription," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford University Press.

[5] Jun, S.-A. & Lee, H.-J. (1998). "Phonetic and phonological markers of contrastive focus in Korean," in *5th International Conference on Spoken Language Processing (ICSLP 1998)*. doi: 10.21437/ICSLP.1998-151.

[6] Jeon, H.-S. & Nolan, F. (2017). "Prosodic Marking of Narrow Focus in Seoul Korean," *Lab. Phonol.*, 8(1). doi: 10.5334/labphon.48.

Individual differences in perceptual cue weighting: behavioral, neurophysiological, and social considerations

Alan Yu

University of Chicago (USA)
aclyu@uchicago.edu

Introduction: Speech categories are defined by multiple acoustic dimensions and their boundaries are generally fuzzy and ambiguous. During speech perception, listeners must determine which cues are relevant and their relative importance. Despite an increasing number of studies documenting systematic and consistent variability in perceptual cue weighting across listeners within the same speech community (e.g., Clayards 2018, Idemaru et al 2012, Kong & Edwards 2016, Schertz et al 2015, Shultz et al. 2012), the processes underlying the variability remain largely underexplored. In this talk, I present results from two studies exploring the behavioral, neurophysiological, and social origins of individual variability in perceptual cue weighting.

Study 1: Recent studies suggest that individual differences in speech processing may stem from differences in the very early stages of speech processing (e.g., Kapnoula & McMurray 2021, Ou & Yu 2021). For example, the nature of VOT categorization among English speakers is found to correlate with how faithfully subcortical responses encode VOT differences, with listeners who showed more uncertainty in categorization exhibiting less faithful encoding of the acoustic differences (Ou & Yu 2021). The present study expands on Ou & Yu 2021 and investigated the role of subcortical encoding as a source of individual variability in cue weighting by focusing on English listeners' frequency following responses (FFR) to the tense/lax English vowel contrast varying in spectral and durational cues. We found that listeners differed in early auditory encoding with some encoding the spectral cue more veridically than the durational one, while others exhibited the reverse pattern. These differences in cue encoding further correlate with behavioral variability in cue weighting, suggesting that specificity in cue encoding across individuals modulates how cues are weighted in downstream processes.

Study 2: Previous studies have found that socio-indexical information influences how listeners process the speech signal (e.g., Strand 1999, Hay et al. 2006). This study investigates specifically how a listener's perception of a speaker's socio-indexical and personality characteristics influences the listener's perceptual cue weighting. In a matched-guise study, three groups of listeners classified a series of gender-neutral /b/-/p/ continua that vary in VOT and F0 at the onset of the following vowel. Listeners were assigned to one of three prompt conditions (i.e., a visually male talker, a visually female talker, or audio-only) and rated the talker in terms of vocal (and facial, in the visual prompt conditions) gender prototypicality, attractiveness, friendliness, confidence, trustworthiness, and gayness. Male listeners and listeners who saw a male face showed less reliance on VOT compared to listeners in the other conditions. Listeners' visual evaluation of the talker also affected their weighting of VOT and onset F0 cues, although the effects of facial impressions differ depending on the gender of the listener.

Conclusions: These findings highlight the fact that the mechanisms underlying individual variation in perceptual cue weightings are multi-dimensional. While listeners may show differential cue encoding, which then affects the reliability and weighting of certain cues that support phonological contrasts, higher order indexical information may nonetheless influence how acoustic cues are utilized in speech processing. The significance of these findings for phonetic theories, theories of sound change, and the nature of phonological knowledge will be discussed.

References

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America* 144, EL172–EL177.

Hay, J., Warren, P., and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*. 34, 458–484. doi: 10.1016/j.wocn.2005.10.001

Idemaru, K., Holt, L. L., and Seltman, H. (2012). Individual differences in cue weight are stable across time: the case of Japanese stop lengths. *The Journal of the Acoustical Society of America* 132, 3950–3964. doi: 10.1121/1.4765076

Kapnoula, E.C. & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language*, 223, 105031.

Kong, E. J. & Edwards, J. (2016) Individual differences in categorical perception of speech: Cue weighting and executive function. Journal of Phonetics 59, 40–57.

Ou, Jinghua and Alan C. L. Yu. (2021). Neural Correlates of Individual Differences in Speech Categorization: Evidence from Subcortical, Cortical, and Behavioral Measures. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2021.1980594

Schertz, J., Cho, T., Lotto, A. & Warner, N. (2015) Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52, 183–204.

Shultz, A. A., Francis, A. L. & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132, EL95–EL101.

Strand, Elizabeth A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Psychology* 18: 86-99.

Poster Presentations

Day 1

(Friday, May 26, 2023)

Investigating MMN Responses to Pitch Contrasts in Monolingual and Bilingual Speakers of Tonal Languages

Chun-Hsien Hsu[1], Wen-Chun Huang[1] & Tong-Hou Cheong[1]

[1]*National Central University (Taiwan)*
kevinhsu@ncu.edu.tw, wjuneh@gmail.com, kenc1998@gmail.com

Among the myriad aspects of language processes, correctly perceiving suprasegmental information is vitally important. However, the majority of the studies of language process assumes a monolingual mindset. Some researchers have highlighted that there are needs to improve the understanding of language processes in multilingual speakers, because working with different languages in everyday lives might require special needs[1]. To contribute to this lack, the present study explored whether lexical tone contrasts of syllables would be processed in the same way regardless of the language background of participants?

In the present study, participants' pitch perception was evaluated in an event-related potentials (ERP) experiment. One group of participants was Mandarin Chinese speakers, and the other one was Hailu Hakka-Mandarin Chinese bilinguals. We measured the mismatch negativity (MMN) which is a unique ERP activity to index discernible changes of acoustic features in a streams of sound and is not influenced by attention[2, 3]. Kuo et al.[4] have noted that there were theoretical explanations for the influence of bilingual experiences on speech perception, such as , such as the cross-language transfer theory and the structural sensitivity theory. Both theories would expect that owing to the simpler phonological structure of Mandarin Chinese than Hailu Hakka, Mandarin speakers may be less sensitive to the phonemic features of Hakka syllables. Therefore, Mandarin speakers may exhibit reduced or insignificant MMN response to the T1/T3 contrast in Hakka syllables.

Seventeen native Mandarin speakers and sixteen Hakka-Mandarin bilinguals were recruited to participate in MMN experiments. Two lexical tones utilized here were the high level tone (T1) the low falling-rising contour tone (T3). One set of speech stimuli were two Mandarin syllables /zu/ with T1 and T3, and they are not real morphemes or words in Hailu Hakka. The other set were Hakka syllable /so/ with T1 and T3, and they are not real morphemes or words in Mandarin. There were four experimental blocks, and each block had five hundreds trials. In each trial, a syllable was presented over two loudspeakers at 70 dB. The stimuli lasted 350 ms with a 400 ms inter-trial interval. The four blocks were orthogonally assigned to one of two languages, and one of two conditions, including the MMN condition and the probability control condition. In the MMN condition, the T1 and T3 syllables were presented 100 and 400 times, respectively. In the control condition, the T1 and T3 syllables were presented with equal probability. During the experiment, participants were watching a movie without its sounds and subtitles, and their EEG signals were simultaneously recorded from 32 scalp electrodes.

A repeated-measure ANOVA model including the condition of probability (control and MMN conditions), language types (Mandarin syllables and Hakka syllables) and electrodes (six electrodes in the frontal scalp) as independent variables was applied to the data of each group. The dependent variable was the mean amplitude of ERPs to T1 syllables. In Hakka-Mandarin bilingual speakers, the main effect of conditions was significant, $F(1, 15) = 15.45$, $p = .001$, suggesting that syllables in MMN blocks yielded more negative activity than in control blocks. The interaction between conditions and language types was not significant, $F(1, 15) = .07$, $p = .796$, suggesting that the difference between MMN and control blocks did not vary across language types. In Mandarin speakers, the results yielded a significant main effect of conditions, $F(1, 16) = 22.01$, $p < .001$, and a significant interaction between conditions and language types, $F(1, 16) = 6.74$, $p = .019$, suggesting that the condition effects differed in terms of language types. Post-hoc tests with Bonferroni-Holm adjustments showed that Mandarin speakers' ERP activity to the syllable /zu1/ in the MMN block was significantly more negative than that in the control block (adjusted $p < .001$).

Finally, there was no condition effect on Mandarin speakers' ERPs to the Hakka syllable /so/ (adjusted p = .283).

In conclusion, the present study demonstrates that monolingual speakers might store a bank of phonological exemplars to perceive the tonal information of native syllables, whereas bilinguals have a more generalized representation of pitch height that allows them to process tonal information from different tonal languages. This study highlights the importance of considering the linguistic background of the participants when studying speech perception in tonal languages, and underscores the need for more research to better understand how linguistic experience can be used to improve language learning.

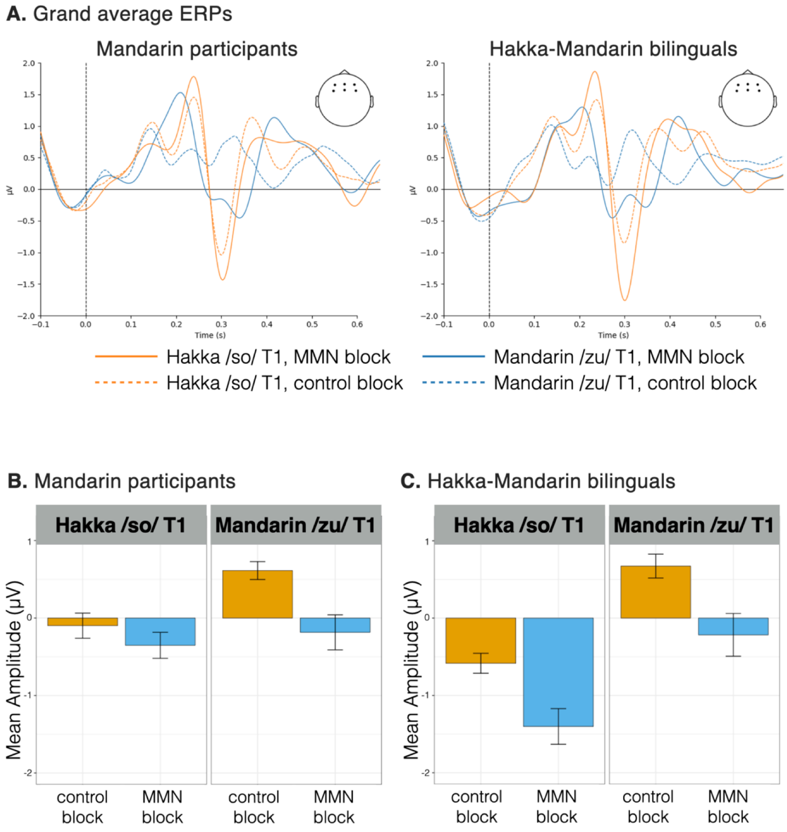


Figure 1. (A) Grand average ERPs elicited by syllables in MMN and control blocks within each participant group at electrodes of interest. (B) Bar plots of the mean amplitudes as a function of the condition of probability (control and MMN conditions), language type (Mandarin syllables and Hakka syllables) within Mandarin participants. The error bars show standard errors. (C) Bar plots of the mean amplitudes within Hakka-Mandarin bilingual participants.

References

[1] Garcia-Sierra, A., Ramirez-Esparza, N., Silva-Pereyra, J., Siard, J., & Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and Language*, 121(3), 194-205.

[2] Naatanen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and Mismatch Negativity. *Psychophysiology*, 30(5), 436-450.

[3] Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432-434.

[4] Kuo, L. J., Uchikoshi, Y., Kim, T. J., & Yang, X. Y. (2016). Bilingualism and phonological awareness: Re-examining theories of cross-language transfer and structural sensitivity. Contemporary Educational Psychology, 46, 1-9.

The realization of lexical tones in Sichuan opera

Jueyu Lu & Albert Lee

The Education University of Hong Kong (Hong Kong)
s1131713@s.eduhk.hk, albertlee@eduhk.hk

The interplay between speech melody and song melody in tonal languages, especially that in Chinese songs, has been of interest to scholars because of the close relationship between pitch and lexical meaning. However, previous studies have mainly focused on popular songs in Standard Mandarin and Cantonese, and research on other Chinese dialects and traditional Chinese opera is scarce. In fact, the tonal-melodic mapping prescribed by the compositional principles of traditional opera differs from the approach taken by popular songs, thus tone-melody correspondence in this form of musical genre and the listeners' strategies for extracting the lexical meaning of the lyrics worth to be investigated.

Pitch sequences, in popular songs, would abandon the ratio scale of fundamental frequency (F_0) of the vocal tones in the lyrics and instead use an ordinal scale [1]. For example, a higher pitch in speech can be realized as any higher F_0 in music, but never a lower F_0. However, the compositional principle of Chinese traditional opera, "*yi zi xing qiang* (Tunes Following Lyrics)", does not strictly follow the patterns of ordinal mapping [2]. This principle states that ordinal mapping is only reserved when the tone sequence is rising or falling. When the tone sequence is consistent, the melody direction will manifest as falling instead of remaining plateaued.

The previous study of traditional opera employed the Suzhou Tanci, a traditional Suzhou opera, as a test case [3]. The study showed that Suzhou Tanci was not in high concordance with the compositional principle (the degree was only 49%). This may be related to the tone sandhi in Wu dialects. In addition, the study did not consider the strategies used by listeners to access the meaning of the lyrics. Therefore, our study has two goals: the first is to add to the body of studies on tone-melody mappings in traditional opera, and find a suitable sample of greater adherence to the composition principle for the perception experiment; the second is to explore the relationship between tone-melody interaction and listeners' perception.

We first calculated the frequency of tone-melody mapping in 30 pieces of Sichuan opera and Beijing opera respectively. Then, a perception experiment was conducted to test whether this principle would affect listeners' accurate recognition of lyrics.

Table 1 shows the tone-melody correspondence data of 30 Sichuan operas chosen. The numbers in parentheses are the percentage of occurrences relative to the total number of cases included in the matrix. The frequency of compliance with composition principle occurred 76.90% of the time (the sum of the grey cells in **Table 1**) in these 30 Sichuan operas.

		Musical sequence		
		Up	Down	Same
Tone sequence	Up	491 (32.78%)	34 (2.27%)	36 (2.40%)
	Down	41 (2.74%)	470 (31.37%)	32 (2.14%)
	Same	60 (4.00%)	191 (12.75%)	143 (9.55%)

Table 1. Number of tone-melody relation in Sichuan opera

This frequency reached only 46.17% in Beijing Opera (**Table 2**), indicating lower frequency than that of Sichuan opera, same as the Suzhou Tanci (49%).

		Musical sequence		
		Up	Down	Same
Tone sequence	Up	125 (18.80%)	87 (13.08%)	40 (6.01%)
	Down	89 (13.38%)	96 (14.44%)	48 (7.22%)
	Same	51 (7.67%)	86 (12.93%)	43 (6.47%)

Table 2. Number of tone-melody relation in Beijing opera

This may be due to the limited diachronic variation in Sichuan Mandarin tones [4]. Furthermore, to avoid sacrificing musicality, it could be desirable to properly abandon strict adherence to the tonal-melodic correspondence [1]. To this end, Sichuan opera was chosen as a test case for the perception experiment.

In the perception experiment, 214 participants were invited. The participants should understand Sichuan Mandarin and possess normal hearing. This experiment followed the design of Wong and Diehl's perception experiment [1], with some modifications. Instead of using a carrier phrase that can carry all potential words, we directly took musical segments from a Sichuan opera excerpt sung by a professional Sichuan opera singer. We measured listeners' perception accuracy as a function of composition principles compliance, which was adjusted by manipulating the melody trend. As the frequency of tone-melody mapping of single sentence in Sichuan opera was concentrated above 50% (M = 0.783, SD = 0.185), the ordinal mapping ratings of chosen stimuli were 100%, 83.3%, and 66.7%. We chose 3 sentences from each scale to compose the 9 original stimuli. These 9 original samples were then resynthesized by the PSOLA method in Praat [5], to obtain the musical variant sample at the other 2 scales. This controlled the variables and ensured a relatively high level of naturalness of sound. A total of 27 clips were obtained, and participants were invited to listen to 9 music clips (random 1 variant stimuli of a sentence * 9 different sentences) and transcribe the lyrics.

A chi-square test of independence found a significant association between the correct recognition of the lyrics and ordinal mapping rate of the stimuli (100% vs. non-100%), χ^2 (1, N = 1926) = 6.350, p = .012. This implied that listeners of Sichuan opera use the principle of traditional opera composition to attain the lexical meaning of the lyrics.

In conclusion, we found that Sichuan opera is composed with a high degree of adherence to the principle of "Tunes Following Lyrics", a partially conforming ordinal mapping. This could be rare in a tonal language with four tones, as it offers limited options for pitch variation [6]. Correspondingly, we found that when listening to Sichuan opera, listeners applied this compositional principle in identifying the lyrics. Although, this principle does not strictly adhere to ordinal mapping, this descending processing is similar to the downdrift contour observed in ordinary conversational speech [7]. Thus, the result of our perceptual experiment may also reflect listeners' experience of ordinary speech patterns.

References

[1] Wong, P. C., & Diehl, R. L. (2002). How can the lyrics of a song in a tone language be understood?. *Psychology of Music*, *30*(2), 202–209. https://doi.org/10.1177/0305735602302006

[2] Miao, T., Ji, L., & Guo, N. (1985). *Zhongguo yinyue cidian* [Chinese music dictionary]. People's Music Publishing House.

[3] Yang, C. (2019). Tunes Following Lyrics in the Singing of Suzhou Tanci Schools. *Art of Music (Journal of the Shanghai Conservatory of Music),* 2019(02). https://doi.org/10.19359/j.cn31-1004/j.2019.02.012

[4] Endo, M., & Ishizaki, H. (2015). *Xiandai hanyu de lishi yanjiu* [Historical Study of Modern Chinese]. Zhejiang University press.

[5] Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication*, *16*(2), 175–205. https://doi.org/10.1016/0167-6393(94)00054-e

[6] Chao, R. (1956). Tone, intonations, singsong, chanting, recitative, tonal composition and atonal composition in Chinese. In M. Halle, H. G. Lunt, H. McLean, & C. H. Van Schooneveld (Eds.), *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday, 11th October 1956*. essay, The Hague, the Netherlands: Monton & Co.

[7] Wong, P. C. (1999). The effect of downdrift in the production and perception of Cantonese level tone. In J. J. Ohala, Y. Hasagawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 2395–2398). ICPhS99, USA.

English listeners' perceptual adaptation to unfamiliar lexical suprasegmental contrast

Hyoju Kim & Jieun Lee

University of Kansas (USA)
kimhj@ku.edu, jieunlee@ku.edu

Introduction. The current study investigates listeners' perceptual adaptation and flexibility to unfamiliar lexical suprasegmental contrast. Listeners are remarkably flexible and rapidly make modifications to accommodate unfamiliar speech patterns in speech perception [1,2,3]. Most prior studies have focused on segmental contrasts (e.g., English stop voicing contrast), whereas listeners' flexibility in the perception of suprasegmental contrasts has been understudied. This study aims to extend the scope of listeners' perceptual flexibility to the suprasegmental contrast by testing native English listeners' processing of unfamiliar lexical stress contrast. More specifically, we encouraged English listeners to adapt to the unfamiliar contrast by increasing their reliance on the secondary dimension (i.e., pitch) and examined individual variability in adaptation patterns (for cue weightings of English stress contrast, see [4,5]). The research questions of this study are as follows: (i) are previously observed listeners' flexible adaptation to unfamiliar segmental contrasts extended to suprasegmental contrasts? (ii) are individual differences in speech adaptation associated with their use of the secondary cue? (iii) are individual differences in speech adaptation and categorical gradience associated with individual listeners' domain-general cognitive abilities (e.g., inhibitory control)?

Methods. Twenty-eight native English listeners completed a Visual Analog Scaling (VAS) task, followed by a language background questionnaire, an adaptation task, a Stroop task (inhibitory control measurement), and a cue-weighting speech perception task. The lexical item used in the VAS, the adaptation, and the cue-weighting tasks was a stress minimal pair in English, *DEsert* vs. *deSSERT* [5]. The vowel quality and pitch of the recorded item were manipulated into seven equidistance steps (step 1 being *DEsert*), respectively. The duration and intensity of the contrast were neutralized across syllables. The stimuli for the Baseline and Exposure of the adaptation task were a subset of the VAS stimuli (Baseline: 14 stimuli with 7 repetitions; Exposure: 12 stimuli with 18 repetitions) (Fig. 1). The Test stimuli (a red square and a blue triangle in Fig. 1) were the most ambiguous step in vowel quality but with canonical steps in pitch.

Results. The results of the VAS task (Fig. 2, left panel) showed extensive individual variabilities, with some being more categorical listeners and others being more gradient listeners. The results of the cue-weighting task (Fig. 2, middle panel) replicated those of earlier studies [4,5] in that English listeners use vowel quality as a primary cue ($\beta = -0.72$, $z = -16.76$, $p < .001$) and pitch as a secondary cue ($\beta = -0.1$, $z = -2.4$, $p < .05$) for perceiving lexical stress contrasts. Notably, for the results of the adaptation task (Fig 2, right panel), the mixed-effects logistic regression model found a marginal interaction of Block (Exposure vs. Baseline 1) × Pitch ($\beta = 0.54$, $z = 1.95$, $p = .05$), indicating that listeners are more likely to use pitch dimension to process lexical stress contrast in Exposure block more than they did in Baseline 1. Additionally, listeners with higher inhibitory control (i.e., lower Stroop interference) made more categorical responses to the VAS task (Fig. 3, left panel). There was no remarkable correlation between inhibitory control and listeners' response in the adaptation task (Fig. 3, right panel).

Discussion. The current results demonstrate that previously observed listeners' flexibility to unfamiliar speech patterns extends to the lexical suprasegmental contrasts, suggesting that the speech perception system adjusts to the acoustic consequences of changes in the relative informativeness of acoustic dimensions. Although the results of this study are in line with the previous findings, the degree of listeners' flexibility was not large enough as compared to the other studies that examined segmental contrast [1,2,3], presumably due to the nature of the acoustic cues involved in the lexical stress contrasts. Our results also suggest that listeners with higher cognitive abilities to suppress goal-irrelevant information are more likely to process the lexical stress contrasts in a categorical manner. However, listeners' perceptual adaptability may not necessarily be associated with their inhibitory control.

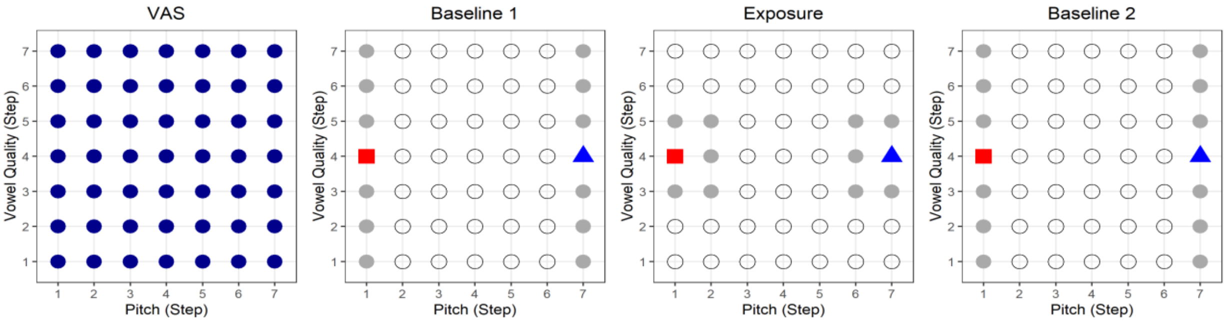


Fig. 1 Illustration of the auditory stimuli of the VAS and the adaptation task

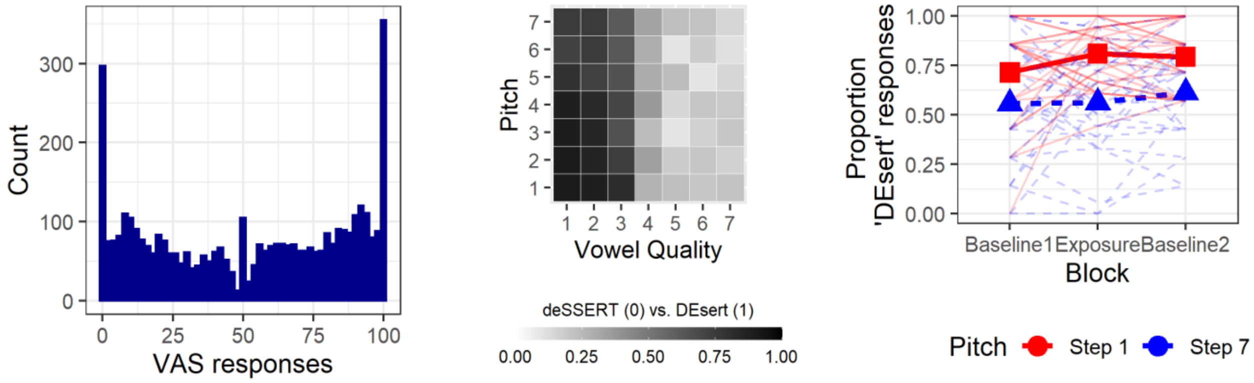


Fig. 2 Participants' responses to the VAS task (left panel), the cue-weighting task (middle panel), and the adaptation task (right panel)

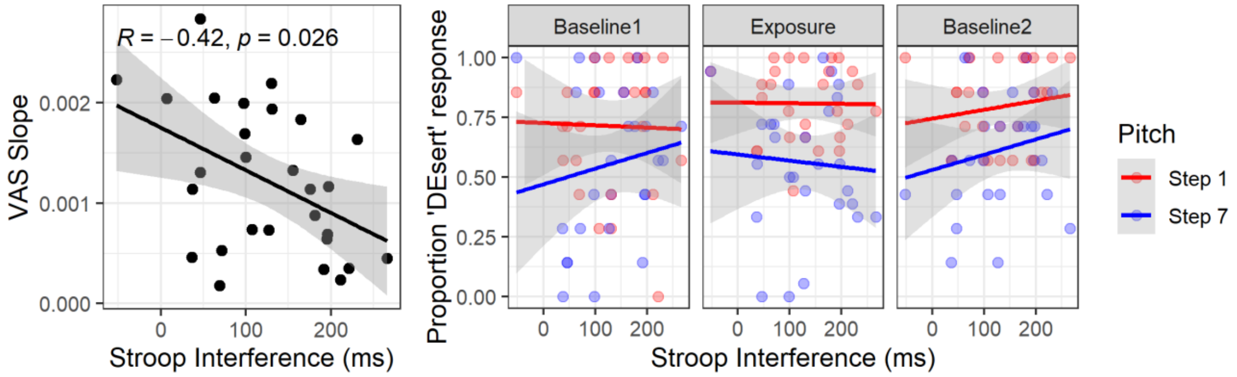


Fig. 3 Relationship between inhibitory control and categorical gradience (left panel) and categorization responses across blocks of the adaptation task (right panel)

References

[1] Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1939-1956.

[2] Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics, 81*, 100984.

[3] Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics, 78*, 355-367.

[4] 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.

[5] Tremblay, A., Broersma, M., Zeng, Y., Kim, H., Lee, J., & Shin, S. (2021). Dutch Listeners' Perception of English Lexical Stress: A Cue-Weighting Approach. *The Journal of Acoustical Society of America, 149*, 3703-3714.

Acoustic analysis of Glottal Stops in Mundari

Pamir Gogoi, Luke Horo & Gregory D. S. Anderson

Living Tongues Institute for Endangered Languages (USA)
pamir.gogoi11@gmail.com, luke.horo@livingtongues.org, gdsa@livingtongues.org

The objective of this paper is to analyze the phonetic realization of glottal stop in Mundari. Glottal stop is a frequently appearing feature in Mundari, an Austroasiatic language spoken in India. While a glottal segment is historically attested in the Austroasiatic language phylum, synchronic studies reveal that glottal articulation surfaces differently in different languages. For instance, in Sora intervocalic glottal stops have three different phonetic realizations- including, a complete vocal fold closure, a complete closure accompanied by creaky phonation and a voiced glottal stop [1]. Likewise, Gorum is also known to have three distinct glottal articulations, including, a complete glottal stop, pre-glottalized obstruents and creaky voice [2]. In the case of Mundari, glottal constriction is known either as an allophonic variant of a word final velar voiced obstruent [3,4] or as a vocalic feature separating identical vowel sequences [5]. However, apart from Sora, there has not been any instrumental study of glottal stops in the Austroasiatic languages spoken in India. The present study analyzes the glottal stop in Mundari and its possible articulatory variations using spectrographic evidence.

Crosslinguistic evidence suggests that glottal stops are often realized partially by exhibiting laryngealization instead of a complete stop and these characteristics may vary based on the context [6]. Also, changes in *f*0, amplitude and spectral measures of source features are some of the widely observed correlates of glottal stops [7]. Moreover, it has been observed that in naturally spoken continuous speech, these features do not strongly correlate to the realization of glottal stops [9]. Therefore, in this study we measure changes in *f*0, amplitude and spectral features both in continuous speech and isolated segments in Mundari. Preliminary observation of Mundari speech data suggests that intervocalic glottal stops in isolated words are primarily produced with either a complete closure of the vocal fold (see Fig.1) or with a creaky phonation (see Fig.2). In some instances, a dip in the *f*0 is observed in the region of the glottal stop. In the word final position, some creakiness is observed in the offset of the preceding vowel.

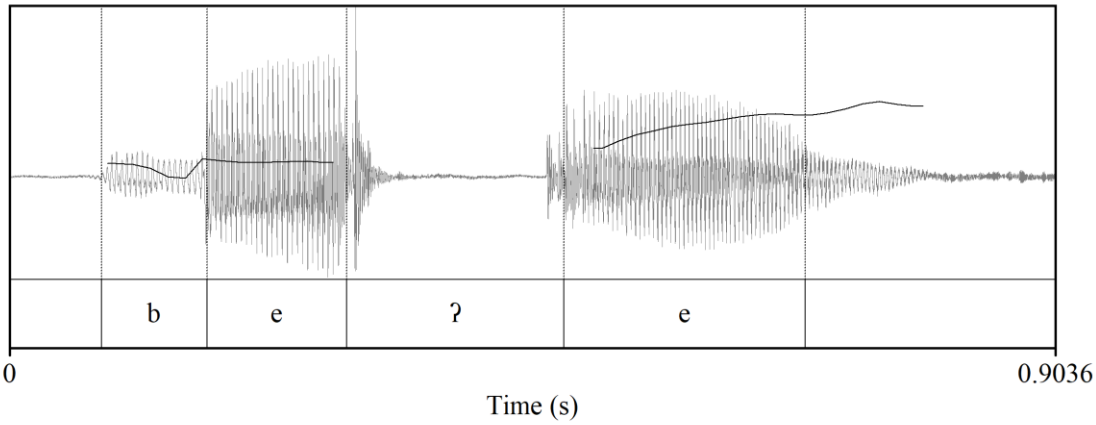


Fig. 1 Intervocalic glottal stop [ʔ] in Mundari realized as a complete stop in the word *beʔe* ‘to spit’.

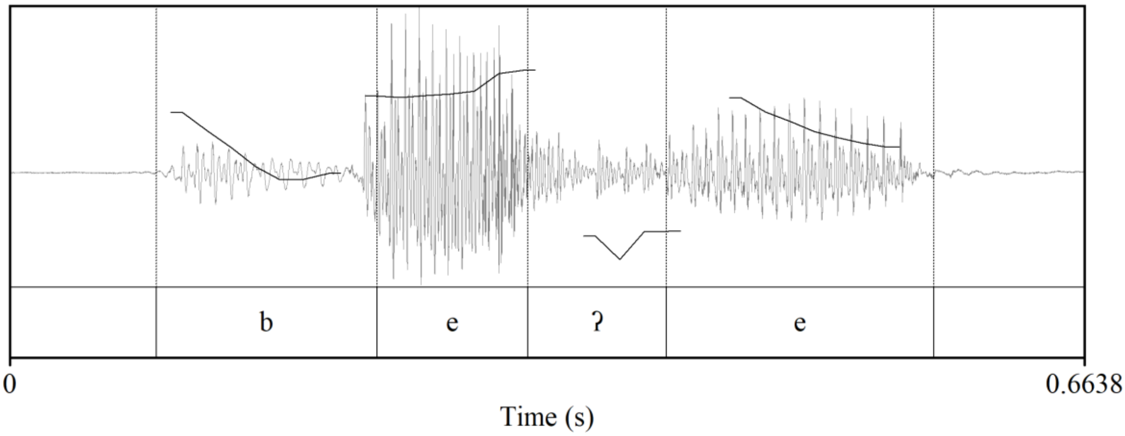


Fig. 2 Intervocalic glottal stop [ʔ] in Mundari realized with creaky phonation in the word *beʔe* ‘to spit’.

References

[1] Kalita, S., Horo, L.,Sarmah, P., Prasanna, S. M., & Dandapat, S. (2016). Analysis of Glottal Stop in Assam Sora Language. In *INTERSPEECH* (pp. 1049-1053).

[2] Rau, Felix. (2011). “Notes on Glottal Constriction in Gorum.” In Sophana Srichampa and Paul Sidwell (eds.) Austroasiatic Studies: papers from ICAAL4. *Mon-Khmer Studies Journal Special Issue No. 2.* Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics. Pp.174-183.

[3] Osada, T. (1992). A Reference Grammar of Mundari. *Institute for the Study of Languages and Cultures of Asia and Africa*, Tokyo University of Foreign Studies.

[4] Osada, T. (2008). Mundari. In Gregory D.S. Anderson (ed.)*The Munda languages*, 99-164

[5] Hoffmann, Johan. (1950) Encyclopedia Mundarica, *Patna: Government Superintendent Printing.* 15 volumes.

[6] Garellek, M. (2013). Production and perception of glottal stops (Doctoral dissertation, UCLA)

[7] Hillenbrand, J. M., & Houde, R. A. (1996). Role of F0 and amplitude in the perception of intervocalic glottal stops. *Journal of Speech, Language, and Hearing Research, 39*(6), 1182-1190.

[8] Ashby, M., & Przedlacka, J. (2014). Measuring incompleteness: Acoustic correlates of glottal articulations. *Journal of the International Phonetic Association*, *44*(3), 283-296.

Pitch accent alignment in Persian

Vahid Sadeghi
Imam Khomeini International University (*Iran*)
vsadeghi@hum.ikiu.ac.ir

The study was intended to examine the effects of various prosodic factors such as syllable structure and segmental composition as well as proximity of the following word boundary and accent on scaling and alignment pattern of prenuclear rising accents in Persian, in order to shed some light on our understanding of the production of the accent gestures and their coordination with the segmental material. Two production experiments were conducted.

The first experiment explored the variability in the timing of the tonal targets as a function of syllable structure and vowel type of the target accented syllable. Results revealed that in proparoxytones, the alignment of the H tones was affected by variations in the duration of the accented syllable in consistent ways: H pre-nuclear peaks were earlier in open and short syllables than closed and long syllables (see Fig. 1 (left panel) for a comparison between open and long syllables). However, when the alignment of the H was measured with reference to the onset of the first post-accentual vowel, syllable structure or vowel type failed to produce significant effects: The H target was aligned on average within 10 ms after the onset of the first post-accentual vowel syllables (see Fig. 1 (right panel) for a comparison between open and long syllables). The results support the Segmental Anchoring Hypothesis, whereby the duration of pitch movements in speech is finely adjusted to the duration of the accompanying segmental material.

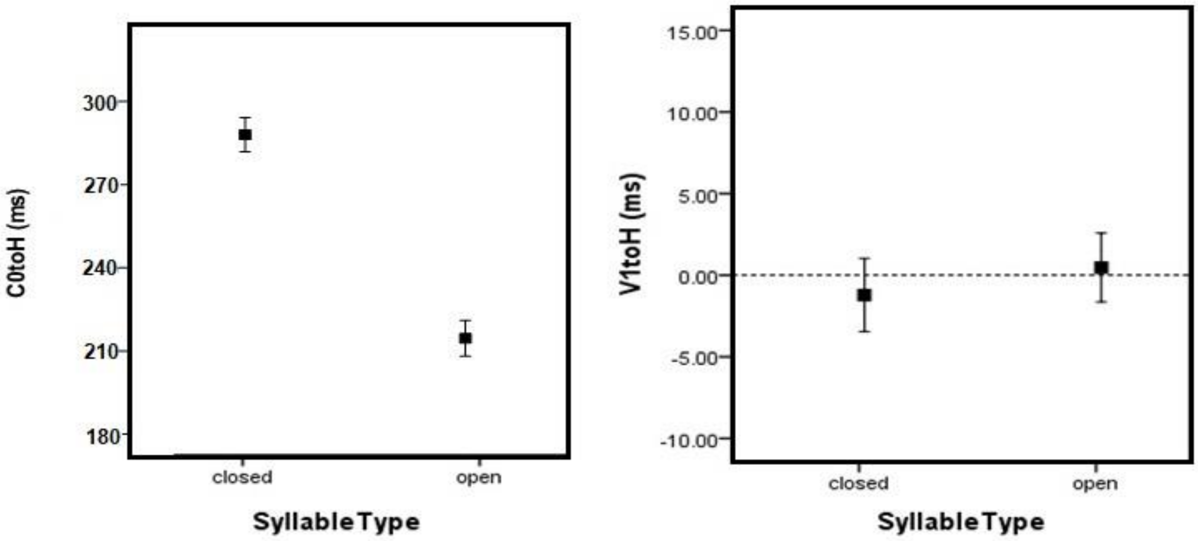


Fig. 1 Left panel: Mean distance from the beginning of the consonant of the accented syllable to H (C0toH); Right panel: Mean distance from the onset of the accented vowel to H (V1toH) (right) as a function of syllable structure.

The second experiment examined the variability in the scaling and timing of pitch accents as a function of the proximity of the word boundary and of the following accent. The alignment data revealed clear effects of stress conditions on H location. Peaks were located earlier as the distance of the target accented syllable and the word boundary decreased. The H target was realized early in oxytones, and progressively later in paroxytones and proparoxytones. L targets, on the other hand, were highly stable across different stress conditions.

In general, our results replicate and extend earlier findings of Arvaniti et al [2] for Greek, Ladd et al., [4] for English and Atterer and Ladd [3] for German showing that in the absence of prosodic pressure from the upcoming material, i.e., when the accented syllable is not in the vicinity of the word boundary or the next accent, the two tones of pre-nuclear pitch accents in Persian are consistently aligned with respect to the segmental material, and the stability effects are pervasive under changes of segmental or syllable structure composition.

A cross-linguistic comparison between the Persian data and data from other languages reveals subtle differences in alignment (as shown in Fig. 13) that cannot be accounted for in terms of phonological specifications of tonal association; rather they might best be interpreted in terms of continuous phonetic alignment rules as proposed by Atterer and Ladd [3], and also advocated by many others, including Arvaniti [1]. Phonetically the pattern of L alignment we find for Persian is slightly later than those of English [4] and Greek [2] and rather earlier than Southern German [3]. For H, the Persian pattern of alignment is quite comparable to what Arvaniti et al. [2] found for Greek and what Atterer and Ladd [3] found for Northern German, and rather later than the findings of Ladd et al. for English and Dutch. One interpretation of this comparison would be that language-specific differences in the alignment of pitch movements may be a matter of what Ladd [5] calls phasing: the same F0 change can be aligned earlier or later. This in turn suggests that the two targets of a bitonal pitch movement are not independently aligned at specific places in structure; rather the whole movements are aligned relative to whole syllables. Thus, Southern German aligns both L and H later than Northern German, Greek and Persian, which in turn align both L and H later than later than Dutch and English. This, in general, may provide some evidence for Xu's idea [7] that the rise is, at some level of analysis, a unitary phonological event, the alignment of which is specified as a whole.

In general, our findings of segmental anchoring provide little evidence for the starred tone interpretation assumed in early auto-segmental theory, according to which one of the two tones of a bitonal accent must be aligned with the accented syllable, while the other tone merely leads (or trails) the starred tone by a fixed temporal interval. In addition, starring one of the two tones in the Persian LH pitch accents would pose problems for Persian intonational system in which there is no contrast of alignment between L+H* and L*+H [6]. Rather, based on both language-specific and cross-linguistic evidence presented in this study, we may suggest that the most appropriate representation of pre-nuclear rising accent in Persian is the starless sequence of a low (L) and a high (H) tone, namely LH, where the L is realized at the beginning of the stressed syllable and the H early in the vowel of the post-tonic syllable according to phonetic implementation rules of alignment for Persian.

References:

[1] Arvaniti, A. (2016). Analytical Decisions in Intonation Research and the Role of Representations: Lessons from Romani. *Journal of the Association for Laboratory Phonology* 7(1), 6.

[2] Arvaniti, A, Ladd, D. R & Ineke M. (1998). Stability of tonal alignment: the case of Greek pre-nuclear accents. *Journal of Phonetics* 26, 3–25.

[3] Atterer, M. & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics* 32, 177-197.

[4] Ladd, D. R., Faulkner, D., Faulkner, H. & Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106, 1543-1554.

[5] Ladd, D. R. (2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Rivista di Linguistica* 18(1), 19-38.

[6] Sadat-Tehrani, N. (2007). *The Intonational grammar of Persian*. PhD thesis, University of Manitoba.

[7] Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179–203.

Language dominance influences L1 attrition and L2 acquisition of lexical tones: Data from Mandarin-speaking immigrants in Hong Kong

Yike Yang, Dong Han

Hong Kong Shue Yan University (Hong Kong)
yyang@hksyu.edu, dhan@hksyu.edu

Language dominance is a broad concept that covers various domains of bilingual speakers' languages [1]. Previous investigations of language dominance have mainly focused on the sociolinguistic or individual perspectives, while the role of language dominance in first language (L1) attrition and second language (L2) acquisition of lexical tones remains underexplored. This study aims to address this issue by exploring the production of L1 and L2 tones by Mandarin-speaking immigrants in Hong Kong, who were late Mandarin-Cantonese speakers with varying degrees of language dominance. There are four lexical tones in Mandarin: Tone 1 (T1) is a high level tone, Tone 2 (T2) is a rising tone, Tone 3 (T3) is a low dipping tone and Tone 4 (T4) is a falling tone [2]. In Cantonese, there are six lexical tones: Tone 1 (T1), Tone 3 (T3) and Tone 6 (T6) are level tones, Tone 2 (T2) and Tone 5 (T5) are rising tones, and Tone 4 (T4) is a low falling tone [3]. Testing the effects of language dominance on Mandarin tone attrition and Cantonese tone acquisition is of theoretical significance due to the obvious differences in the two tonal systems.

A tone production experiment was conducted with 32 Mandarin-speaking immigrants who had spoken Mandarin as their only Chinese dialect prior to their arrival in Hong Kong. According to a language background questionnaire [4], the immigrants were fluent speakers of Mandarin and Cantonese. Two syllables that contained all the possible tones were selected as the target syllables for the stimuli in each language. There were eight monosyllabic target words for Mandarin (2 target syllables * 4 tones) and 12 monosyllabic target words for Cantonese (2 target syllables * 6 tones). The syllables were presented in two contexts: in isolation and embedded in a carrier sentence. All the stimuli were presented twice to each speaker. For each syllable, the vowel portion was segmented, and 20 time-normalised F0 values were extracted using Praat [5]. To eliminate individual differences in the F0 range, the F0 values were converted to a five-point scale ranging from 1 to 5, with 1 representing the lowest F0 value and 5 indicating the highest. In the analysis, generalised additive mixed models (GAMMs) were adopted for modelling time-dependent datasets in R [6].

The 32 participants were divided into two dominance groups according to their scores on the language background questionnaire: a balanced group, who were relatively balanced in their two languages and whose Mandarin was assumed to show a higher degree of attrition, and an unbalanced group, who were still much more dominant in Mandarin than they were in Cantonese. According to the GAMMs, the unbalanced group could clearly distinguish the four Mandarin tones regardless of the context (*p*s < .001), but the balanced group had merged T2 and T3 when the syllables were pronounced in the carrier sentence (p = .176), suggesting that language dominance had led to the attrition of their L1 Mandarin. With regard to Cantonese, the balanced group had merged T3 and T6, both in the isolation (p = .236) and in the sentence (p = .674) conditions. The unbalanced group had merged T2 and T5 in the isolation condition (p = .219), and T3 and T6 in both the isolation (p = .986) and sentence (p = .324) conditions. Both groups were able to distinguish the other Cantonese tone pairs in other conditions. The data for the Cantonese tones indicated that the more balanced bilinguals had acquired the Cantonese tones more successfully, revealing the role of language dominance in L2 tone acquisition. The tone production was plotted in more detail in Figures 1 and 2, which confirmed the statistical analyses above. Moreover, Figure 1 shows that, compared to the unbalanced speakers, the balanced bilinguals had a smaller tonal space in Mandarin, which should have been influenced by Cantonese, as both groups exhibited a smaller tonal space in Cantonese than they did in Mandarin. Based on the data pertaining to L1 and L2 tone production, this study supports the claim of the Prosody Transfer Model, namely that prosodic features can be transferred between an L1 and an L2, even among late L2 learners [7].

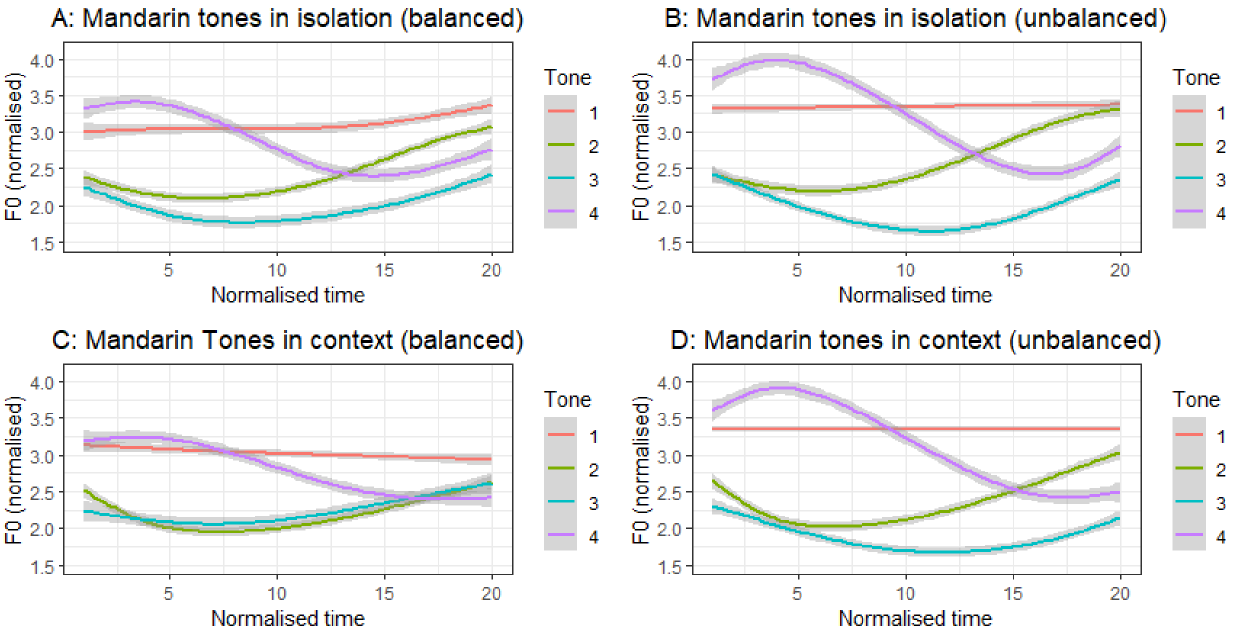


Fig.1 Production of Mandarin tones by the immigrants.

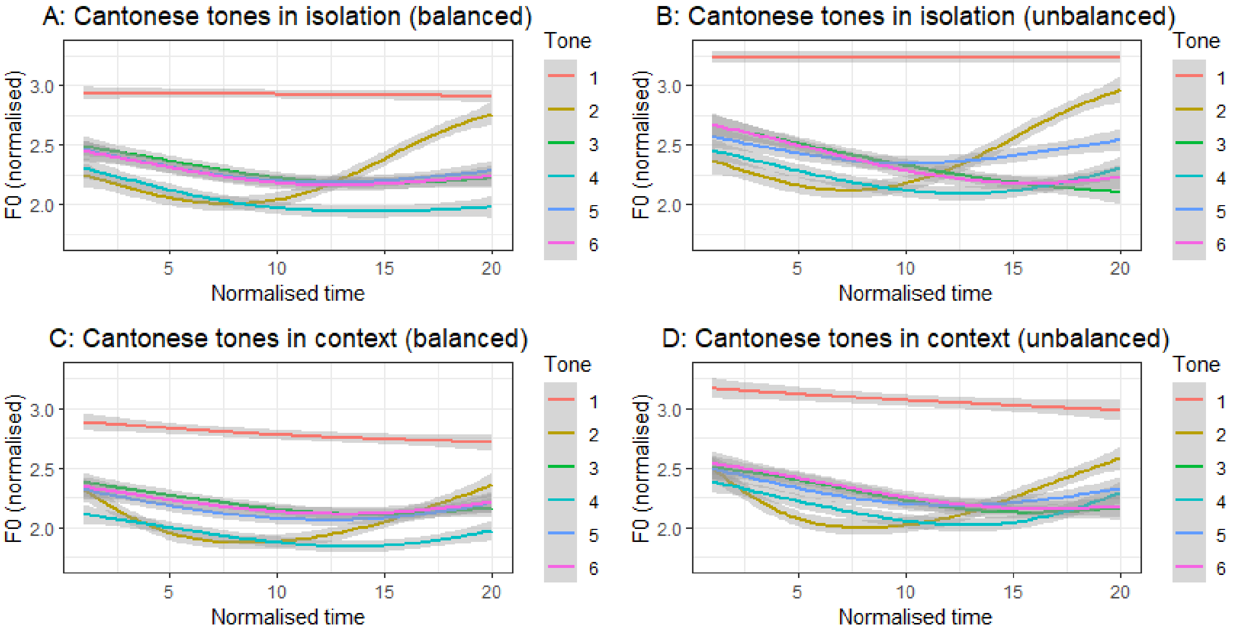


Fig.2 Production of Cantonese tones by the immigrants.

Acknowledgements: This work was supported by an ASA International Student Grant from the Acoustical Society of America and a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No.: UGC/FDS15/H15/22) to the first author.

References

[1] Treffers-Daller, J. (2019). What defines language dominance in bilinguals?. *Annual Review of Linguistics*, *5*, 375-393.

[2] Chao, Y. R. (1948). *Mandarin Primer*. Harvard University Press.

[3] Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Walter de Gruyter.

[4] Birdsong, D., Gertken, L.M., & Amengual, M. (2012). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin. <https://sites.la.utexas.edu/bilingual/>.

[5] Boersma, P. & Weenink, D. (2015). *Praat: doing phonetics by computer*. [Online]. <http://www.praat.org/>.

[6] R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. [Online]. <https://www.r-project.org>.

[7] Yang, Y. (2022). *First Language Attrition and Second Language Attainment of Mandarin-speaking Immigrants in Hong Kong: Evidence from Prosodic Focus*. Doctoral dissertation, The Hong Kong Polytechnic University.

Perceiving speech produced with face masks in competing talker environments

Faith Chiu[1], Laura Bartoševičiūtė[2], Albert Lee[3] and Yujia Yao[2]

[1]University of Glasgow (UK), [2]University of Essex (UK), [3]Education University of Hong Kong (Hong Kong SAR)
faith.chiu@glasgow.ac.uk, laurabartoseviciute@gmail.com, albertlee@eduhk.hk, vera.yujia.yao@gmail.com

Introduction. The COVID-19 pandemic has reshaped speech communication. Face-to-face communication often includes one or both parties sporting a face mask. The listener's comprehension effort now involves adapting to mask-imposed distortions to the acoustic speech signal [1]. Even native speakers [2] struggle with understanding speech produced with a face mask when presented in noise. Everyday speech communication can take place in a noisy background with a competing talker. It is also not uncommon these days to converse in one's second or additional language. Using two experiments, this study aims to understand the difficulty imposed by speech produced with face masks in a multi-talker environment. Target sentences produced with and without a face mask were presented to listeners in the presence of a competing talker. The competing speech either matched or differed in language from target sentences. Participants' linguistic background determined the intelligibility of the competing talker.

Stimuli. The auditory stimuli consisted of target sentences in English and in Lithuanian, and competing speech in English and in Lithuanian. English target sentences were based on the British English version of the International Matrix sentence test [3] using a 50-word base matrix (10 names, 10 verbs, 10 numerals, 10 adjectives, and 10 nouns). Sentences were generated using a random combination of one word of each category in a fixed syntactic structure ('Alan bought two big beds'). Lithuanian target sentences follow the same format and were constructed as original stimuli. Target sentences were recorded by a native female speaker of each language. Individual words were produced, with and without a cotton fabric face mask, then combined acoutically. The competing speech was semantically meaningful sentences in either English or Lithuanian produced by a male speaker without a face mask. The target sentences were presented at a challenging level (−10dB Signal-to-Noise ratio). Male voices were chosen for competing speech and female for target sentences so participants can utilise speaker sex as a segregation cue.

Participants. 24 native Lithuanian listeners (13 female and 11 male, age range: 18–37) took part in *Experiment 1*. Participants for *Experiment 2* were 22 monolingual British English speakers (16 female and 6 male, age range: 18–34) and 22 second language speakers of English with Mandarin Chinese as first language (19 female and 3 male, age range: 20–31).

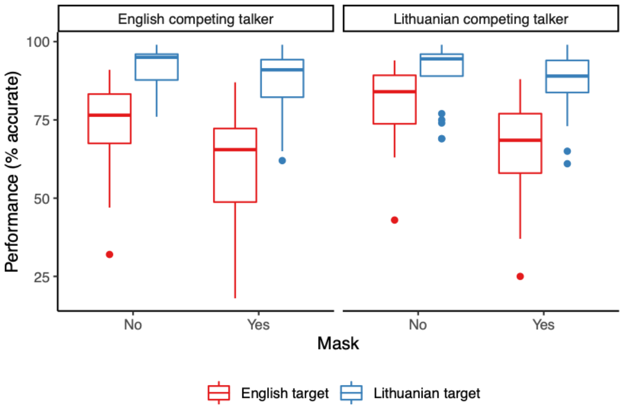
Procedure. *Experiment 1* was conducted online. Participants were instructed that they would hear target sentences by a female talker in the presence of a male competing talker and that they were to listen only to the female voice and ignore the male. They then had to type what they heard after each sentence. Participants heard a total of 160 trials: from 2 TARGET LANGUAGES (English/Lithuanian) × 2 MASK conditions (YES/NO) × 2 COMPETING SPEECH LANGUAGES (English/Lithuanian) with 20 sentences each. *Experiment 2* features an identical procedure to *Experiment 1* except that participants heard only English target sentences produced with and without mask, in the presence of either competing English or Lithuanian speech. Responses were scored based on the number of words accurately reported in each sentence.

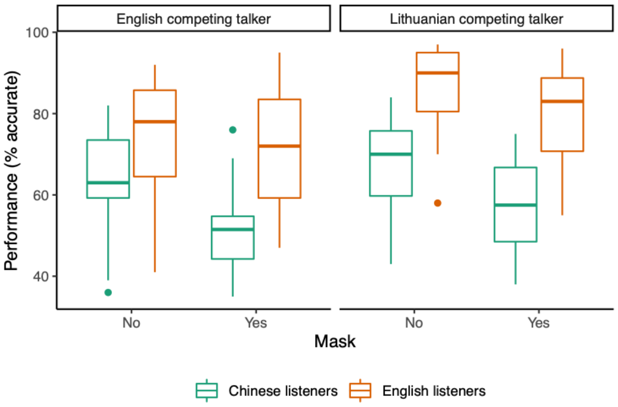
Results. The left figure reveals Lithuanian listeners' performance (*Experiment 1*). The right figure shows the performance of English and Chinese listeners (*Experiment 2*).

Experiment 1. A 2 × 2 × 2 analysis of variance (ANOVA) was conducted on the percentage of accurately reported keywords as a function of language of TARGET (English versus Lithuanian), MASK (with or without a face mask), and language of COMPETING SPEECH (English vs. Lithuanian). The results revealed two significant two way-interactions: TARGET × MASK ($F(1, 23) = 26.001$, $p < .001$, $\eta p^2 = .531$) and TARGET × COMPETING SPEECH ($F(1, 23) = 25.123$, $p < .001$, $\eta p^2 = .522$). Individual 2 × 2 ANOVAs were performed for each target language as follow-up, comparing the %accuracy as a function of MASK and language of COMPETING SPEECH. There was a significant main effect of MASK ($F(1, 23) = 54.448$, $p < .001$, $\eta p^2 = .703$) for English target sentences. More keywords were accurately reported on sentences produced without a face mask, in both English and Lithuanian competing speech. There was also a main effect of COMPETING SPEECH ($F(1, 23) = 31.304$, $p < .001$, $\eta p^2 = .576$). Lithuanians listeners were less accurate when the competing speech was in a language which matches the English target sentences; this was true both when the targets were produced with and without a mask. However, when listening to Lithuanian target sentences, there was only a main effect of MASK (F(1, 23) = 15.544, p < .001, ηp^2 = .403). Lithuanian target sentences produced with a face mask were more poorly perceived, and this was true regardless of whether it was presented in both English and Lithuanian competing speech. Unlike in English target sentences, there was no effect of COMPETING SPEECH. Planned comparisons showed that Lithuanian listeners reported more accurate keywords when listening to Lithuanian than English targets.

Experiment 2. A mixed 2 × 2 × 2 ANOVA was conducted on %accuracy as a function of MASK (with or without a face mask), language of COMPETING SPEECH (English vs. Lithuanian), and GROUP (English vs. Chinese listeners). The results indicated a significant three-way interaction of MASK × COMPETING SPEECH × GROUP (F(1, 42) = 6.497, p = .015, ηp^2 = .134). Planned comparisons showed English listeners outperforming Chinese listeners in all conditions. Individual 2 × 2 ANOVAs performed for each listener group revealed that for English listeners, there was a main effect of MASK (F(1, 21) = 5.439, p = .030, ηp^2 = .206) as well as a main effect of COMPETING SPEECH (F(1, 21) = 78.729, p. < .002, ηp^2 = .789). Chinese listeners were similar with both a main effect of MASK (F(1, 21) = 21.960, p < .001, ηp^2 = .511) and COMPETING SPEECH (F(1, 21) = 19.869, p < .001, ηp^2 = .486).

Discussion. In sum, masked speech is always more poorly perceived across all listener groups in all conditions. This finding echoes existing reports of decreased perception performance when listening to speech produced with a face mask and presented in noise [2]. This across-the-board effect could be due to attenuation of the acoustic signal from mask-wearing in the form of dampening. In particular, high frequency information is lost [1]. Additionally, perception accuracy was higher when listening in one's first language, echoing previous work showing that speech perception with a competing talker is more difficult in one's non-native language [4]. Finally, a competing talker in a language which matches the target sentences had more of a detrimental effect on perception accuracy compared to a mismatched one. This replicates findings of a benefit of linguistic mismatch between target and competing speech for non-native speakers [5]. Exceptionally in our study, when Lithuanian participants (with both English and Lithuanian knowledge) listened for Lithuanian targets there was there no added challenge from matching language of target and competing speech. We conclude that acoustic distortions from face masks present an across-the-board difficulty while linguistic knowledge can reduce distraction from competing speech.





References

[1] Corey, M. R., Jones, U., Singer, C. A. 2020. Acoustic effects of medical, cloth, and transparent face masks on speech signals. *J. Acoust. Soc. Am.* 148, 2371–2375.

[2] Yi, H., Pingsterhaus, A., Song, W. 2021. Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic. *Front. Psychol.* 12, 682677.

[3] Hall, S. J. 2006. *The development of a new English Sentence in Noise Test and an English Number Recognition Test.* MSc thesis. University of Southampton.

[4] Bradlow, A. R., Alexander, J. A. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust. Soc. Am.* 121, 2239–2249.

[5] Brouwer, S., Van Engen, K.J., Calandruccio, L., Bradlow, A.R. 2012. Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. *J. Acoust. Soc. Am.* 131, 1449–64.

Dependent Pitch Cues in Tone Perception: Evidence from Mandarin Chinese

Ok Joo Lee[1] & Kyungmin Lee[2]

[1]Seoul National University (Korea), [2]Seoul National University (Korea)
ojlee@snu.ac.kr, leegmf4@snu.ac.kr

Pitch is known to be a primary phonetic and acoustic property of tone, while other articulatory features such as phonation, duration, and amplitude may also play an important role. Tones, therefore, are often defined by contrastive pitch height or movement such as ‘high’, ‘low’, ‘rising’, and ‘falling’ [1, 2, 3]. However, it has been witnessed that the production of tonal pitch is heavily influenced by neighboring tones, prosodic conditions, segmental properties as well as a speaker’s pitch range and the tonal inventory of a native language [4, 5, 6]. Despite an increasing body of recent perception studies on tones [7, 8, 9], the questions of how listeners utilize seemingly unstable pitch cues and how different pitch cues interact with each other in perception are not fully understood.

The present study investigates two types of pitch cues in tone perception in Mandarin Chinese, namely, pitch height cues and pitch movement cues. As well known, Mandarin has four lexical tones, which are conventionally described as ‘55’, ‘35’, ‘214’, and ‘51’ in Chao’s five-number scale or named Tone 1 through Tone 4 for ease of reference [10, 11, 12]. Among the four tones, Tone 3 surfaces as low tone when followed by Tone 1, Tone 2, or Tone 4. Therefore, pitch height cues can be crucial in identifying the high and low level tones (i.e., Tone 1 and Tone 3) in the non-final position, while pitch movement cues distinguish the rising and falling tones (i.e., Tone 2 and Tone 4). Of particular interest to this study is to explore how level and rising tones are identified when the pitch height and rising slope cues vary and how the pitch cues interact with each other in perception. Perceptual differences between native listeners (NM hereafter) and non-native listeners (i.e., native Korean listeners of advanced-level L2 Mandarin and those of beginning-level L2 Mandarin, NKA and NKB hereafter) are also examined.

A perception experiment that instructed a total of 91 participants to identify the tones in real disyllabic words of four tonal combinations (i.e., T1+T1, T1+T2, T3+T1, and T3+T2) was conducted for the study. The stimuli were created by synthesizing the F0 height on the level tone syllable and the F0 slope on the rising tone syllable, both on a 7-step continuum. The initial F0 height of the rising tone was also manipulated on a 7-step continuum. In total, 259 stimuli and 74 fillers were used in the experiment. Results of the study show: (1) the pitch height and slope cues were crucial in identifying level tones and rising tones, respectively, in NM and NKA. Neither cues were fully utilized by NKB who had a tendency to yield more Tone 1 judgements. (2) Nonetheless, the pitch cues on the adjacent tone had a significant impact on the tone identification, in that (a) the Tone 1 and Tone 3 judgements were influenced by the pitch slope of the following tone, and (b) the level tone and rising tone distinction patterns were strongly influenced by the pitch height of the preceding level tone. The pitch slope cues appeared to bear less importance in non-native perception. (3) Importantly, as shown in Figure 1, the F0 differences between the neighboring tones were the most salient cues by which both NM and NKA categorically distinguished two level tone types as well as level tones from rising tones. The findings suggest that although the pitch height cues and pitch slope cues may be the most crucial cues in identifying level tones and rising tones, respectively, in isolation, the tone perception is guided by the phonetic interactions between different types of pitch cues in connected speech. It is further suggested that the native language and L2 experience of a listener bear a substantial impact on the use of pitch cues in perception.

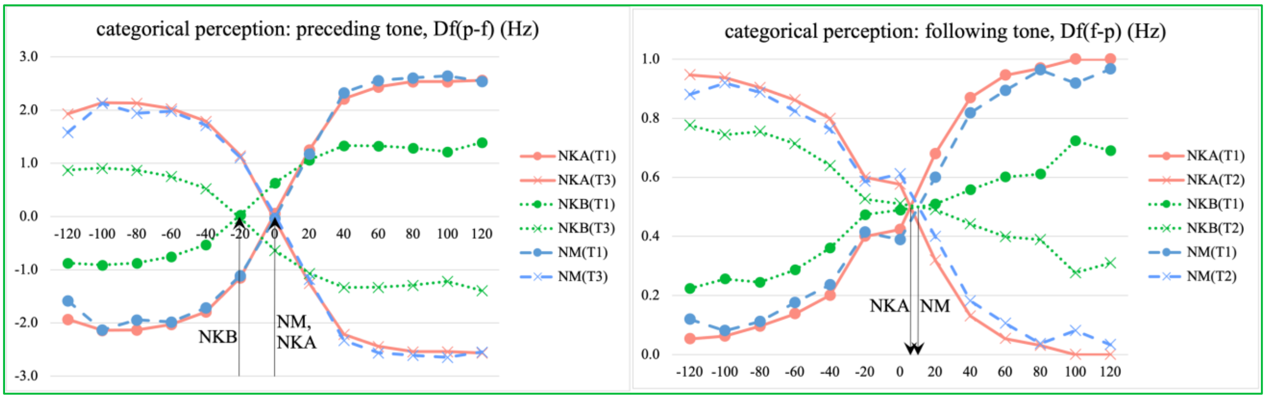


Fig.1. F0 difference cues in tone perception

References

[1] Pike, K. L. (1948). *Tone Languages: A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*. Ann Arbor: University of Michigan Press.

[2] Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

[3] Lee, O. J. (2022). Tones of Asian Languages: A Comparative Overview of Tonology. In C. Shei & S. Li (eds.), *The Routledge Handbook of Asian Linguistics*. (pp.261-278). London: Routledge.

[4] Cao, J., & Maddieson, I. (1992). An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics, 20*(1), 77-92.

[5] Peng, G., Zhang, C., Zheng, H.-Y., Minett, J. W., & Wang, W. S.-Y. (2012). The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *Journal of Speech, Language, and Hearing Research, 55*(2), 579-595.

[6] Gao, J., & Hallé, P. (2017). Phonetic and phonological properties of tones in Shanghai Chinese. *East Asian Languages and Linguistics*, *46*(1), 1-31.

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

[8] Sjerps, M. J., Zhang, C., & Peng, G. (2018). Lexical tone is perceived relative to locally surrounding context, vowel quality to preceding context. *Journal of Experimental Psychology: Human Perception and Performance, 44*(6), 914-924.

[9] Lee, K., & O. J. Lee. (2022). Native and non-native perception of Mandarin level tones. *Linguistic Research*, *39*(3), 567-601.

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

[11] Chao, Y. R. (1956). Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In M. Hallé (ed.), *For Roman Jakobson: Essays on the occasion of his sixtieth birthday*. (pp.52-59). The Hague: Mouton.

[12] Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Alignment of Prosodic and Syntactic Junctures and Vowel-initial Glottalization in Syntactic Disambiguation: A Preliminary Report

Jae-Eun Jennifer Shin[1], Sahyang Kim[2] & Taehong Cho[1]

[1]*Hanyang Institute for Phonetics & Cognitive Sciences of Language (HIPCS), Hanyang University (Korea),*
[2]*Hongik University (Korea)*
jaeeunjshin@gmail.com, sahyang@gmail.com, tcho@hanyang.ac.kr

Significant evidence supporting the role of syntax-prosody boundary mapping in syntactic disambiguation has been accumulated in the existing literature [1, 2, 3]. While prosodic boundaries are commonly treated as phonologically defined and categorical elements involved in syntax-prosody mapping, our proposed approach [4] takes a more nuanced perspective that considers fine phonetic details across suprasegmental and segmental dimensions. In this study, we aim to investigate the phonetic granularity of syntax-prosody mapping, specifically focusing on the voice quality (degree of glottalization) of word-initial vowels aligned with prosodic and syntactic junctures, as well as durational measures encompassing both preboundary and postboundary lengthening. To explore the interaction between syntax, prosody, and focus, we examine syntax-prosody mapping in three focus contexts: Broad, Narrow, and Contrastive. By doing so, we seek to enhance our understanding of how these intricate phonetic aspects contribute to resolving syntactically ambiguous coordinate structures in American English.

An experiment involving acoustic recordings was conducted with a group of fourteen native speakers of American English (7 male, 7 female) aged between 19 and 35. The participants were presented with speech materials (refer to Table 1) in three different focus contexts. They read 'answer' sentences containing ambiguous coordinate structures, such as 'Anna and Annie or Angie,' which can be interpreted in two possible syntactic structures: [N1] and [N2 or N3] (Early Closure) or [N1 and N2] or [N3] (Late Closure). The analysis of prosodic boundaries involved the utilization of the ToBI system to code Intonational Phrase (IP) and Word (Wd) boundaries, as well as the presence of pitch accent. Glottalization was examined using established parameters, including H1*-H2*, CPP, and HNR [5, 6], extracted from three equi-interval segments of vowels (Time points 1~3) using VoiceSauce [7]. Additionally, the duration of the entire syllable preceding and following the prosodic juncture was measured.

Table 1. Speech Materials according to Focus Type. Only the Answer category has been recorded. Early Closure has a syntactic juncture before 'and' whereas Late Closure (LC) has one before 'or'. Note narrow and contrastive focus put emphasis respectively on the whole utterance or syntactic structure than a single lexical item

	Question	Answer	
Broad Focus	What is going on?	Well, (Anna) and (Annie or Angie) are coming.	**EC**
		Well, (Anna and Annie) or (Angie) are coming.	**LC**
Narrow Focus	WHO will come to the party?	Well, (Anna) and (Annie or Angie) will.	**EC**
		Well, (Anna and Annie) or (Angie) will.	**LC**
Contrastive Focus	Did they say (Anna and Annie) or (Angie) will come?	No. They said, (Anna) and (Annie or Angie) will.	**EC**
	Did they say (Anna) and (Annie or Angie) will come?	No. They said, (Anna and Annie) or (Angie) will.	**LC**

The results of the syntax-prosody boundary mappings generally align with the assumptions made in the existing literature [1, 2, 3]. Specifically, major syntactic junctures occurring before conjunctions (e.g., [N1] # and [N2 or N3]; [N1 and N2] # or N3) were consistently aligned with Intonational Phrase (IP) boundaries, referred to as "**critical junctures**", with no exceptions. Furthermore, an "**optional**" occurrence of IP boundary after conjunctions (about 13%) was observed (e.g., [N1] # and (#) [N2 or N3]). Interestingly, the degree of preboundary lengthening was significantly greater before the critical IP boundary (Fig. 2d) compared to the optional one (Fig. 1d). Regarding glottalization, we perceived noticeable glottalization qualitatively for *both* IP-initial and IP-medial occurrences of the word-initial vowels. However, spectral tilt measures indicated prosodic boundary effects at the critical junctures, with an interaction with pitch-accent. Significant interactions between Boundary and Accent were found for H1H2c, HNR, and CPP. Specifically, conjunction vowels (‘and/or’) displayed increased glottalization at the critical juncture when pitch-accented (Fig. 2a). Conversely, at the optional IP boundary, no IP boundary effects were observed for vowels of nouns (e.g., [N1] # and (#) [N2 or N3]). This finding was further supported by the HNR and CPP measures for 'and,' where lower HNR and CPP values (indicating more noise) corresponded to increased glottalization at the critical juncture (Fig. 2b-c, upper panels). However, the directionality of HNR and CPP for 'or' exhibited inconsistency at certain points (refer to Fig. 2b-c, lower panels), indicating that noise-related measures did not yield clear results regarding the boundary effect on glottalization. It is also noteworthy that there was no evidence to suggest that pitch accent alone increased the degree of glottalization.

The findings of this study provide strong evidence for distinct phonetic variations in the realization of prosodic structure, which are dependent on different syntactic structures. Specifically, significant differences in glottalization and temporal expansion were observed between the critical IP boundary and the optional one. These results suggest that the phonologically-defined IP category [8] is phonetically modulated through the interaction of prosodic boundaries with syntax and prominence structure [3]. Overall, these findings highlight the intricate relationship between syntax, prosody, and phonetics. The observed fine-grained phonetic differences underscore the dynamic nature of language production, revealing that prosodic cues play a crucial role in disambiguating syntactic structures in a more nuanced manner than previously assumed in traditional syntax-prosody mapping.

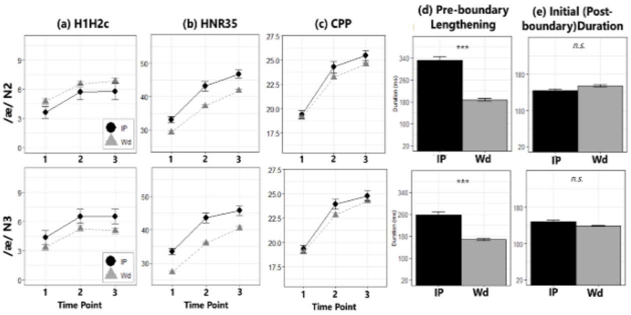


Fig. 1. Boundary effects with an *optional* IP boundary on H1H2c, HNR35, CPP, Pre-boundary lengthening and Initial (post-boundary) duration for the vowel /æ/ of the second and third names (N2, N3). Error bars represent standard errors. The lower the spectral tilt values, the more glottalized (creakier). Note pause duration was not included.

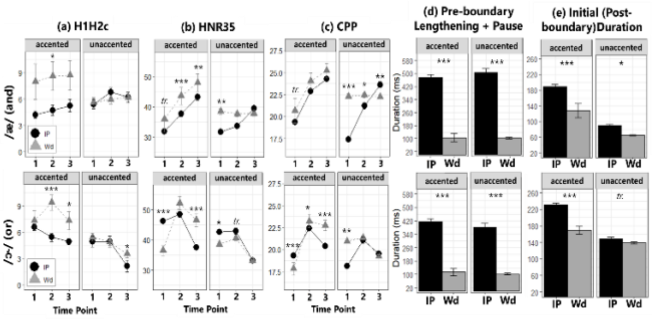


Fig. 2. Boundary effects at the *critical* juncture on H1H2c, HNR35, CPP, Pre-boundary lengthening and Initial (post-boundary) duration for the vowel /æ/ of ‘and’ and /ɚ/ of ‘or’ regarding pitch accent. Error bars represent standard errors. The lower the spectral tilt values, the more glottalized (creakier).

References

[1] Mitterer, H., Kim, S., & Cho, T. 2021. The Role of Segmental Information in Syntactic Processing Through the Syntax–Prosody Interface. *Language and Speech*, 64(4), 962-979.

[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] Elfner, E. 2018. The syntax-prosody interface: Current theoretical approaches and outstanding questions. *Linguistics Vanguard*, 4(1), 1-14.

[4] Cho, T. 2022. The phonetics-prosody interface and prosodic strengthening in Korean. In S. Cho and J. Whitman (Eds.), *The Cambridge Handbook of Korean Linguistics* (1st ed., pp. 248-293). Cambridge University Press.

[5] Davidson, L. 2021. The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews: Cognitive Science*, 12(3), e1547.

[6] Garellek, M. 2022. Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics*, 94, 101155.

[7] Shue, Y. L. 2010. *The voice source in speech production: data, analysis and models*. Los Angeles: University of Calinfornia

[8] Ladd, D. R. 2008. *Intonational phonology*. Cambridge University Press.

Variable realization of consonant clusters in Seoul and Gyeongsang Korean

Soohyun Kwon[1], Tae-Jin Yoon[2], Sujin Oh[3] & Jeong-Im Han[4*]

[1]Seoul National University (Korea), [2]Sungshin Women's University (Korea),
[3]University of Wisconsin-Milwaukee (USA),[4]Konkuk University
soohyunkwon@snu.ac.kr, tyoon@sungshin.ac.kr, sujinoh@uwm.edu, jhan@konkuk.ac.kr

Background In Korean, consonant clusters are realized either as the first (C1) or second consonant (C2) of the cluster before another consonant (*e.g. kulk.ta* 'to be thick'), conforming to the phonotactic constraints disallowing clusters in syllable coda as well as onset positions. The early studies showed that Consonant Cluster Simplification (CCS) in Korean is conditioned by cluster type ([1]) and dialect ([2]). Recent acoustic studies show further that an innovative variant of preserving both consonants (C1C2) has emerged and increased its frequency in /l/ + consonant (lC) clusters ([3,4]). These studies suggest that younger speakers show a preference to preserve C1, realizing CC as C1 or C1C2, despite the Standard Pronunciation rule dictating that CC should be pronounced as C2 in some cluster types. The present study aims to examine the realization of lC in Seoul and Gyeongsang Korean, using a large-scale spontaneous speech corpus.

Data & Methods The data come from the NIKL Korean Dialogue Corpus ([5]), a corpus of conversational interviews conducted with 2739 speakers from all parts of South Korea in 2020. From this 500-hour-long transcribed corpus, 1107 tokens of lC clusters (/lh, lt^h, lk, lp, lm/) were extracted using the Korean forced aligner ([6]). The coding was performed based on three phonetically-trained coders' auditory and acoustic analyses. Each token was coded as C1, C2, or C1C2 as well as for the factors known to condition the patterns of CCS. Mixed effects logistic regression models with random intercepts for speaker and verb stem were fit using the *glmer*() function in R to evaluate the effects of all factors as well as their interaction terms on the occurrence rate of C1-preserved variants (C1 & C1C2).

Results The results show that /lp/, /lk/ and /lm/ surface variably as C1, C2 or CC, whereas /lh/ and /lt^h/ clusters are realized categorically as C1. Also, CC in nominal stems (*e.g. talk* 'chicken') was categorically realized as C2. Further analyses, therefore, focus on the variable realizations of /lp/, /lk/ and /lm/ in verb stems. The distributional and statistical analyses provide the three main findings.

First, the realization patterns differ drastically by cluster type (p<0.001). Figure 1 shows that the occurrence rate of C1-preserved variants is significantly higher for /lp/ than /lk/(β = –2.32, p<0.05) and /lm/(β = –7.46, p<0.001) in both dialects.

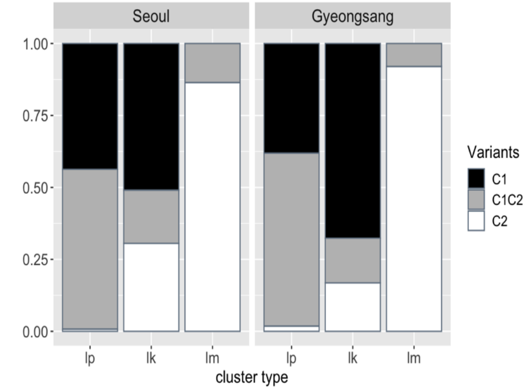
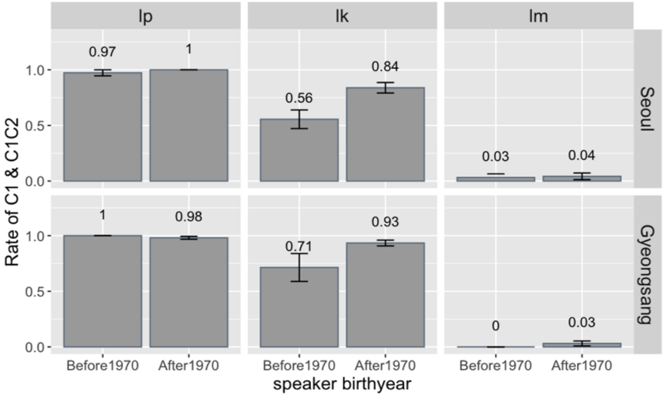


Figure 1. Relative frequency of variants by cluster type and dialect



Second, the trend toward preserving C1 shows an asymmetry depending on cluster type and dialect. Figure 2 reveals that, for /lp/, preserving C1 is prevalent both among older and younger speakers in both dialects; for /lk/, speakers born after 1970 show a significantly higher rate of preserving C1 than those born before 1970 (β =–1.75, p<0.01) in both dialects; for /lm/, preserving C1 is rarely observed.

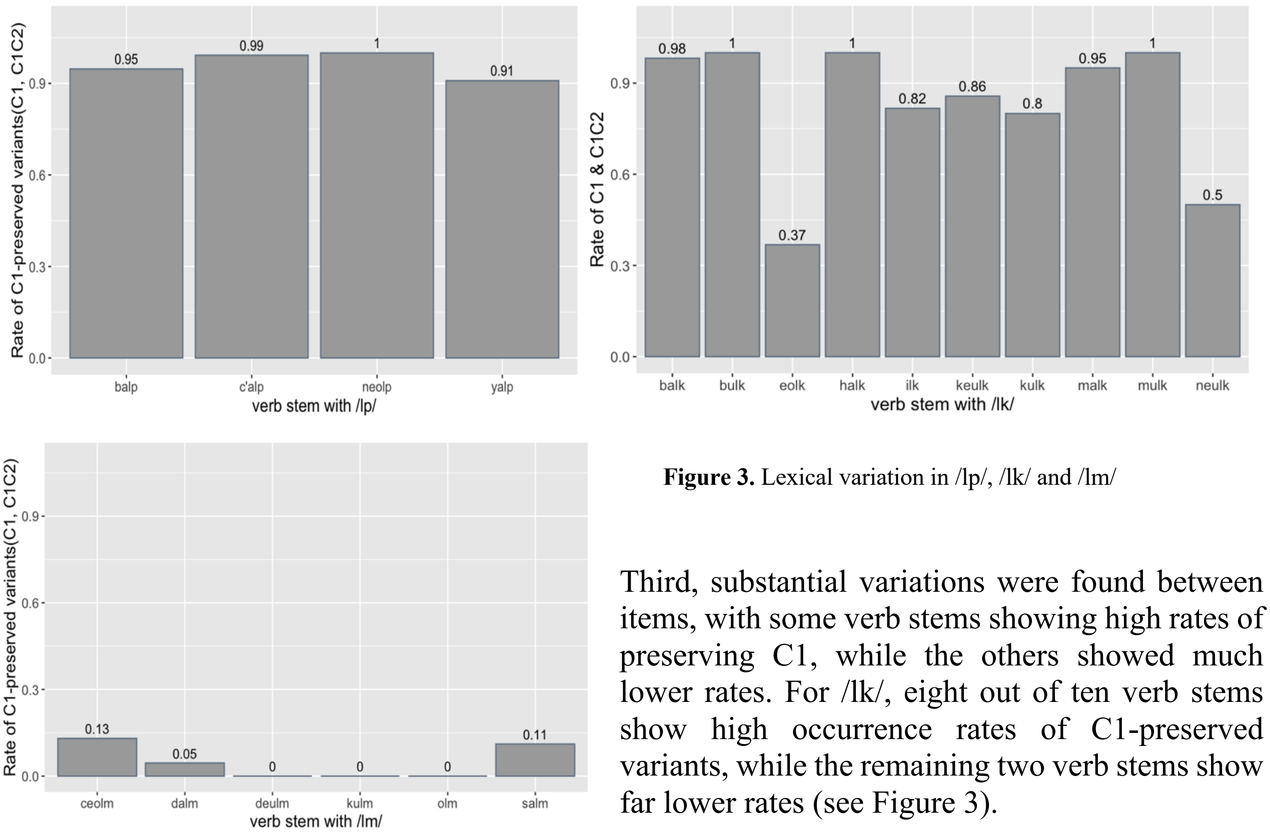


Figure 3. Lexical variation in /lp/, /lk/ and /lm/

Third, substantial variations were found between items, with some verb stems showing high rates of preserving C1, while the others showed much lower rates. For /lk/, eight out of ten verb stems show high occurrence rates of C1-preserved variants, while the remaining two verb stems show far lower rates (see Figure 3).

Discussion We argue that these findings represent mid-course patterns of an ongoing lexical diffusion that involves the expansion of a phonological process from one context to the next and from one item to another within each new context ([7], [8]). We speculate that the change was initiated in a limited context where the preference of C1 over C2 has a strong phonetic motivation, such as/lp/. Following Steriade's ([9],[10]) licensing by cue hypothesis, we suggest that speakers may prefer to select /l/ over /p/ on the ground that /l/ is perceptually more salient. The change toward preserving C1 in /lp/ appears to have almost gone to completion, with C1 preserved almost categorically. The innovation could have extended to /lk/ where the rate of preserving C1 appears to be vigorously increasing among younger speakers in both dialects. Within this new context, it is presumed to affect certain lexical items earlier than others. The change in /lm/ where the difference between two consonants in the strength of acoustic cues is small appears to be at a very early stage, with C1-preserved variants appearing in some verb stems only.

References

[1] Cho, T. (1999). Intra-dialectal variation in Korean consonant cluster simplification: A stochastic approach. *Chicago Linguistics Society* 35, 43–57.

[2] Whitman, J. (1985). Korean clusters. *Harvard Studies in Korean Linguistics* 1, 280–290.

[3] Cho, T. and Kim, S. (2009). Statistical patterns in consonant cluster simplification in Seoul Korean: Within-dialect interspeaker and intraspeaker variation. *Phonetics and Speech Sciences* 1(1), 33–40.

[4] Kim, J. J. and Kang, E. (2021). Phonetic variation of Korean stem-final consonant clusters beginning with a liquid. *Studies in Phonetics, Phonology and Morphology* 27(2), 161-192.

[5] National Institute of Korean Language (2022). NIKL Korean Dialogue Corpus (audio) 2020 (v. 1.3).

[6] Yoon, T-J. and Kang, Y. (2013). The Korean Phonetic Aligner Program Suite.

[7] Labov, W. (1994). *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

[8] Kiparsky, P. (1995). The phonological basis of sound change. *The Handbook of Phonological Theory*, 640, 670.

[9] Steriade, D. (1993). Neutralization and the expression of contrast. Paper presented at NELS 24.

[10] Steriade, D. (1997). Phonetics in phonology: the case of laryngeal neutralization. Ms., UCLA, Los Angeles.

Phonological Status of Voiced Fricatives in Fanchang Wu

Zihao Wei
zihaowei@cuhk.edu.hk
Department of Linguistics and Modern Languages, Chinese University of Hong Kong

This study investigates the phonological status of voiced fricatives in Fanchang Wu, a Xuanzhou group Wu dialect, using both empirical methods, cross-linguistic comparisional methods, and panchronic phonological methods.

The phonetic nature of voicing contrast of initial consonants in Wu dialects is long being a debatable question. Former research mainly focuses on Wu dialects distributed in coastal areas (i.e., Taihu group, Taizhou group, Oujiang group). Early impressionistic phonologists have different point of view on the phonetic quality of voiced fricatives in Xuanzhou Wu dialects, varying between aspirated voiceless fricatives (Fang, 1966; Zhengzhang, 1987; Meng 1988) and fricatives with a voiceless unaspirated first half and a voiced unaspirated second half (Shen, Hu & Meng, 1962). Zhu (2009) reports that the voiced fricatives in Jingxian Wu are voiceless fricatives followed by a strong aspiration based on spectrograms. Yuan (2019) reports that the voiced fricatives in Xinbo dialect are short devoiced fricatives with an aspiration behind. Hou and Chen's research (2019) shows that the existence of strong aspiration behind the voiceless fricative is a distinctive feature of voiced fricatives in dialects of Xuanzhou group. As a comparison, they point out that voiced initials in the surrounding dialects of Piling Wu group (coastal Wu) do not have a significant strong aspiration. Song (2012) studied the acoustic and articulatory features of voiced and voiceless fricatives in Xianju Wu (Taizhou group) and Wenzhou Wu (Oujiang group). The results show that voiced fricatives are different from voiceless fricatives in three aspects: shorter duration of closure, lower F0, and higher open quotient. Ling's research (2017) on Shanghainese voiced fricatives shows that Shanghainese, and many other Wu dialects are common in the acoustic features of voiced fricatives that the fricative segment can be largely or completely voiceless fricative.

Based on the discussion above, we can form two initial impressions: 1) Both voiced fricatives in coastal Wu dialects and in Xuanzhou Wu dialects are phonetically devoiced. 2) Voiced fricatives in Xuanzhou Wu dialects often cooccur with an aspiration, which is not seen in coastal Wu dialects. Since the phonetic nature of voiced fricatives is different in coastal Wu dialects and in Xuanzhou dialects, this raises a question: how to define he phonological status of voiced fricatives in Xuanzhou Wu dialects. The current study provides a better understanding of the phonological status of voiced fricatives Xuanzhou Wu dialects, using Fanchang Wu as an example. Simultaneous audio and electroglottographic recordings were made from six native Fanchang Wu speakers. A series of cue weighting processes using Linear Discriminant Analysis were done to investigate the relative importance of each phonological features. Based on the results, we point out that voiced fricatives in Fanchang Wu are devoiced in monosyllables, and the duration of the fricative segment is the primary cue in distinguishing voiced and voiceless fricatives. The existence of aspiration following the fricative segment of voiced fricative should only be treated as a secondary cue. A comparison of cross-linguistic experimental reports is then conducted with languages have a similar contrast between two phonetically devoiced fricatives like Korean (Cho, 2002; Chang, 2007) and Hmong (Wu & Wu, 2018) based on detailed and credible results using methods of experimental phonetics. The similar contrast in Korean and Hmong is treated by former scholars as a contrast between aspirated and unaspirated rather than between voiced and voiceless because of the primary importance of aspiration in these languages. Our results show that the contrastive pattern of fricatives in Fanchang Wu is significantly different from the pattern in Korean and Hmong because of the relative unimportance of aspiration, and should be treated as a voicing contrast rather than aspiration contrast. We further discuss the contrastive pattern in Fanchang Wu should not be treated as aspiration contrast based on the paradigm of Panchronic Phonology (Jacque, 2011). The aspiration element of voiced fricatives in Fanchang Wu neither has a transparent correspondence to a diachronic origin, nor can it represent the voiced fricatives as a whole phonological category. The present study strongly revises the views of previous scholars on voiced fricatives in Xuanzhou Wu dialects: These phonemes still form an independent "voiced" category and should not be treated as "voiceless aspirated". The present study also provides a new idea when dealing with the problem of defining phonological status of specific phonemes in a given language: combining empirical research, cross-linguistic comparison and panchronic phonology is not a bad idea to try.

Reference

[1] Chang, C. B. Korean fricatives: Production, perception, and laryngeal typology, *UC Berkeley Phonology Lab Annual Report*, 2007.

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

[3] Fang, J., An Impressionistic Description of Fangcun Phonology in Wuhu County, *Chinese Languages*, 1966, 2.

[4] Fu, G., Cai, Y., Bao, S., Fang, S., Fu, Z., Zhengzhang, S., The Dialectal Distribution in Southern Anhui Province, *Dialect*, 1986, 1.

[5] Hou, C., Chen, Z. M., Xuanzhouwuyu zhuoyinshengmu de shengxueshiyanyanjiu——yi suwan jiaojiediqu fangyan weili (An Acoustic Study on the Voiced Initials in Xuanzhou Wu Dialects——An Example from Dialects on the Boarder of Jiangsu and Anhui), *Chinese Language*, 2019, 6.

[6] Jacques, G. A panchronic study of aspirated fricatives, with new evidence from Pumi, *Lingua, Elsevier, 2011*, 121(9), 1518-1538.

[7] Ling, F. Xinpai shanghaihua quanzhuocayin zhong de qingyinchengfen fenxi (An Analysis on the Voiceless Element of the Voiced Fricatives in New Shanghainese), *Chinese Language*, 2017:746.

[8] Meng, Q., The Relation between Tongling-Taiping Dialects in Southern Anhui Province and Wu Dialects, *Thesis Series on Wu Language*, Shanghai: Fudan University Press, 1988.

[9] Shen, S., Hu, Zh., Meng, Q., *The Profile of Anhui Dialects*, Hefei: Hefei Normal University dialect working group, 1962.

[10] Song, Y., *An Articulatory Research on Obstruent in Wu Language based on EGG Data*, Beijing: World Publishing Corporation, 2012.

[11] Wu, S. Y., Wu, L. Ganrong Hmongyu songqicayin de shengxuetedian yu gongshibianyi (The Acoustic and Variations of Aspirated Fricatives in Ganrong Hmong), *A collection of Language studies*, 2019.

[12] Yuan, D. Wannanwuyu tongjingpian songqicayin s^{h}-ɕ^{h}- de laiyuan ji yinbian (The Origin and Sound Change of Aspirated Fricatives in Wu dialects of Su-Wan Boundary, *Chinese Language*, 2018, 1.

[13] Zhu, L., On the Tongling-Jingxian Type in the Change of MC Voiced Initials of Wu Dialect in Xuanzhou cluster in Anhui Province, *Dialect*, 2009, 2.

Relations between Opinion Convergence, Acoustic Convergence and Movement Convergence in Interlocutors

Charlize Ma, Effie Kao, Raechel Kitamura, Stephanie Wang, Jahurul Islam, Gillian De Boer, Bryan Gick
Department of Linguistics, University of British Columbia (Canada)
sy.charlize@gmail.com

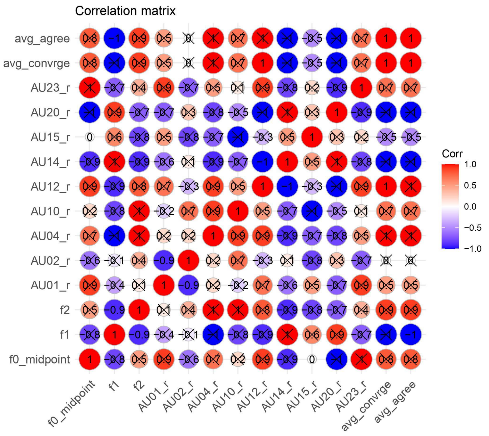
Background: Speakers in a conversation (interlocutors) can exhibit convergent behaviours in a variety of ways, including influencing one another's speech acoustics, movements, and opinions. Past research shows that interlocutors appear to converge in a descending F_0 pattern nearing the end of a conversation [1]. Additional research has also shown that speakers tended to imitate each other's changes in F_0 across turns during a turn-taking reading task [2]. Notably, individuals who perceived a Voice User Interface (VUI) as having the same opinion and characteristics as themselves had an increased likelihood of convergence [3]. Furthermore, the degree of closeness in the relationship between interlocutors appeared to be a factor in the polarization of their opinions [4]. Most of the research into speech convergence has been focused on acoustics, but there have been few attempts to assess if the same applies to visual cues, like lip and eyebrow movement. Past studies have found that our facial movements change during speech depending on our interlocutor. Lip movements were observed to increase significantly during infant-directed speech [5] and in congenitally blind speakers [6]. We sought to discover whether facial movement and speech convergence could be linked to the convergence of opinions.

Methods: 36 participants (M:9, F:27) above the age of 18 were recruited. Each participant was randomly paired with another participant to have a short conversation (3-5 mins) in a Zoom meeting where they discussed their views of online vs. in-person schooling. At the end of the conversation, they completed a questionnaire asking how much they thought their opinion converged with their partners' (convergence), and how much they agreed with each others' ideas (agreement), on a 7-point Likert scale. The whole conversation process was videotaped and recorded using Zoom's recording system.

OpenFace 2.0 [7] software was used to extract lip and eyebrow movement information from the video data. The first and last minutes of the conversation were selected to generate differences in action units (AUs) in each dyad. 9 AUs were targeted (brows: 1, 2, 4; lips: 10, 12, 14, 15, 20, and 23). Audio data was transcribed and force-aligned using Montreal Forced Aligner [8]. Acoustics values (F_0, F1 values etc.) were extracted from the vowel midpoints using Praat [9]. Acoustic data was synchronized with the facial movement data from OpenFace 2.0 using timestamps.

From this acoustic data, plots for seven vowels (ɪ, i, ɛ, ɑ, ɔ, ʊ, u) were examined to aid in visualizing the relationship between specific vowels and dependent variable values from the experiment, namely the Likert scale data taken from the questionnaire after the discussion and facial movement differences. The average agreement and convergence values from the Likert scale after the conversation for each dyad was then calculated.

Results: A correlation matrix was run on the acoustics values from Praat, the AU values from OpenFace, and the average Likert scale data. In the matrix in Figure 1, circles that are crossed out denote non-significance. A significant positive correlation was found between lip corner pull (AU12_r) and average convergence (avg_converge) (r = 1, p < .001), and a significant negative correlation was found between F1 values and average agreement (avg_agree) (r = -1, p = .018). However, there was no overall difference observed between F_0 values and AUs within participants in a conversation from their first to last minutes of conversation. A box plot was generated to display the differences between the first and last minutes of conversation for each AU as well as F_0 (Figure 2). Additionally, a U-Test was run using R [10], that indicated no significant difference between F_0 values and AUs (*p* > .05 for all comparisons).



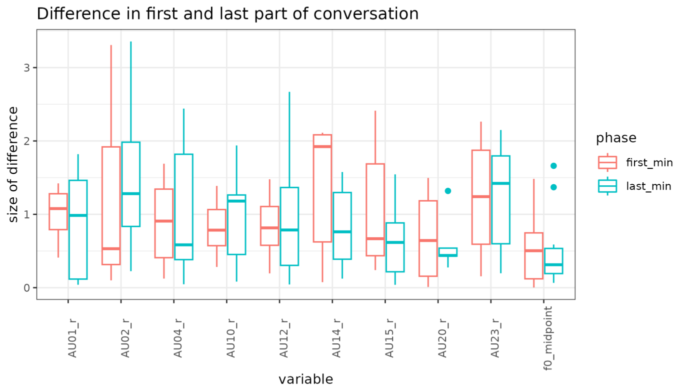


Fig. 1 Correlation Matrix of all variables

Fig. 2 Box plot of AU differences in first vs. last min

Discussion: Our initial analysis shows a correlation between a consensus of agreement among participants and increased lip corner pulling (AU 12). This could possibly demonstrate a relationship between opinion convergence and facial movements (in this case, smiling). Additionally, the correlation between participants who agreed more and those who exhibited higher vowel height (through acoustic analysis) could indicate that participants expended more effort in trying to converge with their interlocutor. However, the vast majority of facial action units analyzed did not appear to be affected by opinion convergence, suggesting that speech convergence and opinion convergence appear to work largely independently. The lack of significant F_0 convergence shows different results from that of previous literature [1], but there is room for further investigation with regard to interactions between facial movements and opinion convergence.

References

[1] Yang, Li-Chiung. (2013). Prosodic convergence, divergence, and feedback: Coherence and meaning in conversation. 27th Pacific Asia Conference on Language, Information, and Computation, PACLIC 27. 85-91.

[2] Aubanel V, Nguyen N. Speaking to a common tune: Between-speaker convergence in voice fundamental frequency in a joint speech production task. PLoS One. 2020 May 4;15(5):e0232209. doi: 10.1371/journal.pone.0232209. PMID: 32365075; PMCID: PMC7197779.

[3] Farr, C., Purnomo, G., Cardoso, A., Shamei, A. & Gick, B. (2021). Speaker accommodations and VUI voices: Does human-likeness of a voice matter? In Proceedings of the XVIIth Conference of Associazione Italiana di Scienze della Voce, 63-64.

[4] Balietti, S., Getoor, L., Goldstein, D. G., & Watts, D. J. (2021). Reducing opinion polarization: Effects of exposure to similar people with differing political views. Proceedings of the National Academy of Sciences, 118(52), e2112552118. https://doi.org /10.1073/pnas.2112552118

[5] Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. Journal of Speech, Language, and Hearing Research, 53(6), 1529-1542. https://doi.org/10.1044/1092-4388(2010/09-0005)

[6] Ménard, L., Leclerc, A., & Tiede, M. (2014). Articulatory and acoustic correlates of contrastive focus in congenitally blind adults and sighted adults. Journal of Speech, Language, and Hearing Research, 57(3), 793-804. https://doi.org/10.1044/2014_JSLHR-S-12-03

[7] OpenFace: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.

[8] MFA: McAuliffe, Michael, Michaela Socolof, Elias Stengel-Eskin, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger (2017). Montreal Forced Aligner [Computer program].

[9] Boersma, Paul & Weenink, David (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.07, retrieved 6 February 2023 from http://www.praat.org/

[10] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Effects of consonantal contexts on L2 English tense-lax vowel perception and production

Yung-hsiang Shawn Chang

National Taipei University of Technology (Taiwan)
shawnchang@mail.ntut.edu.tw

Studies that examine cross-language speech perception and production often use stimuli with the target segments embedded in single or limited phonetic contexts [1]. However, L2 perceptual and production accuracy have been found to vary greatly as a function of the phonetic contexts where the target L2 sounds occur (e.g., [2, 3]). Therefore, as argued in Beddor et al. ([4]), generalizations made based on perception and production patterns of the target sound in one phonetic context, without systematically exploring contextual effects, may be problematic. In this light, this study used the case of English tense and lax vowels, which Mandarin speakers generally find difficult to learn (e.g., [5, 6]), to examine whether L2 vowel perception and production performances vary as a function of consonantal contexts.

Fifteen native speakers of Taiwan Mandarin (between 19-20 years of age) participated in this study. The stimuli were 58 CVC real words in English, produced by two native American English speakers (1 male; 1 female). The stimuli contained one of the tense /i, e/ vowels or their lax counterparts /ɪ, ɛ/, flanked by consonants of different places (bilabial, alveolar, velar) and manners of articulation (liquid, nasal, sibilant, stop). The perception test was an AX discrimination task, where listeners judged whether the two words they heard were the same or different. The 58 words were arranged into 116 AX trials; within each trial, the two words were always produced by different native speakers. The perceptual accuracy was automatically logged in E-Prime 3. The production test was a repetition task, where the listeners repeated what they heard through the headphones. The production accuracy was determined by whether a given token was correctly identified by two native American English listeners. Any disagreement between the two listeners was resolved by resorting to a third listener's judgment.

The perceptual and production accuracy data were separately analyzed with generalized logistic mixed-effects regression models. The full model included vowel, onset consonants' place of articulation (POA), onset manner of articulation (MOA), coda POA, coda MOA, and their interactions as fixed effects, and the participant intercept as the random effect. The results for the perceptual data analysis revealed a significant coda MOA effect and a coda MOA*coda POA interaction effect. Post-hoc comparisons showed that vowels followed by alveolar nasal codas were discriminated with significantly lower accuracy than all other place-manner combinations (see Figure 1). The results for the production data analysis indicated significant vowel and coda MOA effects and their interaction. In particular, the accuracy of /e/+nasal coda and /ɪ/+nasal coda productions was significantly lower than any other vowel+consonant combinations (see Figure 2).

Nasalization as a result of coarticulatory influences of nasal consonants modifies vowel formant frequencies (e.g., [7, 8]) and can cause certain vowels in pre-nasal position to be less distinct from neighboring vowels. While native speakers may compensate for nasality-related changes in formant structure by recruiting variation in additional acoustic properties [9], this study shows that the nasal context (i.e., coda) was where Mandarin-speaking learners had most trouble discriminating and producing English tense and lax vowels. These findings are in line with [1] that cross-language perception and production of vowels is context-dependent and cannot be predicted from patterns observed in single or limited contexts. The findings also have pedagogical value in that target vowel contrasts that differ in difficulty levels can be generated for teaching or training stimuli.

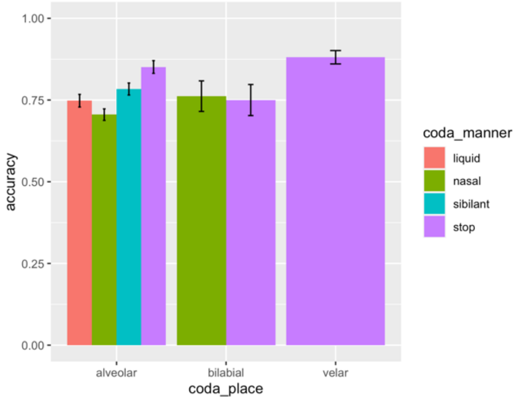


Fig.1. Vowel perception accuracy by coda consonants' place and manner of articulation

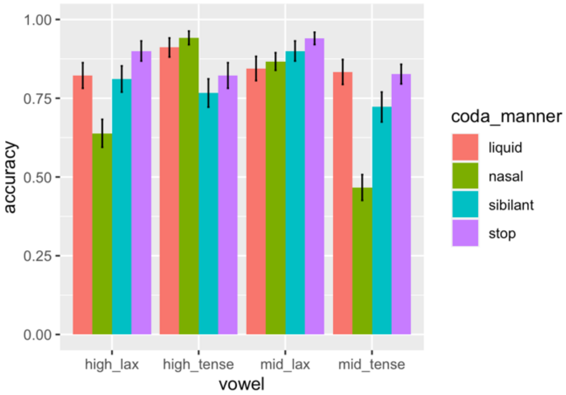


Fig.2. Vowel production accuracy by vowels and coda consonants' manner of articulation

References

[1] Bohn, O. S., & Steinlen, A. K. (2003). Consonantal context affects cross-language perception of vowels. In *Proceedings of the 15th International Congress of phonetic Sciences*.

[2] Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America, 128*(3), 1290-1305.

[3] Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America, 109*(4), 1691-1704.

[4] Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*(4), 591-627.

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

[6] Wang, X. (2002). Training Mandarin and Cantonese speakers to identify English vowel contrasts: Long-term retention and effects on production. Doctoral dissertation, Simon Fraser University.

[7] Carignan, C. (2018). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *The Journal of the Acoustical Society of America, 143*(5), 2588-2601.

[8] Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America, 142(*4), 2469-2482.

[9] Zellou, G., Barreda, S., & Ferenc Segedin, B. (2020). Partial perceptual compensation for nasal coarticulation is robust to fundamental frequency variation. *The Journal of the Acoustical Society of America, 147*(3), EL271-EL276.

The Effect of Cue-specific Lexical Competitors on Hyperarticulation of VOT and F0 Contrasts in Korean stops

Cheonkam Jeong[1] & Andrew Wedel[1]

[1] *University of Arizona (USA)*
cheonkamjeong@arizona.edu, wedel@arizona.edu

Standard Korean has an unusual three-way laryngeal distinction in stops: aspirated (e.g., /p^h/, 'grass'), lenis (e.g., /pul/, 'fire'), and fortis (e.g., /p*ul/, 'horn'). VOT is historically a primary cue distinguishing these contrasts, with fundamental frequency (F0) of the following vowel as a secondary cue: in accentual phrase-initial positions, aspirated stops are associated with the longest VOT and higher F0 on the following vowel, lenis stops with an intermediate VOT value and lower F0, and fortis stops with the shortest VOT and higher F0 [1]. This work focuses on the aspirated~lenis contrast.

Many, particularly older, speakers of Seoul Korean produce both robust VOT and F0 distinctions between aspirated and lenis stops [2]. However for many speakers of Seoul Korean a transphonologization is in progress in which the VOT cue progresses toward neutralization between aspirated and lenis stops, with a concomitant expansion of the F0 contrast [2], leaving lexical contrasts intact. A variety of factors have been found to predict aspects of this sound change in progress. At the speaker level, lower age and female gender are associated with a more advanced position in this sound change [2], while at the lexical level, higher word frequency and lower following vowel are associated with more advanced position, that is, a smaller VOT and greater F0 contrast [3].

The potential role of lexical functional load in this sound change, however, has not yet been investigated. Theoretical models of speech production and perception within a community predict that phonemic contrasts will be relatively hyperarticulated when they carry more lexically disambiguating information, which over diachronic time can influence the trajectory of sound change [4]. In support of these models, a greater number minimal pairs (i.e., pairs of words distinguished by a single phonemic contrast) has been shown to be significantly correlated with preservation of that phonemic contrast over time [5]. At the level of individual speech production, contrastive hyperarticulation in minimal pairs has been found in English consonant VOT contrasts (e.g., 'pat'~'bat') as well as in vowel formant contrasts (e.g., 'lift'~'left') [6]. The same effect has also been found in Japanese singleton and geminate consonants (e.g., /kata/, 'frame'~/katta/, 'bought') [7]. The ongoing sound change in Korean stops provides an opportunity to further explore the predicted connection between the lexicon, usage-level variation and systemic change in the sound system. Here, we use a production experiment to ask whether minimal pair status influences variation in production of VOT and F0 in aspirated versus lenis stops, and whether the degree of variation is influenced by a speaker's position in this sound change. Based on the previous body of theoretical and experimental work, we expect that both VOT and F0 will be hyperarticulated, but that speakers more advanced in the sound change may show greater hyperarticulation of F0 relative to VOT.

We identified minimal pairs of the same syntactic category distinguished by the aspirated~lenis stop contrast in word-initial position from the Korean National Database, as well as a set of aspirated/lenis stop-initial words balanced for following vowel that do not have minimal pair competitors. A total of 78 word sets were inserted into 220 meaningful, declarative sentences, controlling honorific degree. A shorter version was also made by removing 92 sentences from the set for elder speakers who had difficulty producing the larger set. Participants comprised 41 adult Seoul Korean speakers, 26 females and 15 males, born between 1932 and 1996. All but one female speaker and nine speakers who read the short list version produced all of 220 sentences, thereby totaling 7,703 utterances. VOT was hand-annotated, and F0 extracted using Praat's built in function. F0 was semitonized and normalized relative to the F0 of the vowel in the second syllable in the word to control for prosodic context. To identify a speakers' position within the sound change, we introduce a novel measure comparing the degree of use of the VOT contrast to the degree of use of the F0 contrast. This measure was calculated over a speaker's set of non-minimal pair word productions as in Equation (1), where F is the difference between the median of the z-scored F0 values of a speaker's aspirated stops and that of their lenis stops in non-minimal pairs, and V is the corresponding difference in aspirated and lenis VOTs.

$$\text{Position} = F/(V + F) \qquad (1)$$

Values of 'Position' for this set of speakers range from near 0.5, indicating that the speaker produces a balanced F0 and VOT contrast between aspirated and lenis stops, to near 1, indicating that the speaker produces solely an F0 distinction.

VOT and normalized F0 were separately modeled as a function of age, gender, lexical frequency, speech rate, vowel height, 'position', and presence of a minimal pair, using Bayesian hierarchical models fitted in Stan via the R package *brms*. Based on initial analysis of global models, we analyzed models with individual laryngeal class and vowel height levels fitted with the same syntax as that of the global models. Analysis shows that the aspirated~lenis contrast is significantly hyperarticulated in minimal pairs for both VOT and F0 cues, but only in aspirated stops with following high vowels. In addition, speakers advanced in the sound change produce significantly lower F0 values for lenis stops with following non-high vowels if there is an aspirated minimal pair competitor. These results extend our understanding lexical contrast-driven hyperarticulation by showing contrastive VOT and F0 hyperarticulation in minimal pairs in a language other than English. We find that the F0 distinction in lenis stops is relatively more hyperarticulated in speakers who are more advanced in the sound change consistent with predictions, but that speakers appear to hyperarticulate VOT regardless of how much of a VOT contrast they normally produce themselves. This may be related to the fact that VOT is still a robust cue to the aspirated~lenis distinction in much of the speech community - if so, this result is consistent with listener-oriented models of contrastive hyperarticulation [8].

References

[1] Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

[2] Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*(2), 287-308.

[3] 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.

[4] Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition*, *4*(4), 319-355.

[5] Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179-186.

[6] Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, *100*, 61-88.

[7] Sano, S. (2018). Durational contrast in gemination and informativity. *Linguistics Vanguard: Multimodal Online Journal, 4*(2), Linguistics vanguard : multimodal online journal, 2018, Vol.4 (2).

[8] Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers’ subsequent pronunciations. Journal of Memory and Language, 89, 68–86.

Effects of Fine Phonetic Detail on Speaker Identification from Japanese Nasal Consonants

Ai Mizoguchi[1,2], Mark K. Tiede[3,4] & D. H. Whalen[3,4,5]

[1]Maebashi Institute of Technology (Japan), [2]NINJAL (Japan), [3]Haskins Laboratories (USA),
[4]Yale University (USA), [5]City University of New York (USA)
aimizoguchi@maebashi-it.ac.jp, mark.tiede@yale.edu, douglas.whalen@yale.edu

The acoustic characteristics of nasal consonants are more complicated than those of oral consonants due to involvement of both the oral and nasal vocal tracts. Theoretically, the oral cavity acts as a side-tube because the oral articulators block the airflow for a nasal consonant and the airflow is emitted only from the nostrils. As a result, the acoustic information of the place of articulation (PoA) achieved in the oral cavity appears only as anti-resonances or anti-formants in nasal acoustics [1]. The formants, not anti-formants, of nasal consonants are considered to be rather stable, reflecting the shapes of the pharynx and nasal cavity, which are largely dependent on the anatomy of each speaker. This is supported by research that indicates nasal consonants were better perceived for speaker identification [2, 3].

However, acoustic analyses have revealed some tendencies of nasal formants depending on the PoA in several languages: frequency values of the first formant (N1) are the highest for [ŋ] and lower for [ɲ], [n], and [m] in that order [4 for review]. [3] assumed the larger role for the oral cavity and examined the acoustics of /m/ and /n/ produced by Dutch speakers. It was shown that phonetic context and syllabic position affected the nasal acoustics, especially on the second formant (N2) and the spectral center of gravity (CoG) (N1 was unmeasurable due to the loss of frequencies below 300 Hz in telephone utterances).

The inter-speaker variability of the articulation of the nasal consonant /N/ (moraic nasal) in Japanese has been reported [5]. In this study, the acoustics and articulation of nasal consonants in Japanese were investigated to determine whether speaker variability predicts speaker identity.

Ten native speakers of Japanese (6 females, 4 males) participated in the experiment. The target phonemes were intervocalic nasal consonants in the words, /amata/ 'many', /anata/ 'you', /aŋa/ 'lay down', and /kaNaN/ 'consideration'. The participants read aloud the words displayed on the screen one at a time in a random order and each word was repeated 10 times. The audio signal, ultrasound video showing the midsagittal image of the oral tract, and motion measurement data were recorded simultaneously. The Haskins Optically Corrected Ultrasound System (HOCUS) [6] was used to align tongue contours obtained from ultrasound images to palatal hard structure.

N1, N2, N3, and CoG were measured at the midpoint of each target phoneme using Praat [7]. The tongue contour of the midpoint was traced using GetContours [8]. The highest point of the tongue contour was identified and the values on the horizontal axis (highest_X) and vertical axis (highest_Y) were used for statistical analyses. Multinomial logistic regression analyses were performed using the 'nnet' package [9] in R [10].

Table 1 shows the result of the multinomial analysis predicting speakers from acoustic variables (N1, N2, N3, and CoG) and articulatory variables (highest_X and highest_Y). All the variables except N2 showed multiple significant effects on identifying speakers. The multinomial analysis was also carried out for phoneme prediction, using the same variables as for the speaker prediction. Fig. 1 shows the correct prediction rates for speakers and phonemes by each variable. The correct prediction rate was 89.1% for speakers and 79.0% for phonemes when all variables were used. The acoustic variables were more relevant for the speaker prediction than the articulatory variables and vice versa for the phoneme prediction. N1 seems to reflect some PoA information as it predicted 40.8% of the phonemes correctly. Speaker specificity seems to be most evident on N3, as N3 itself predicted speakers better than the other variables. Relationships between acoustics and articulation will be investigated in future analyses.

Table 1. Statistical output for multinomial analysis of speaker predictions.

*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: Speaker								
Coefficients:	JF03	JF04	JF05	JF06	JF07	JM01	JM02	JM03	JM04
N1_Hz	0.099***	0.058***	0.063***	0.016	0.153***	0.017	0.079***	-0.006	0.457***
N2_Hz	-0.001	0.003	-0.006	-0.005	-0.029***	-0.0003	-0.006	0.006	-0.017
N3_Hz	-0.045***	-0.042***	-0.058***	-0.140***	-0.165***	-0.046***	-0.052***	-0.051***	-0.278***
N_cog	-0.123***	-0.023	-0.058***	0.081	-0.535***	-0.005	-0.249***	-0.092***	0.259***
highest_X	0.375***	0.405***	0.310***	0.637*	0.954***	0.515***	0.405***	0.412***	3.783***
highest_Y	-0.961***	-0.531***	-0.770***	-1.104**	-1.343***	0.374	0.085	0.318	3.703***

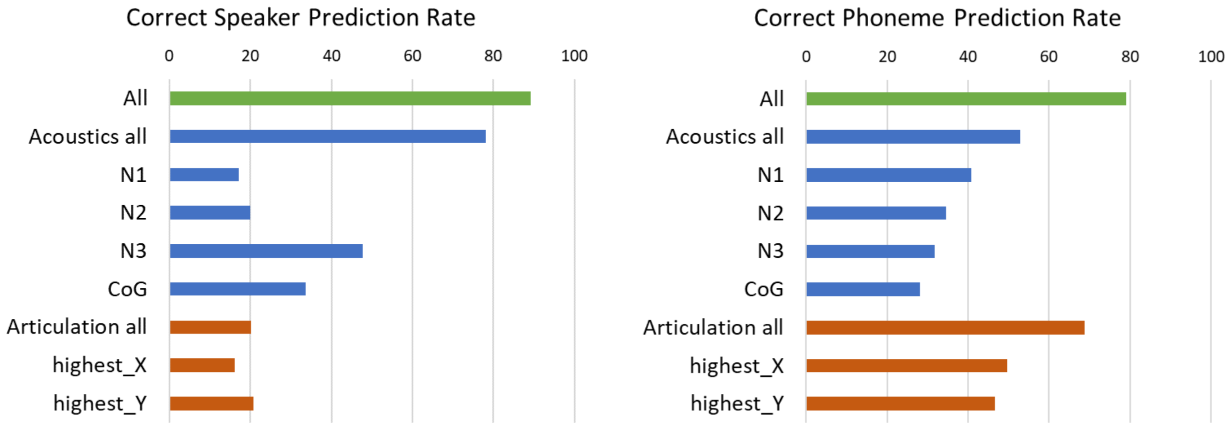


Fig.1 Correct prediction rate for speakers (left) and for phonemes (right).

References

[1] Johnson, K. (2012). *Acoustic and auditory phonetics, 3rd ed*. Malden, MA: Wiley-Blackwell.

[2] Amino, K., & Arai, T. (2009). Speaker-dependent characteristics of the nasals. *Forensic Science International, 185(1–3)*, 21–28.

[3] Smorenburg, L., & Heeren, W. (2021). Acoustic and speaker variation in Dutch /n/ and /m/ as a function of phonetic context and syllabic position. *The Journal of the Acoustical Society of America, 150(2)*, 979–989.

[4] Recasens, D. (1983). Place cues for nasal consonants with special reference to Catalan. *The Journal of the Acoustical Society of America, 73(4)*, 1346–1353.

[5] Mizoguchi, A., Tiede, M. K., & Whalen, D. H. (2022). Inter-speaker variability of articulation for the Japanese moraic nasal: An ultrasound study. *Phonological Studies, 25*, 121–132.

[6] Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research, 48(3),* 543–553.

[7] Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer. Version 5.2.46, retrieved 7 October 2011, http://www.praat.org/.

[8] Tiede, M. K. (2016). GetContours. GitHub repository, https://github.com/mktiede/GetContours.

[9] Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S, 4th ed*. New York: Springer.

[10] R Core Team. (2019). R: A language and environment for statistical computing.

A generative phonetic approach to the ongoing sound change in Kyengsang Korean

Yeong-Joon Kim[1]

[1]*Massachusetts Institute of Technology (USA)*
joonkim@mit.edu

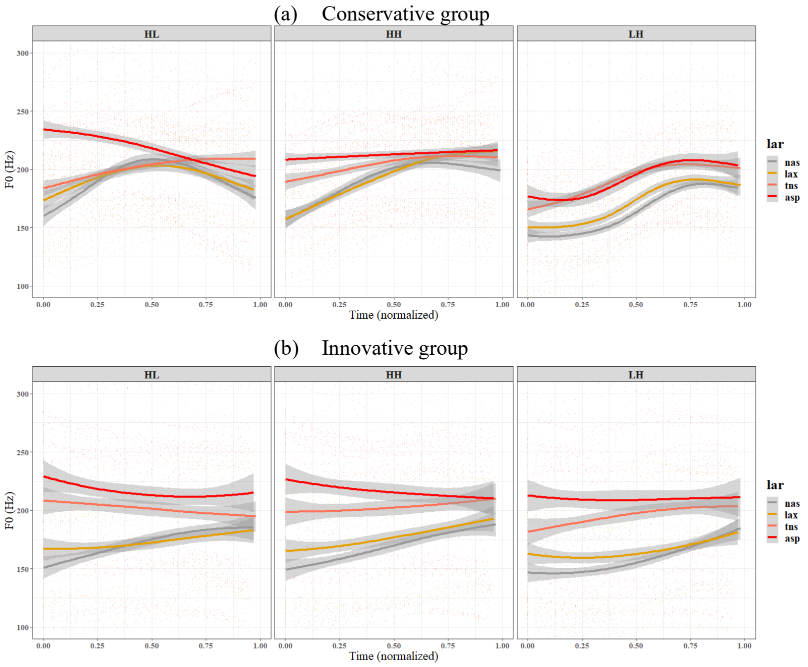
Outline: While the redistribution of cue weights from VOT to F0 in three-way laryngeal contrast has been well documented in many Korean dialects [1], its effect on dialects with a pitch-accent system, such as Kyengsang Korean, is less clear. However, it is expected that the F0 cues for both laryngeal and tonal contrasts will have mutual influence given the difficulty of rapid pitch changes in a limited time span [2, 3], and this interaction indeed has been reported for Kyengsang Koeran [4, 5]. This study aims to test two competing hypotheses regarding the long-term consequences derived from this physiological constraint. The first hypothesis predicts that the tonal contrast in Kyengsang Korean will not be able to use F0 to maintain unambiguous pitch distinctions as the cue becomes devoted to discriminating the laryngeal categories. In contrast, the second hypothesis proposes that the laryngeal contrast in the dialect will not successfully utilize F0 as a substitute for VOT, as F0 is dedicated to maintaining the tonal contrast.

Experiment: To evaluate these hypotheses, an experiment with three different age groups was conducted. 24 North Kyengsang Korean speakers from Taykwu (Daegu) or nearby regions participated in a speech production experiment, divided into three age groups: 20s-19 (innovative), 40s-50s (transitional), and 70s (conservative), each with four female and four male speakers. C1V1C2V2(C) disyllabic stimuli were balanced with regard to the C1 laryngeal categories (nasal, lax, tense, and aspirated) and the accentual categories (HL, HH, and LH). Subjects who participated in a speech production experiment were exposed to the test items embedded in a sentential frame "*kulimey*___*pointa* (___is seen in the picture)." Participants repeated a sequence of 24 words twice.

Results: Four phonetic dimensions were measured: LarF0 (onset F0 of V1 voicing), AccF0 (F0 at accentual peak), AccPT (accentual peak timing), and VOT of C1. Generational differences were observed in all these dimensions. The results showed that the realizations of LarF0, AccF0, and AccPT were increasingly dependent on the laryngeal categories as age decreased, while the difference in the measurements based on the accentual categories became less distinct (Figure 1). These changes in F0 realizations were accompanied by the gradual VOT merger between lax and aspirated stops. Mixed-effects regressions confirmed this trend. For instance, among the significant statistical results, there were significant differences in the mean AccF0 values between aspirated stops and other stop types across age groups: The transitional group was distinct from the conservative group ($\beta = 12.42$, $t = 4.51$), and the innovative group was distinct from the two older groups ($\beta = 5.78$, $t = 2.46$). The overall findings supported the first hypothesis that the tonal contrast will become less distinctive as the laryngeal contrast takes advantage of the F0 cue.

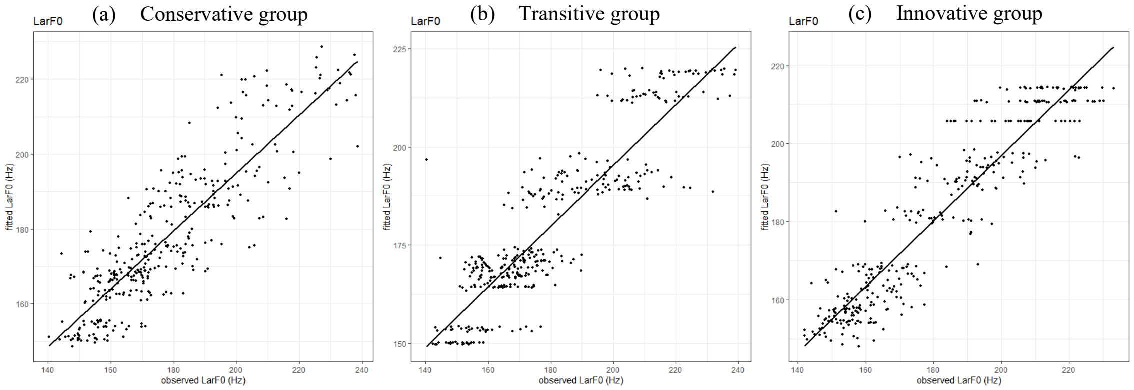
Modeling: The results were analyzed using weighted constraints within a generative phonetics framework [3, 6] to explain how these cue redistributions lead to a potential loss of the language's tonal contrast. In this model, phonetic goals are subjected to compromise. For instance, the realizations of LarF0 and AccF0 need to be reconciled with each other as a function of their temporal distance to avoid a marked F0 change in a short period of time. This model can adequately capture the two hypotheses expected under the physiological constraint on the realization of F0 contours. If the first hypothesis is true, the realization of the LarF0 target incrementally outweighs that of the AccF0 target in a model; if the second hypothesis is true, a model can be constructed with reverse weighting. The overall model trained based on the experimental results captured the different patterns in F0 realizations over the generations relatively well (Figure 2).

Summary: This research confirms the presence of a sound change taking place in Kyengsang Korean. The experimental results over the generations showed that the laryngeal contrast based on F0 enhanced while the tonal contrast weakened as age decreased. This trend was explained through a phonetic grammar formalized with weighted constraints: The ongoing sound change can be seen as a process in which the realization of LarF0 becomes relatively prioritized under the grammatical controls.


(a) Conservative group

(b) Innovative group

Fig.1 Schematic illustrations of the pitch contours for the conservative group (a) and the innovative group (b). The observed F0 values were plotted against the normalized time. The representative contour shapes in colored lines were derived from the data by loess method.


(a) Conservative group

(b) Transitive group

(c) Innovative group

Fig.2 Scatter plots of fitted values (y-axis) against observed values (x-axis) for LarF0 over three generations. The model used was $w_{\mathrm{LarF0}}(\mathrm{LarF0} - \mathrm{T_{LarF0}})^2 + w_{\mathrm{AccF0}}(\mathrm{AccF0} - \mathrm{T_{AccF0}})^2 + w_{\mathrm{AccPT}}(\mathrm{AccPT} - \mathrm{T_{AccPT}})^2 + w_{\mathrm{S}}((\mathrm{M/AccPT}) - \mathrm{T_S})^2 + w_{\mathrm{M}}(\mathrm{M} - \mathrm{T_M})^2$, where w, T, S, and M stand for weight, target, slope, and magnitude, respectively.

References

[1] Lee, H., J. Holliday, & E. Kong. 2020. Diachronic change and synchronic variation in the Korean stop laryngeal contrast. *Lang. Linguist. Compass* 14(7): 1-12.

[2] Xu, Y. & X. Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *JASA* 111: 1399-1413.

[3] Flemming, E. & H. Cho. 2017. The phonetic specification of contour tones: evidence from the Mandarin rising tone. *Phonology* 34: 1-40.

[4] Kenstowicz, M. & C. Park. 2006. Laryngeal features and tone in Kyungsang Korean: A phonetic study. Studies in *Phonetics, Phonology and Morphology* 12: 247-264.

[5] Lee, H. & A. Jongman. 2019. Effects of sound change on the weighting of acoustic cues to the three-way laryngeal stop contrast in Korean: diachronic and dialectal comparisons. *Language and Speech* 62: 509-530.

[6] Flemming, E. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7-44.

Production and perception of ejective stops in Hul’q’umi’num’ and Q’anjob’al

Maida Percival[1], Pedro Mateo Pedro[2] & Sonya Bird[3]

[1]University of Toronto (Canada), [2]University of Toronto (Canada), [3]University of Victoria (Canada)
maida.percival@mail.utoronto.ca, pedro.mateo@utoronto.ca, sbird@uvic.ca

INTRODUCTION: This study presents production and perception experiments from two languages, Hul’q’umi’num’ (Coast Salish) and Q’anjob’al (Mayan), which have been impressionistically associated with typologically different strong and weak ejectives, respectively, as described in [1] and [2]. No other research has systematically investigated the perceptual cues to ejectives, and so a major contribution of this work is to examine how the acoustic dimensions which characterize ejectives in production are used by listeners in perception of the ejective – plain stop contrast. The findings also address whether there are differences across the languages in production or perception which may pattern along [1] and [2]’s typological classification of languages’ ejectives as strong and weak.

PRODUCTION: Acoustic analysis was done for each language separately to determine how ejectives are characterized as opposed to plain stops in each language and the extent to which their acoustics aligns them as strong or weak.

Methodology: 9 L1 speakers of Hul’q’umi’num’ (aged 65-87), and 25 of Q’anjob’al (aged 21-61) were recorded reading a word list of plain - ejective (near-)minimal pairs in their language covering all stop places of articulation across three word positions (word-initial, intervocalic, word-final). Annotations were made in Praat (2717 stop tokens for Hul’q’umi’num’ and 4754 for Q’anjob’al) and measurements taken based on [3]. Linear mixed effects models were done in R to determine for each acoustic measurement whether ejectives significantly differed from plain stops.

***Results**:* As summarized in Table 1, Hul’q’umi’num’ ejectives typically had long releases with a period of silence after loud bursts, characteristics of strong ejectives. Plain stops were aspirated and followed by vowels with more raised onset F0 and higher onset H1-H2 than ejectives’, but a similar amount of jitter suggesting breathy voice on them more so than the creaky voice following ejectives that is characteristic of weak ejectives. Q’anjob’al ejectives were equally likely to have or not have a period of silence following the burst (characteristics of strong vs. weak ejectives), and except word-finally were longer than plain stops, which were unaspirated. Ejective bursts were similar in intensity to plain bursts and following vowel onsets had lower H1-H2, greater jitter, and lower F0 for ejectives which suggests the presence of creaky voice, a characteristic of weak ejectives. Neither language’s ejectives fit perfectly into the strong – weak typology, but on average Hul’q’umi’num’ ejectives had more strong characteristics and Q’anjob’al more weak.

PERCEPTION: Forced choice identification tasks with language-specific stimuli were given to listeners of each language. Q’anjob’al listeners had an additional task, one of the Hul’q’umi’num’ stimulus sets, to more directly compare perception without differences in stimuli acoustics.

***Methodology**:* Participants were 25 L1 listeners of Q’anjob’al and 26 listeners of Hul’q’umi’num’, of which 7 were L1 – Hul’q’umi’num’ only has about 35 L1 speakers and so for this reason L2 speakers were included. The stimuli were minimal pairs manipulated along acoustic dimensions found to characterize ejectives in production and cross-spliced to make all combinations: for Hul’q’umi’num’ there were 4 bursts (2 burst types: baseline ejective and baseline plain and 2 intensities: 40 dB and 50 dB), 5 releases (0 ms/burst only, 50 ms silence, 50 aspiration, 120 ms silence, 120 ms aspiration), and 4 vowels (2 vowel types: baseline ejective and plain and 2 F0 patterns: onset raised and onset lowered by 0.286 barks). Q’anjob’al stimuli had the same manipulations except no aspirated releases. Logistic mixed-effects regression models were performed in R on the 6160 responses from Hul’q’umi’num’ listeners, 4008 responses from Q’anjob’al listeners to their own language stimuli and an additional 2000 Q’anjob’al responses to Hul’q’umi’num’ stimuli to determine whether listeners’ percent of ejective responses differed across levels of the dimensions.

***Results**:* Listeners from both languages were very similar in perception: they used as primary cue to the perception of ejectives the presence of silence after the burst. 75% of Hul’q’umi’num’ stimuli with silence were perceived as ejective by Hul’q’umi’num’ listeners, 95% of Q’anjob’al stimuli by Q’anjob’al listeners, and 94% of Hul’q’umi’num’ stimuli by Q’anjob’al listeners. In contrast, about 50% of stimuli with no silence or 0 ms of post-burst release duration were perceived as ejective for each. Properties of the stop burst and coarticulation in the following vowel were secondary cues in both languages, that listeners relied on more in the stimuli with 0 ms of post-burst release duration, the results for which are presented in Table 2. One difference between the languages is that Q’anjob’al listeners seemed more sensitive to baseline burst type and vowel type.

DISCUSSION: The results did not find complete correspondence between production and perception. Q’anjob’al listeners used silence in the release as a cue to ejectives slightly more despite having a lesser percent of releases with silence in production. Both languages used burst intensity to a similarly small extent in perception despite Hul’q’umi’num’ but not Q’anjob’al’s ejectives’ bursts being louder than plain bursts in production. Neither language used lowered vowel onset F0 as a cue to ejectives even though ejectives differed from plain stops in this in production.

The overall similarities in perception suggest that differences in ejective stop production may not relate straightforwardly to any typological differences across languages. One explanation which contextualizes the findings is sociolinguistic factors related to language context: Hul’q’umi’num’ participants were teachers and students with varying levels of fluency at a language school aimed at language revitalization. Strong ejective characteristics in production may reflect a shift in the stop system due to hyperarticulation in the context of language teaching [4] and less robust perceptual cue usage may reflect variation in production in this language context.

Table 1. Summary of acoustic results. > = “significantly greater than”; , = “not significantly greater than”; ej = ejective; pl = plain; #_ = word-initial; V_V = intervocalic; _# = word-final, asp = aspiration, sil = silence, w/ = with, w/o = without

Acoustic dimension		**Hul’q’umi’num’**	**Q’anjob’al**
Release type (% of tokens by stop type with post-burst release content)	Plain	60% w/ asp, 37% w/o	13% w/ asp, 69% w/o
	Ejective	62% w/ sil, 28% w/o	43% w/ sil, 45% w/o
Release duration (from onset of burst to onset of voicing or offset of release for _#)		ej > pl (#_) ej, pl (V_V, _#)	ej > pl (#_ and V_V) pl > ej (_#)
Burst intensity (maximum intensity in dB)		ej > pl	pl, ej
Following vowel onset phonation (normalized against midpoint)	H1-H2	pl > ej	pl > ej
	Jitter	ej, pl	ej > pl
Following vowel F0 perturbation		pl > ej	pl > ej

Table 2. Summary of % ejective responses for Hul’q’umi’num’ and Q’anjob’al listeners for each manipulated dimension in stimuli with 0 ms of post-burst release duration. > = “significantly greater % ejective responses than”; , = “not significantly greater % ejective responses than”; (%) is the difference in % ejective response between the first and second level (a higher % = more cue usage)

Dimension	**Own-language stimuli**		**Shared Hul’q’umi’num’ stimulus set**	
	Hul’q’umi’num’	**Q’anjob’al**	**Hul’q’umi’num’**	**Q’anjob’al**
Burst type	ej > pl (10%)	ej > pl (35%)	ej > pl (13%)	ej > pl (32%)
Burst intensity	loud > quiet (9%)	loud, quiet (4%)	loud > quiet (9%)	loud > quiet (10%)
Vowel type	ej > pl (19%)	ej > pl (21%)	ej > pl (15%)	ej > pl (33%)
Vowel F0	raised, lowered (1%)	raised > lowered (5%)	lowered, raised (3%)	lowered, raised (6%)

References

[1] Lindau, M. (1984). Phonetic differences in glottalic consonants. *Journal of Phonetics, 12*, 147–155.

[2] Kingston, J. (1985). The phonetics and phonology of the timing of oral and glottal events (Ph.D. dissertation). University of California, Berkeley.

[3] Wright, R., Hargus, S., & Davis, K. (2002). On the categorization of ejectives: Data from Witsuwit’en. *Journal of the International Phonetic Association, 32*, 43–77.

[4] Leonard, J., Bird, S. & Gerdts, D. (2017). Exploring L2 pronunciation features in Hul’q’umi’num’ and SENĆOŦEN. Presentation given at NowPhon 2. UBC, May 19.

Phonetic Targets in the Clear Speech Vowel Productions of Native Chinese Learners of Korean

Shiyu Zhang[1] & Jeffrey J. Holliday[2]

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

Most previous studies on L1 Chinese learners of Korean attribute errors in L2 Korean vowel production to differences in the phonological vowel inventories between Chinese and Korean. It is explained that errors occur because when L1 Chinese learners speak Korean, they simply substitute the Chinese vowel phonemes most similar to Korean vowels. However, a difference in phonological inventory does not always cause a problem in L2 vowel learning [1]. One question raised by these previous studies is what L1 Chinese learners' Korean vowel targets actually are: are their phonetic targets themselves non-native like, or merely implemented in a way such that the target is not fully realized (e.g. undershot)?

According to [2], it has been shown that the phonetic targets of vowels are hyperarticulated: that is, the phonetic target of a vowel in the mind of the speaker is more peripheral in the vowel space than what is typically produced. In the current study, we aim to characterize the phonetic targets by eliciting citation speech and clear speech, which is a form of hyper-articulation. The main question we aim to address is the following: When L1 Chinese speakers are asked to produce Korean vowels more clearly, do they get closer to native-like targets? We additionally looked at these speakers' native Chinese vowel productions, to confirm whether the clear speech enhancement strategy used in the L2 is the same as what they use in their L1.

The participants were 20 L1 speakers of Chinese and 20 L1 speakers of Korean. Speakers first read a list of 21 disyllabic Korean words in a carrier sentence, and then were asked to repeat the list but speaking as clearly as possible, as if to a child learning the word. Each word contained a target vowel in the first syllable. The Chinese speakers did the task again with a list of 15 Chinese words, too. The recordings were labeled in Praat and the mean F1 and F2 were measured over a 20 ms window in the steady state interval of each target vowel. The format values were normalized using the Lobanov method. Statistical significance between vowels was tested using separate linear mixed effects models of normalized F1 and F2.

The vowel spaces for both groups' Korean vowel productions are shown in Figure 1. The clear speech distributions are very similar for the vowels /ɑ/, /ɛ/, /i/, /ɨ/, and /ʌ/, with the only significant differences between groups being that L1 Chinese speakers' /i/ and /ʌ/ are slightly higher (i.e. lower F1; /i/, *p* = .015; /ʌ/, *p* = .037) than that of L1 Korean speakers. In the high back area of the vowel space, compared to native Korean speakers, L1 Chinese /o/ is lower and fronter (i.e. higher F1, *p* = .011; higher F2, *p* = .027), and /u/ is higher and backer (i.e. lower F1, *p* = .013; lower F2, *p* = .003). In terms of particular contrasts, it was found that the L1 Chinese speakers made the /o/-/u/ contrast using F1 alone, whereas L1 Korean speakers used both F1 and F2.

When the /o/ and /u/ targets are compared across speaking styles but within each L1 group, it was found that L1 Chinese speakers produce their clear targets by lowering and backing /o/, and backing /u/. In other words, for L1 Chinese speakers, both /o/ and /u/ should be in the back of the vowel space, and /o/ is lower than /u/. But for L1 Korean speakers, /u/ is already fronter than /o/ in citation speech, and in their clear speech both /o/ and /u/ are lowered. This result was unexpected, because the lowering of both /o/ and /u/ does not result in an overall expanded vowel space. But it demonstrates a pattern in which speakers of both languages use one dimension to make the /o/-/u/ contrast (F1 for L1 Chinese, primarily F2 for L1 Korean), but then use the other dimension to produce it more clearly (F2 for L1 Chinese, F1 for L1 Korean).

So why do L1 Chinese speakers move their /u/ target back in clear speech, when L1 Korean speakers' /u/ target is actually fronter? Comparing the clear speech phonetic targets of Chinese speakers' Korean /u/ and Chinese /u/, it was found that Chinese /u/ was both lower (i.e. higher F1; p < .001) and backer (i.e. lower F2; p < .001), suggesting that the backing of Korean /u/ in the clear speech of L1 Chinese speakers could be the result of their phonetic target being aligned with their back Chinese /u/.

In conclusion, although there are differences in the Chinese and Korean vowel inventories, these results suggest that L1 Chinese speakers are not merely substituting a Chinese /u/ for a Korean /u/: they do produce a difference in /u/ between the two languages. But rather, their phonetic target for Korean /u/ remains tied to their Chinese /u/, leading them to move /u/ further back in the vowel space in clear speech, despite the fact that Korean /u/ is actually fronter. These results further demonstrate that both native and non-native speakers can make enhancement adjustments by recruiting non-primary acoustic cues. L1 Korean speakers signal the /o/-/u/ contrast primarily using F2, but increased their reliance on F1 to produce it more clearly. L1 Chinese speakers, on the other hand, rely primarily on F1 to signal the /o/-u/, but used both F1 and F2 to enhance the contrast in clear speech.

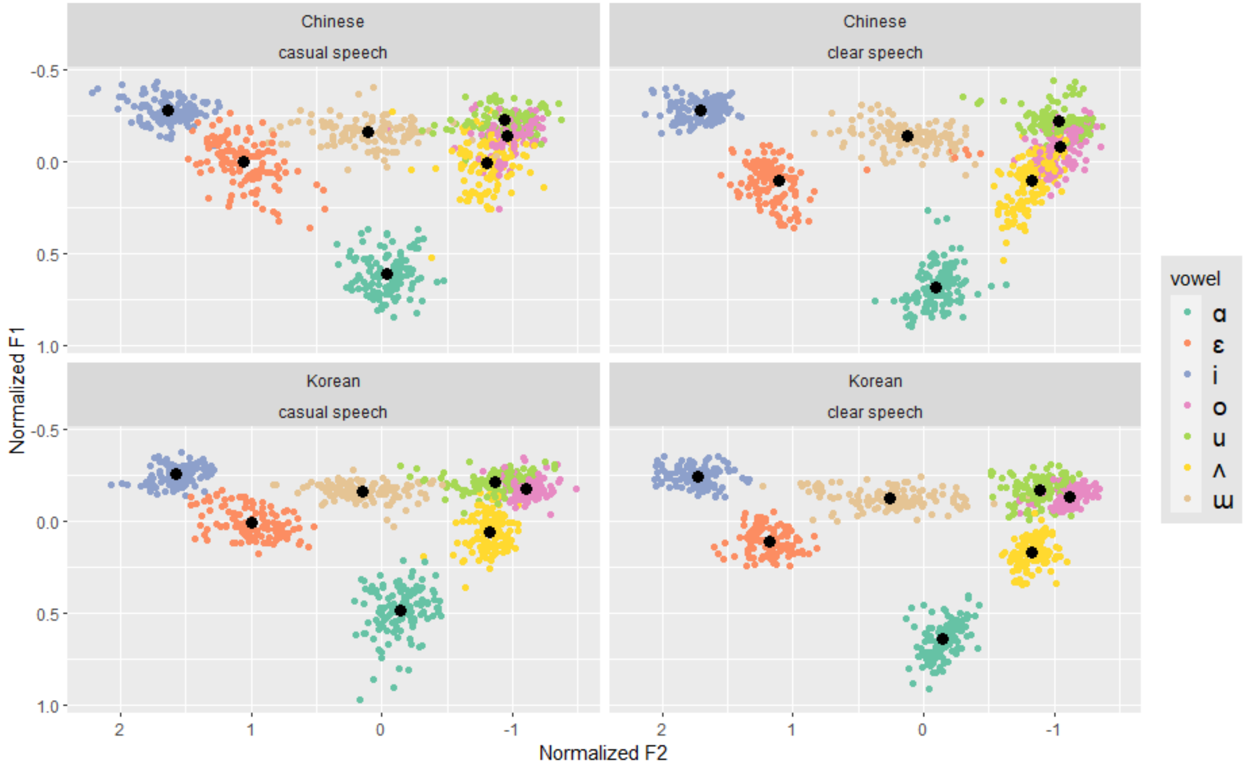


Figure 1. L1 Chinese and L1 Korean speakers' vowel spaces

References

[1] Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.) *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). Cambridge University Press.

[2] Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69(3)*, 505–528.

Influence of research tasks and linguistic factors on phonetic convergence in language alternation

Ernesto R. Gutiérrez Topete[1]

[1]*University of California, Berkeley (USA)*
ernesto.gutierrez@berkeley.edu

Recent studies have reported on the dynamic nature of bilingual speech (i.e., two languages produced in a single discourse event) compared to monolingual speech (i.e., one language produced in a single discourse event) by multilingual speakers. For example, in monolingual speech, multilingual speakers “maintain language-specific phonetic categories”; meanwhile, in bilingual speech, the speakers displayed “phonetic convergence” [1, pg. 1280-1]. Seemingly, the active usage of multiple languages increases phonetic convergence in the production of multilingual speakers. Other studies on the production of voice onset time (VOT) during language alternation (i.e., language switching and code switching) corroborate the convergence effect observed in bilingual speech [e.g., 2, 3, 4, 5]. However, some inconsistencies have been pointed out [6, 7]. Namely, while some studies find bidirectional convergence, others only find effects in one language. And among those studies with a unidirectional effect, some researchers claim the effect occurs only in the speakers’ *L2* (as opposed to the L1) while others claim the *dominant language* is more susceptible to convergence. The wide array of research tasks used to study this phenomenon makes it all the more difficult to pinpoint the cause of the aforementioned inconsistent directionality of the convergence effect. For example, some studies relied on word list reading tasks [6] or phrase/passage reading tasks [2, 3], and others used speech spontaneously produced during sociolinguistic interviews [4], inter-subject group conversations [4, 5], or puzzle tasks [5].

In order to shine light on the possible effect of task type on the directionality of convergence effects in bilingualism research, the present study analyzes acoustic productions across four of the most popular research tasks (i.e., word list reading, passage reading, puzzle—spot the difference—and casual interview tasks) from a single group of Spanish-English bilingual speakers to obtain VOT measurements for word-initial voiceless stops /p t k/. A total of 60 Spanish-English bilingual subjects participated in the four tasks, which yielded nearly 100 hours of recorded bilingual speech. Data collection took place in a sound booth at the Berkeley PhonLab. The audio for the word list and passage reading tasks were annotated by hand. The transcriptions for the puzzle and interview tasks were automated with OpenAI’s Whisper automatic speech recognition model. All data were then processed with the Montreal Forced Aligner [8], and VOT measurements for voiceless stop-initial words were obtained in both languages using AutoVOT [9].

A mixed-effects linear regression model in R was performed on the VOT values of a majority of the corpus data. The independent variables under examination were task type (word list, passage, puzzle, interview), place of articulation (POA; bilabial, coronal, velar), and across-boundary context in English (/sC/, /nC/, other), to analyze the potential of resyllabification (more precisely, switch sites with /s/-final Spanish words followed by voiceless stop-initial English words, compared to /n/-final Spanish words or words ending in other sounds). Lexical item and subject were included as random intercepts. Study results show an influence of task on English data, with passage reading experiencing the highest level of convergence (i.e., lower VOT values), followed by the interview, puzzle, and word list (see figure 1). Spanish data showed no difference across tasks. POA results revealed that in English /p/ VOT productions are lower than /t k/ VOT, whereas in Spanish /p t/ VOT values are lower than /k/ VOT (see Figure 2). Finally, there is no evidence of resyllabification, given that we do not see shorter VOT values in the /sC/ linguistic context (compared to other contexts), which is linked to shortened VOT productions in American English when found as an onset cluster.

On their own, these results suggest that convergence is linked to attention to speech, where tasks that draw more attention to speech are less prone to convergence, meanwhile those that require less attention to individual sounds present higher levels of convergence. These results suggest that methodological choice has an influence on the acoustics of (code-switched) speech, potentially resulting in methodological artifacts in studies that do not take this into consideration. Moreover, VOT productions per POA vary in English and Spanish (English: /p/ < /t k/; Spanish /p t/ < /k/), a pattern that can be explained by the distinctions of coronal stops in each language; Spanish has dental stops, and English has alveolar stops. In other words, in Spanish /p t/ are produced closer together than their English counterparts, resulting in more similar VOT productions between these two sounds in Spanish speech; these findings underlie the importance of analyzing articulatory differences when comparing language pairs in language alternation research. Finally, the fact that /sC/ linguistic contexts in switch sites were not correlated with lower VOT values in English stops indicates that speakers do not engage in a resyllabification process across languages during code-switched speech, suggesting that this form of phonotactics (ie, resyllabification) does not transfer from one language to another, even during bilingual speech.

All in all, this study provides (1) a comparative analysis of research methodologies that are commonly used in code-switching studies to uncover the inadvertent task effects in production studies and (2) a better understanding of the language processing mechanisms that are engaged during bilingual speech to better inform our methodological choices.

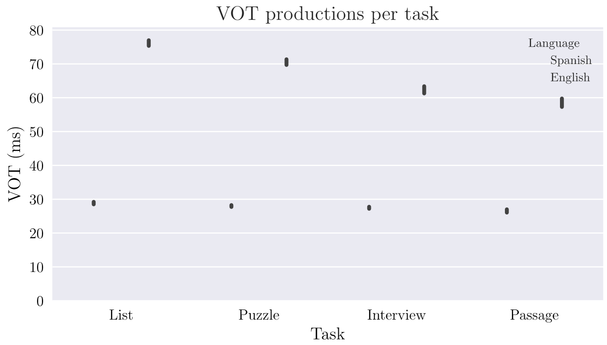


Fig.1 VOT productions per methodological task in English and Spanish

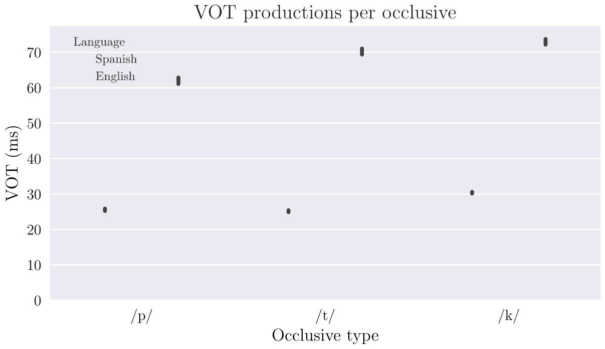


Fig.2 VOT productions per place of articulation in English and Spanish

References

[1] Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *The Journal of the Acoustical Society of America*, *149* (2), 1271–1284.

[2] Toribio, A. J., Bullock, B. E., Botero, C. G., & Davis, K. A. (2005). Perseverative phonetic effects in bilingual code-switching. In R. Gess & E. Rubin (Eds.), *Theoretical and experimental approaches to romance linguistics* (pp. 291–306). John Benjamins Publishing Company.

[3] Bullock, B. E., Toribio, A. J., González, V., & Dalola, A. (2006). Language dominance and performance outcomes in bilingual pronunciation. In M. Grantham O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th generative approaches to second language acquisition conference* (pp. 9–16).

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

[5] Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish–English spontaneous code-switching. *Journal of Phonetics*, *52*, 121–137.

[6] Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in vot production. *Journal of Phonetics*, *41* (6), 407–420.

[7] Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language*, *89*, 110–137.

[8] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. *Interspeech*, *2017*, 498–502.

[9] Keshet, J., Sonderegger, M., & Knowles, T. (2014). AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [Computer program]. [Version 0.94, retrieved June 2020 from https://github.com/mlml/autovot/].

Sound change reverse via short-term phonetic accommodation: evidence from an in-progress tonal sound change toward the prestigious accent

Liu Huangmei[1] & Zhang Chunmei[2]
University of Shanghai for Science and Technology (China)
Lauraliu4321@163.com, zhangchunmeiw@163.com

Phonetic accommodation is defined as a phenomenon that speakers spontaneously adjusting their speech when talking to another talker, which could lead a result of convergence (being more similar to the counterpart) or divergence (being more different from the counterpart). When hearing a new sound, no matter the hearer choose to spontaneously imitate the new sound or stay different with it, he or she will always play a role in the progress of language development. The decision of the hearer either could become a new trend of sound change, or could reinforce the prestige of the original one. Therefore, phonetic accommodation of the speakers in contact is one of the powerful forces that drive our language into sound changes. Phonetic accommodation is adopted to investigate a huge amount of possible sound changes within one language (Babel et al., 2013; Kwon, 2021), between two languages (Tobin, Nam, & Fowler, 2017), or even in conditions between human and virtual talkers (Gijssels et al., 2016; Wynn & Borrie, 2020). Empirical findings suggested that phonetic accommodation could be affect by various of factors, such as age, gender, social statues, and other social factors (Labov, 2001), linguistic experience (Lee, Politzer-Ahles & Jongman, 2013), stage of a change in progress (Lin et al., 2021), motor activation condition, cognitive sensitivity, and other cognitive factors.

The present study endeavors to find whether shadowers could reverse back to their dialectal accent from an in-progress sound change toward the standard sound. On one hand, despite the fruitful achievement of the previous finding, most studies on phonetic accommodation used an imitation target which might not be on either end of in-progress sound changes. However the shadowers knowledge of the variant (the imitation target) may affected the result of phonetic imitation task. Therefore, our study tries to contribute in exploring imitation effect on sound change phenomenons that is occurring in progress now. On the other hand, though sound change which had completed rarely reverse its development, those in-progress ones may show uncertainty of sound change directions (reversible or irreversible). For example, Babel et al. (2013) find a reversing effect of phonetic accommodation of New Zealand English speakers shadowing Australian vowels. Yao and Chang (2016) also found empirical evidence for Shanghainese speakers could reverse some sound change towards Mandarin vowels. Lin et al. (2021) evidenced that a tone merger-in-progress in Hong Kong Cantonese could be reversed via short-term imitation tasks. Moreover, the above researches only investigated imitation effect on mergers. Mergers usually are sound changes between two or more categorical phonemes / tonemes. The present study here tries to extend the research targets to in-progress sound change towards the standard accent, which is a non-categorical change.

Therefore, the present studies was designed to include both production and perception activities of phonetic accommodation. The in-progress tonal sound change is T2 of Shanghai accented Mandarin1. T2 in Shanghai accented Mandarin is under a sound changing progress from a low flat tone to a high rising tone2 (Gu, 2007). The progress gives a unique and thrilling opportunity to study non-categorical sound change. Therefore, 60 Shanghainese (younger group: 30 young adult, aged from 20-30; older group: 30 old adult, aged from 45-55 with no hearing impairment) participant in the shadowing experiments. Only monosyllabic words in Shanghai accented Mandarin is used as models in the experiment. Stimuli for pre and post shadowing perception tests are the same with those in the shadowing task. In the experiment, the perception task follows the production task both in of pre-shadowing phase and post-shadowing phase with no break in between. The main procedure includes "a baseline production, a baseline perception, shadow task 1, shadow task 2, a post shadowing production, a post shadowing perception", and "a familiarity & attitude test to sound change" in separate in temporal order. Both production and perception results of the shadowing task of participants are recorded and analyzed as phonetic accommodation results. The familiarity and attitude tests are two separate tests performed consecutively. The familiarity test asks the participant to make lexical decision on 20 monosyllabic Mandarin stimuli. The accuracy and latency results are analyzed as the familiarity data. The attitude test asks participants to score pleasantness (1: the least pleasant; 5: the most pleasant) on the same monosyllabic stimuli in the familiarity tasks, with the scoring result and latency analyzed as attitude data. The difference between baseline and model tone production of participants (Baseline difference) is calculated as the main factor. A between group factor (age) and two individual factors (familiarity and attitude) are added to see individual effect on sound change.

Results show interesting findings. Both groups show tendency of reversing, with shadowers having larger baseline difference from the model showing a more noticeable imitation. This is consistent with other empirical evidence that bigger difference in baseline promote imitation. Familiarity shows a strong interactive effect with Baseline difference in both group, while attitude only interact with Baseline different within the younger group. Our finding suggested two important evidences: i) talkers are sensitive to non-categorical difference in phonetic accommodation activities; ii) a reversing development for an in-progress non-categorical tone change toward a prestigious accent is possible by phonetic accommodation, which still undergo impact from talkers' sound change statues, age, linguistic familiarity, and sound change attitude. Still future studies are needed to illuminate the role of present factors on other tonal sound changes, as the ample materials of non-categorial tonal sound change in Mandarin present a unique and thrilling opportunity of study like ours.

Keywords: Mandarin tone; sound change; phonetic imitation; phonetic accommodation

References

[1] Babel, M. , McAuliffe, M. , & Haber, G . (2013). Can mergers-in-progress be unmerged in speech accommodation?. *Frontiers in Psychology*, *4*(SEP), 1– 14.

[2] Kwon , H . (2021). A non-contrastive cue in spontaneous imitation: Comparing mono- and bilingual imitators. *Journal of Phonetics*, *88*.

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

[4] Gijssels, T ., Staum Casasanto, L ., Jasmin, K . , Hagoort, P ., & Casasanto, D . (2016). Speech Accommodation Without Priming: The Case of Pitch. *Discourse Processes*, *53*, 233–251.

[5] Wynn, C . J. , & Borrie, S . A . (2020). Methodology matters: The impact of research design on conversational entrainment outcomes . *Journal of Speech, Language, and Hearing Research*, *63*, 1352– 1360 .

[6] Labov, W . (2001). *Principles of linguistic change*, *Volume 2: Social factors*. Blackwell.

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

[8] Lin, Y . , Yao, Y . , & Luo, J. (2021). Phonetic accommodation of tone: reversing a tone merger-in- progress via imitation. *Journal of Phonetics*, *87*(5), 101060.

[9] Yao, Y . , & Chang, C. B. (2016). On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese. *Language*, *92*(2), 433–467.

[10] Gu, Q. (2007). *The influence of language contact on the phonetic evolution of the dialects in Shanghai urban districts*. Doctoral Dissertation, Shanghai Normal University, Shanghai.

[1] Shanghai accented Mandarin and Northern Mandarin share the same phonological system with four tones but are different in acoustical realization [T1: Yinping, a high flat tone in acoustics; T2: Yangping, a high rising tone (*a lowflat tone*); T3: Shang sheng, a low dipping tone (*a lowfalling tone*); T4: Qu sheng, a high falling tone (the ones with different acoustical realization in Shanghai accented Mandarin is display in blanket in Italic)].

[2] Though there is a noticeable tone shape difference, the tonal system is stable, therefore it is more like a non-categorical change.

An articulatory study of word-level prominence in two Mandarin dialects

Jing Huang, Feng-fan Hsieh, Yueh-chin Chang
National Tsing Hua University
xiaokuidaren94@163.com, ffhsieh@mx.nthu.edu.tw, ycchang@mx.nthu.edu.tw

Introduction: The existence of word-level prominence/lexical stress in Mandarin Chinese is controversial. In this study, we investigated whether word-level prominence is attested in two relatively understudied dialects of Mandarin, Southwestern Mandarin (SWM) and Taiwanese Mandarin (TWM). Unlike Beijing Mandarin, SWM and TWM do not have "unstressed" (neutral tone) syllables in lexical words, meaning that no apparent strong-weak patterns have been reported for them, at least impressionistically. Nevertheless, little is known if there is any articulatory difference between the syllables in a word. Given that stressed syllables may involve longer, larger, and faster gestures than their counterparts in unstressed positions (e.g., Katsika and Tsai 2021), we set out to assess the supralaryngeal kinematic correlates of word-level prominence/lexical stress in these two Mandarin dialects, using electromagnetic articulography (EMA).

Method: We analyzed and reported data from four SWM speakers (1 female) and two TWM speakers (1 male) in their twenties. The stimuli are personal names with two or three identical syllables (note that these are *not* reduplicated forms): {papa, pepe, tutu, titi} in SWM (all with the mid tone) and {pi, pai, tai, two/papapa, tatata, kakaka} in TWM (all with the high-level tone). Seven repetitions for each token were collected for SWM and ten repetitions for TWM. For SWM, the target words were embedded in the carrier phrase, "p^he A, *pu* p^he B", meaning "(Please) pat A! not B." The tokens in the unfocused position B were analyzed. For TWM, the target words were embedded in X in the carrier phrase "wǒ xiǎng X bèi Y", meaning "I think X memorizes Y." The target words were produced in a position with the *least* higher-level prominence (i.e., phrasal/sentential stress) and without the well-established effect of domain-initial strengthening (Keating et al. 2003) to highlight the word-level prominence.

Results: In disyllables, the initial syllables are significantly longer in duration (p<.05) in SWM, while the final syllables are significantly longer (p<.05) in TWM. See (1). Note that the raw data are presented here for ease of interpretation only.

(1) SWM-Initial σ	SWM-Final σ	TWM-Initial σ	TWM-Final σ
209 ms (SD: 56)	172 ms (SD: 55)	135 ms (SD: 22)	**153 ms** (SD: 25)

Regarding gesture duration, the onset gestures (i.e., Lip Aperture and Tongue Tip along the vertical dimension) are significantly longer (p<.05) in initial position than in final position in TWM. In contrast, no significant difference is found for SWM. See (2).

(2) SWM-Initial σ	SWM-Final σ	TWM-Initial σ	TWM-Final σ
55 mm (SD: 29)	51 mm (SD: 27)	**78 mm** (SD: 39)	42 mm (SD: 16)

Regarding the peak velocities of the onset gestures, our results reveal that there is no significant difference between the initial and final syllables in SWM and TWM (3).

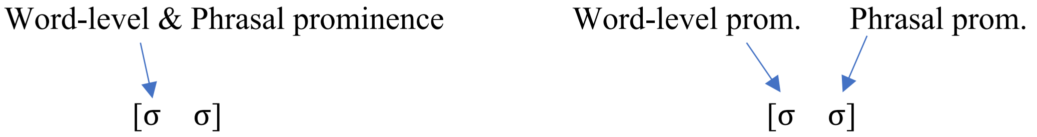
(3) SWM-Initial σ	SWM-Final σ	TWM-Initial σ	TWM-Final σ
13 cm/s (SD: 7)	12 cm/s (SD: 8)	15 cm/s (SD: 7)	15 cm/s (SD: 6)

The acoustic and articulatory results of the trisyllables in TWM are given in (4). It appears that there is a "mismatch" in acoustics and articulation: the non-initial syllables are significantly longer in duration (*p*<.05); on the other hand, the onset gestural duration is the longest in word-initial position (*p*<.05). Finally, other acoustic cues such as F_0 and intensity are not significantly different across the board (not shown here).

(4) TWM	Initial σ	Medial σ	Final σ
Syllable duration	134 ms (SD: 28)	**150 ms (SD: 31)**	**159 ms (17ms)**
Gestural duration	**83 mm** (SD: 41)	37 mm (SD: 17)	37 mm (SD: 18)
Peak Velocity	13 cm/s (SD: 7)	17 cm/s (SD: 7)	15 cm/s (SD: 8)

Discussion: The present results reveal a novel cross-dialectal difference: both word- and phrase-level prominences are cued by phonetic duration in SWM because the initial syllable is longer, although no kinematic differences (gestural duration/peak velocity) were found. On the other hand, TWM shows mixed results; the non-final syllables are longer in acoustics, but articulatorily, the onset gestures are significantly longer in word-initial positions. This cross-dialectal difference might be attributed to the possibility that in TWM, the word-level prominence does not coincide with the phrasal/edge prominence. In other words, the longer gesture duration in word-initial positions can be regarded as an exponent of word-level prominence/lexical stress and phrasal prominence (longer duration) falls on the final syllables in TWM (see also Duanmu 2007 for a similar view). The cross-dialectal distinction is schematized in (5).

(5) a. Southwestern Mandarin b. Taiwanese Mandarin



Word-level & Phrasal prominence

[σ σ]

Word-level prom. Phrasal prom.

[σ σ]

From a broader perspective, both the SWM and TWM results further suggest that the gestural duration in articulation might be orthogonal to the syllable duration in acoustic measurements.

Conclusion: In this study, we have shown that there are different sources of prosodic prominence in Mandarin: specifically, word-level prominence in TWM is more reliably detected in kinematics but not in acoustics. In sum, the present results indicate that the articulatory properties also contribute to the realization of prosodic prominence, thus casting doubt on the widely accepted claim that Mandarin has no lexical stress.

References

[1] Katsika, A. and K. Tsai. 2021. The supralaryngeal articulation of stress and accent in Greek. *JPhon* 88.

[2] Keating, P., Cho, T. H., Fougeron, C., & Hsu, C. S. 2003. Domain-initial strengthening in four languages. In Local J. et al. (Eds.*), Laboratory Phonology VI: Phonetic Interpretation*, pp.145-163. CUP.

[3] Duanmu, S. 2007. *The Phonology of Standard Chinese*. OUP.

Cross-language Perception of Parallel Encoded Emotional and Linguistic Prosody by Chinese Learners of English

Qianyutong Zhang[1], Shanpeng Li[2] & Lei Zhu[1]

[1]Shanghai International Studies University (China), [2]Nanjing University of Science and Technology (China)
zqyt1227@163.com, shanpeng@njust.edu.cn, zhulei@shisu.edu.cn

Background: Prosody is used to convey not only emotional (e.g., angry or happy emotion) but also linguistic information (e.g., question or statement) [1, 2, 3], both of which are usually encoded in parallel in the same utterance at the same time. Previous studies have demonstrated an interaction effect in the perception of parallel-encoded prosody [3, 4, 5], whereby listeners may have difficulty identifying its emotional and linguistic function. To be specific, linguistic prosody such as sentence-type intonation will affect the recognition of emotion types [6], and conversely, emotional prosody also interferes with the perception of statement/question contrasts [3]. However, these studies are scarce and are mostly focused on native speakers of Indo-European languages like English and German. Very few have considered tonal-language speakers, let alone such speakers who are learning a non-tonal language as L2.

The cross-language study between Mandarin Chinese and English is of great significance since tonal languages have differences from non-tonal languages in using prosodic features such as F0. Since our brain's processing of F0 is closely related to language experience [7], Chinese English learners' processing of their L2 English prosody may have some differences with that of Mandarin prosody. Therefore, the present study aims to explore the interaction between emotional and linguistic function during the perception of Mandarin Chinese and English prosody by Chinese English learners, and to examine how L2 proficiency moderates this interaction.

Method: Forty-four Chinese native speakers participated in this experiment. They all learned English as a second language and were divided into high-level and low-level groups according to their CET-4 and CET-6 scores. The materials consisted of 130 syntactically similar and semantically neutral sentences with the parallel sentences in both English and Chinese versions (e.g., "Mark is watching TV. /?", "小马正在看电视。/?"), and 130 filler sentences. All target sentences were read by a female Mandarin speaker and a female English native speaker, and were recorded in four conditions: emotionally neutral statements, emotionally angry statements, emotionally neutral questions, and emotionally angry questions. In the emotion-identification task, participants were asked to ignore the sentence-type intonation and to identify emotions by pressing the keyboard ("1" for "angry", "2" for "neutral" and "3" for "others"). And in the intonation-identification task, participants should ignore the emotions and recognize the intonations ("1" for "question", "2" for "statement" and "3" for "others").

Results: Participants' average identification accuracy under two tasks and four conditions were analyzed by the linear mixed-effect model in R [8]. The results showed that emotional prosody and linguistic prosody have interactions in the prosody identification process, but in different ways under different prosodic conditions. 1) On the one hand, linguistic intonation affects the perception of emotional prosody. Specifically, in the emotion identification of both English and Mandarin sentences, question intonation reduces the accuracy of neutral emotion identification. The only difference between Mandarin and English lies in the perception of angry emotion. In English, statement/question does not affect the perception of anger, while in Mandarin, the accuracy of angry statements is significantly lower than that of angry questions. 2) On the other hand, emotional prosody also affects linguistic prosody perception, and the results are consistent in Mandarin and English. Angry emotion interferes with the perception of statement intonation, while angry/neutral emotion does not affect the perception of question intonation. 3) In addition, English proficiency has no significant influence on Chinese English learners' perception of English prosody.

Conclusion: In sum, our results proves an interaction effect between different functions of prosody, and this interaction shows a generally similar pattern between Mandarin and English. In both languages, question intonation reduces the accuracy of neutral emotion identification, and angry emotion impedes the perception of statement intonation, indicating that the pitch variability associated with emotion realization interferes with the pitch direction related to linguistic prosody. On the other hand, the perception of questions is not influenced by emotions in both languages, indicating a stable perception of questions. The only difference between Mandarin and English lies in the perception of angry emotion under different intonations, indicating a different mechanism in the perception of emotional prosody in tonal and non-tonal languages. Overall, the present study describes the interaction effect between emotional and linguistic prosody in Chinese English learners' L1 and L2 prosody processing, contributing to enriching the cross-language prosody study.

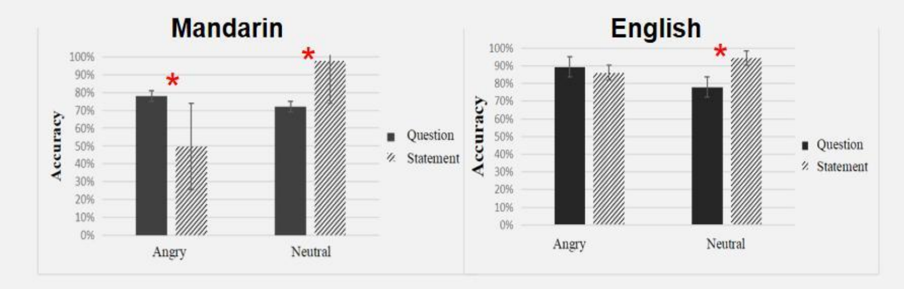


Fig.1 The mean identification accuracy of angry and neutral emotions under different intonations in Mandarin and English utterances.

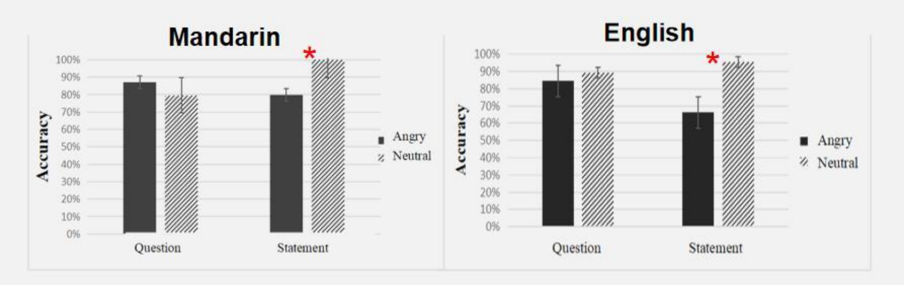


Fig.2 The mean identification accuracy of question and statement intonations under different emotions in Mandarin and English utterances.

References

[1] Eckstein, K., & Friederici, A. D. (2006). It's Early: Event-related Potential Evidence for Initial Interaction of Syntax and Prosody in Speech Comprehension. *Journal of Cognitive Neuroscience, 18*(10), 1696–1711

[2] Pihan, H., Tabert, M., Assuras, S., & Borod, J. (2008). Unattended emotional intonations modulate linguistic prosody processing. *Brain and Language, 105*(2), 141-147.

[3] Paulmann, S., Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia, 50*(7), 1609-1620.

[4] Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *The Journal of the Acoustical Society of America, 109*(4), 1668-1680.

[5] Zora, H., Rudner, M., & Montell Magnusson, A. K. (2019). Concurrent affective and linguistic prosody with the same emotional valence elicits a late positive ERP response. *European Journal of Neuroscience, 51*(11), 2236-2249.

[6] Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America, 76*(5), 1346-1356.

[7] Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *NeuroReport, 9*(9), 2115-2119.

[8] R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.2) [Computer software]. Retrieved from https://www.R-project.org/

Structured Suprasegmental Variation: Marking Prominence in Australian Languages

Sarah Babinski[1]

[1]*University of Zürich (Switzerland)*
sarah.babinski@uzh.ch

Recent work has investigated patterns of phonetic variation with respect to phenomena that are not phonologically contrastive in a language, finding structure even in areas that were once thought below the control of the speaker [1]. Speakers have been observed to converge on patterns of phonetic variation that are consistent within languages but variable cross-linguistically for the same phonological phenomenon. One area where principled phonetic variation is well-established and expected is the domain of prosody and stress, which may have one or more phonetic cues, commonly including features such as duration, pitch, and intensity, although the exact correlates of something like lexical stress is highly language specific and may vary substantially [3].

The present study considers variation in acoustic features, particularly in reference to word-level stress marking, in sixteen Australian languages. The results of this study support the claim that the phonetic markers of a prosodic phenomenon such as lexical stress varies in structured ways that indicate these markers vary and change in a principled way, and thus can be studied similarly to linguistic studies of segmental change. Even though the acoustic correlates to stress– phonetic factors such as duration, intensity, f0– all serve to mark the same type of phonological event, the phonetic variation in this marking is still structured in the way that a phonologized factor such as phonemic stop voicing might be. This talk explores the structure in this variation, finding evidence for historical links between related languages, as well as sociolinguistic variation within languages that further support the claim that speakers are in control of suprasegmental cues just as they are for segmental phenomena.

While the position of word-level prominence marking in most Australian languages is phonologically stable at the beginning of each word, the phonetic factors used to mark this prominence varies widely [4,5]. Past studies of the largest family on the continent, Pama-Nyungan, have stated that the primary correlate of stress in these languages is usually pitch [6,7,8]; while this study does find pitch to be a common correlate of stress, duration is about equally as common, and in some languages neither of these correlates highly with word-initial prominence. Within languages, however, speakers usually are consistent in which phonetic cues they use to mark word prominence, suggesting language-internal consensus on these cues. The question that remains is how this variation arises over time. What I propose in this talk is that phonetic correlates of prosodic phenomena are both stable and variable – there is often a primary correlate or set of correlates that speakers are very consistent on, in addition to ‘secondary’ correlates that are more variable across speakers and may serve as the catalysts of change in many situations.

The study presented here is an investigation into structured variation of the acoustic correlates of stress and prosody in sixteen Indigenous languages of Australia that all have consistent initial stress placement, with a focus on the source(s) of variation in these factors cross-linguistically. The data used in this dissertation are narrative speech recordings sourced from language archives, collected in varying field settings. Natural speech data, along with being the only available data sources for some languages that are no longer spoken, also have the advantage of showing more variation than careful lab speech and revealing the multidimensionality and variation inherent in language and in prosody more specifically. Original audio and transcripts were time-aligned using the Montreal Forced Aligner with manual correction, and subsequent phonetic measurements were extracted with Praat and analyzed in R.

The acoustic correlates of stress show significant cross-linguistic variation, both in the presence or absence of a particular cue to stress and the size of these effects, despite the phonological uniformity present in these languages with respect to initial stress placement. The phonological uniformity of stress assignment allows for a more controlled comparison of the acoustic correlates of stress across these languages, since the placement of stress marking remains constant. Acoustic correlates investigated are vowel duration, pre-tonic and post-tonic consonant duration, intensity, f0 (maximum and range), and vowel peripherality. These cues are identified using a series of mixed effects linear regression models, and the sources of variation are identified using Analysis of Molecular Variance [9] (AMOVA).

Almost all the languages in this study have multiple acoustic factors that correlate with lexical stress. Likewise, in all languages that have more than one speaker, at least one of these factors showed interspeaker variation. These results show that stress is often marked by multiple cues, and not all these cues are only doing the work of marking the stress contrast. Some factors that show interspeaker variation are additionally conveying some information about the speaker, be it age, gender, or social status. This sociolinguistic variation could potentially be an example of change in progress, where one group has taken up the change and another has not yet, or it could be stable social variation within a community, in line with recent work on socio-prosodic variation in other languages [10].

Speakers are evidently sensitive to the patterns of prosodic marking in their language, and they learn this phonetic variation in a consistent way. Furthermore, the systematicity of this variation suggests that these patterns should change over time systematically as well. The results of this study indicate that the phonetic correlates of stress are shared among related languages in some cases, while in other cases cross-linguistic variation is substantial. I argue that the changes that give rise to this variation come from either regular sound change or contact situations, similarly to many types of segmental change. Changes in the phonetics of prosody can also occur within subpopulations of a language, creating variation along sociolinguistic lines. Such observations speak to the nature of the language faculty and the cognitive organization of language, even below the abstract level of the phoneme, and to our theories of phonetic change and the phonetic precursors to phonological change.

References

[1] S. M. Kakadelis, “Phonetic Properties of Oral Stops in Three Languages with No Voicing Distinction,” PhD, CUNY, 2018.

[3] M. Gordon and T. Roettger, “Acoustic correlates of word stress: A cross-linguistic survey,” Linguist. Vanguard, vol. 3, no. 1, 2017.

[4] J. Fletcher and A. R. Butcher, “Sound patterns of Australian languages,” in The Languages and Linguistics of Australia: A Comprehensive Guide, R. Nordlinger and H. Koch, Eds. De Gruyter, 2014.

[5] R. Goedemans, “2. An overview of word stress in Australian Aboriginal languages,” in A Survey of Word Accentual Patterns in the Languages of the World, H. van der Hulst, R. Goedemans, and E. van Zanten, Eds. Berlin, New York: DE GRUYTER MOUTON, 2010. doi: 10.1515/9783110198966.1.55.

[6] C. Simard, “The prosodic contours of Jaminjung, a Northern Australian language,” PhD Dissertation, University of Manchester, 2010.

[7] J. Bishop, “Aspects of intonation and prosody in Bininj Gun-wok: an autosegmental-metrical analysis,” PhD Dissertation, University of Melbourne, 2003.

[8] J. Fletcher and N. Evans, “An acoustic phonetic analysis of intonational prominence in two Australian languages,” J. Int. Phon. Assoc., vol. 32, no. 2, pp. 123–140, Dec. 2002, doi: 10.1017/S0025100302001019.

[9] L. Excoffier, P. E. Smouse, and J. M. Quattro, “Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.,” Genetics, vol. 131, no. 2, pp. 479–491, Jun. 1992, doi: 10.1093/genetics/131.2.479.

[10] M. Armstrong, M. Breen, S. Gooden, E. Levon, and K. M. Yu, “Sociolectal and Dialectal Variation in Prosody,” Lang. Speech, vol. 65, no. 4, pp. 783–790, Dec. 2022, doi: 10.1177/00238309221122105.

Assessment of finger tapping for rhythm control in performative speech synthesis

Christophe d'Alessandro, Grégoire Locqueville

Institut Jean le Rond ∂'Alembert, Sorbonne Université - CNRS, Paris (FRANCE)
christophe.dalessandro@sorbonne-universite.fr, gregoire.locqueville@ sorbonne-universite.fr

Performative Vocal Synthesis (PVS), developed initially as a new type of musical instrument, allows the real-time gestural control of synthesized speech through the modulation of voice pitch, syllable timing, vocal effort, and vocal quality [1,2]. As an interaction technique, PVS gives a "prothetic" voice whose gestures for control are explicit and externalized, contrary to the natural voice. It has been used for computer-aided intonation training in second language learning [3] and voice and motor reeducation (see the Gepeto project http://gepeto.dalembert.upmc.fr/).

The present research aims to assess the precision of finger taping for rhythm production, as it is used in the PVS system Voks [4]. The taping rhythm control paradigm is based on an analogy with speech production's frame/content theory [5]. Syllables are considered time "frames" (cycles of articulators open-close alternation) where segmental "contents" (phonemes) take place. These opening and closing cycles can be exploited for rhythmic control with the help of Syllabic Control Points (SCP). Vocalic Points (Pv) correspond to vocalic nuclei, and Intervocalic Points (Pi) correspond to intervocalic consonants (syllabic attack and coda). When the finger depresses the button, a Pv is triggered; when the finger raises the button, the Pi is triggered. Figure 1 shows the placement of Pv and Pi on the signal and the process of rhythm control using tapping.

The precision achieved by finger taping for rhythmic control is assessed using a prosodic imitation paradigm and a prosodic synchronization paradigm. A set of 8 French sentences ranging from 2 to 9 syllables, recorded by a male and a female speaker, is presented randomly to 8 subjects. In the imitation task, subjects listen to a sentence and are asked to reproduce it using Voks. The subjects' task is to reproduce as accurately as possible the prosody of sentences after listening to them, tapping the rhythm with a MacBook keyboard space bar (a non-synchronized motor repetition task [6]). In the synchronization task, they are asked to play the synthetic and natural sentence in synchrony (sensorimotor synchronization task [6]). The question whether biphasic control points (using Pv and Pi) or monophasic control points (using Pv or P-center only) are needed is investigated: sensorimotor synchronization experiments assume that each tap aims at only one rhythmic anchor [7] although motor control gestures are intrinsically biphasic (lowering and rising the finger, opening and closing the vocal tract).

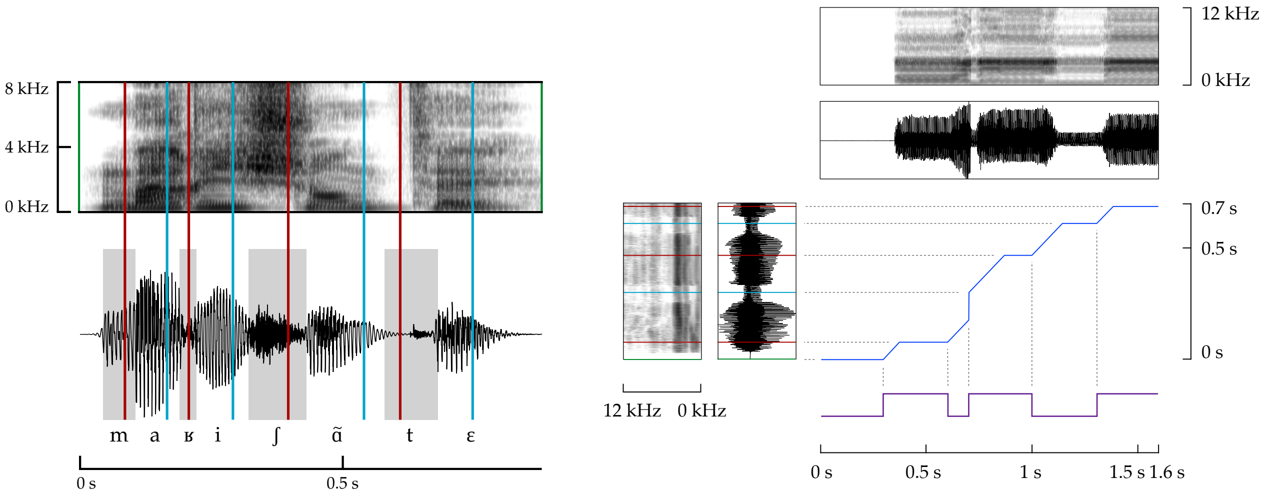


Figure 1. (Left) French sentence *Marie chantait*. Vocalic control points are in red; intervocalic control points are in cyan. (Right) Principle of biphasic taping rhythm control. Y-axis: original sound. Y-axis: tap sequences (bottom and synthesized sound. The time index is displayed as the synthesis time as a function of the original timeline.

Rhythmic precision is measured in this preliminary experiment for :1. comparing vocal and tapping rhythmic precision; 2. studying the effect of dominant/non-dominant hand; 3. studying the effect one or two fingers tapping; 4. studying the effect of simultaneous rhythmic and pitch control; 5. studying the effect of one vs two SCP per syllable. The tested conditions are: **Nat.** natural vocal production; **Cont. Pts**: monophasic tasks (one point per syllable, using Pv); **P-cent**. Monophasic tasks (one point per syllable, using P-centers); **Base**: biphasic task using Pv and Pi; **Hnd 2**: using the subjects' nondominant hand; **2 Fngs**: Using two fingers; **Tblt**: Simultaneously controlling pitch (as a distractor to the rhythmic task). The mean distance between production and stimulus phoneme boundaries is used to assess the precision obtained for each condition. Results are reported in Figure 2 (left: imitation, right: synchronization).

This preliminary experiment suggests that: 1. subjects are better at reproducing and synchronizing rhythm with their natural voice than with tapping; 2. the best tapping condition is biphasic with one finger; 3. one finger performs better than two; 4. the preferred hand performs slightly better; 5. simultaneous pitch and rhythm control is impairing precision; 5. monophasic conditions with P-centers and Pv are comparable; 6. the synchronization task is slightly less precise than the imitation task; 7. Biphasic control il more precise than Monophasic control.

In summary, the biphasic tapping paradigm shows good precision for performative rhythm control in speech. The experiments support the hypothesis that tapping using biphasic control points is a motor control process somewhat analogous to articulatory oscillation for syllable production in speech. Further studies are needed to extend this preliminary study to more subjects and to other language. Anticipation and of taps and individual strategies in rhythmic control must also be investigated.

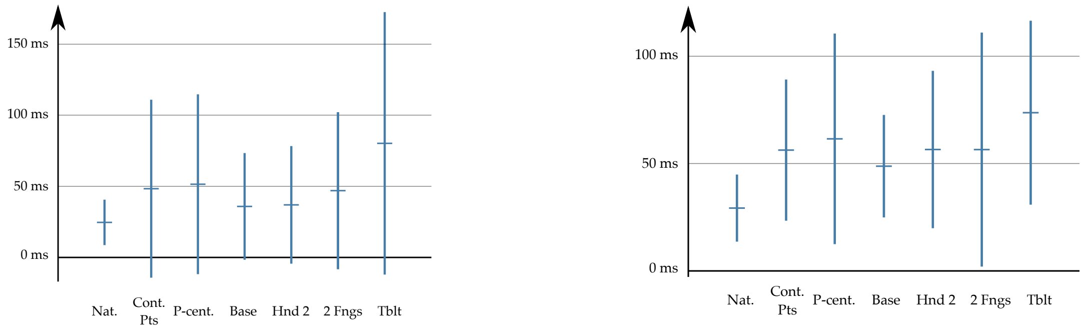


Figure 2. Mean and standard deviation of the absolute shift for imitation (left) and synchronization (right) tasks for all sentences with all subjects.

[1] Christophe d'Alessandro, Albert Rilliard, and Sylvain LeBeux (2011). Chironomic stylization of intonation. Journal of the Acoustical Society of America, 129(3):1594-1604, 2011.

[2] Samuel Delalez and Christophe d'Alessandro (2017). Adjusting the Frame: Biphasic Performative Control of Speech Rhythm. Proc. Interspeech 2017, 864-868, Stockholm, Sweden.

[3] F Xiao Xiao, Nicolas Audibert, Grégoire Locqueville, Christophe d'Alessandro, Barbara Kuhnert, and Claire Pillot-Loiseau (2021). Prosodic disambiguation using chironomic stylization of intonation with native and non-native speakers. Proc. Interspeech 2021,516-520. Brno, Czech Republic.

[4] Grégoire Locqueville, Christophe d'Alessandro, Samuel Delalez, Boris Doval, and Xiao Xiao (2020). Voks: Digital instruments for chironomic control of voice samples. Speech Communication, 125:97-113.

[5] Peter F. MacNeilage (1998). The frame/content theory of evolution of speech production. Behavioral and Brain Sciences, 21(4):499-511.

[6] Ho Tamara Rathcke, Chia-Yuan Lin, Simone Falk and Simone Dalla Bella (2021), Tapping into linguistic rhythm. Laboratory Phonology: Journal of the Association for Laboratory Phonology 12(1): 11, pp. 1–32, 2021.

[7] Bruno H. Repp (2005). Sensorimotor synchronization: A review of the tapping literature. Psychonomic Bulletin & Review, 12(6):969-992.

Taiwan Min Nan Checked Tones Sound Changes

Ho-hsien Pan[1] & Shao-ren Lyu[1]

[1]*National Yang Ming Chiao Tung University (Taiwan)*
hhpan@nycu.edu.tw, shaorenlyu@gmail.com

This study investigates the current state of sound change in Taiwan Min Nan (TMN) with regards to its seven lexical tones, which are related by tone sandhi chains. The sandhi rules are 55 → 13 → 33 → 31 → 51 → 55 and 3 → 5 → 3. Within a sandhi domain, a base tone surface in the tone sandhi domain final position, whereas sandhi tones surface in the non-final positions. The checked tones 3 and 5 are undergoing sound change that spans over generations. Fieldworks based on auditory impression noticed (1) deletion of glottal stops [ʔ], (2) vowel lengthening among base tone /5/ [5] (Liao, 2004, Chen, 2010, Ang, 2003), (3) lowering of f0 onset for base tone /5/ [5] (Chen, 2010, Ang, 2003) and (4) de-glottalized for base tone /5/ [5] (Chen, 2010).

To evaluate the current state of TMN checked tone change, native speakers of TMN, who were college students, discriminated AX stimuli pairs in citation forms containing sandhi-sandhi (SS), base-base (BB), and sandhi-base (SB) tonal pairs. Checked tones [3] and [5] were the most confusing, followed by level tones [55] and [33], and then low level and low falling tones [33] and [31]. Aside from similar tonal contours between two level tones or similar tonal registers between low tones [31] and [33], it is proposed that the tone sandhi relationship can induce perceptual confusion as well. For example, between the sandhi-related tonal pair [33] and [55], where the base tone /55/ [55] changes to the sandhi tone /55/ [33], is more prone to perceptual confusion than the sandhi-unrelated tone pair which includes the base tone /33/ [33] and the sandhi tone /51/ [55]. The two checked tones with two cyclic sandhi rules, 3 → 5 → 3, are most confusing perceptually.

This study also explores the acoustic similarities between checked tones. Read speech of disyllabic words containing sandhi tone in the first syllables and base tone in the second syllables were recorded from 40 TMN speakers, including four males and four females from each of the five dialect regions. Half of the speakers were under 30 years of age, whereas the other half were above 40 years of age. Following Pan (2017), the study compared the normalized f0, H1-A3c, and duration of sandhi and base checked tones with coda stops /p, t, k, ʔ/ realized as full stop, energy damping, irregular glottal pulse or complete deletion. As shown in Figure 2, and 3, in read speech, base tone 3 and 5 are merged in duration, f0 and H1-A3c domains, but sandhi tones 3 and 5 remain distinctive in terms of f0 and H1-A3c domains. As shown in Figure 4, deletion of coda stops was most prevalent among the glottal stops produced deep in the larynx, followed by the velar coda /k/ produced in the back of the mouth, and then by the /t/ produced in the front of the mouth. Coda stop /p/, which is visible, is least likely to be deleted. After coda deletion, vowels preceding deleted codas tend to be lengthened.

In summary, TMN checked tones are gradually losing their marked features, such as short duration and CV[stop] syllable structure. F0-wise, the merging of tonal registers gradually proceeds from checked base tones to checked sandhi tones. The loss of coda stops proceeds from glottal stops to velar stops, and then onto alveolar stops. Duration-wise, compensatory lengthening was proposed to cause duration lengthening after coda deletion. The lengthening of checked tones may cause further confusion between checked and unchecked syllables. Additionally, this tonal merging process begins with the less frequently occurring base tones and progresses towards the more commonly observed sandhi tones.

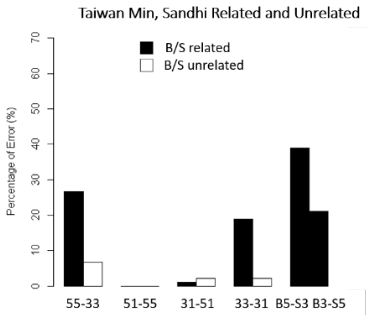


Fig.1 Percentages of error on sandhi related and unrelated tonal pairs in AX discrimination test.

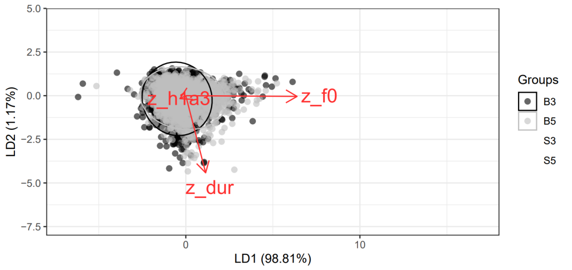


Fig.2 Results of Linear Discriminant Analysis on f0, duraiton and H1-A3c of base tones 3 and 5.

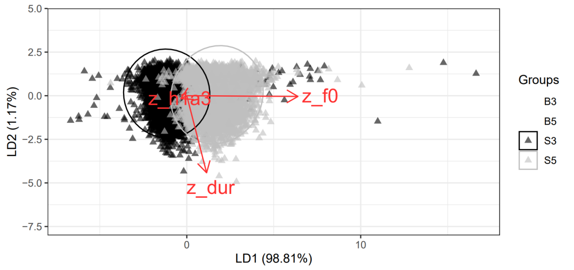


Fig.3 Results of Linear Discriminant Analysis on f0, duration and H1-A3c of sandhi tones 3 and 5.

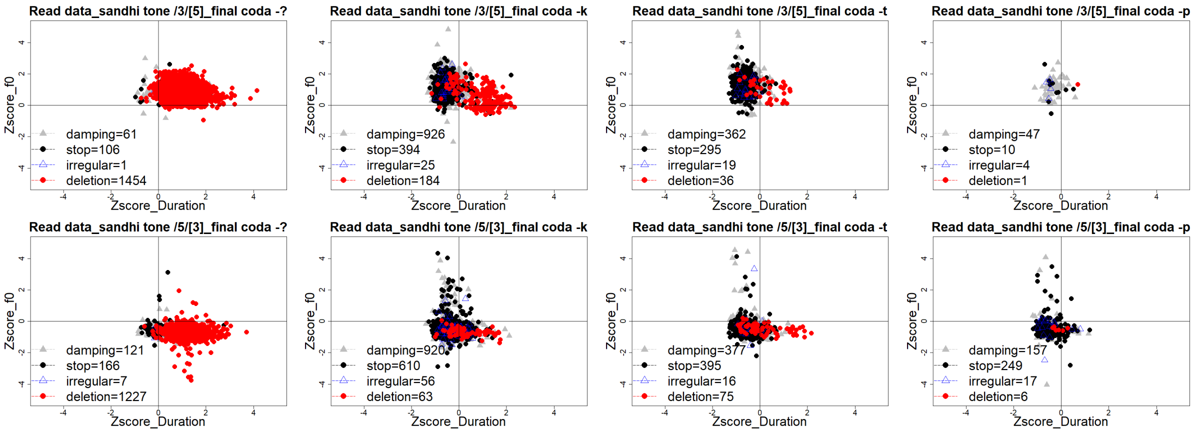


Fig.4 F0 and duration of sandhi tones 3 and 5 in different stages of coda weakening (▲: energy damping, ●: full stop, △: irregular glottal pulse, ●: complete deletion).

References

[1] Liao, J. C. (2004). *台語入聲調之現況分析* [*An investigation on the entering tones in Taiwanese*]. MA thesis, Taiwan : National Hsinchu University of Education.

[2] Chen, S.C. (2010). 台灣閩南語新興的語音變異 :台北市、彰化市及台南市元音系統與陽入原調的調查分析 [New sound variation in Taiwan Southern Min: Vowel system and the lower register entering tone in Taipei, Changhua and Tainan]. *Language and Linguistics, 11(2),* 425–468.

[3] Ang, U. J. *音變的動機與方向:漳泉競爭與台灣普通腔的形成* [*The motivation and direction of sound change: On the competition of Minnan dialects Chang-chou and Chuan-chou, and the emergence of General Taiwanese*]. Ph.D. dissertation, Taiwan : National Tsing Hua University.

[4] Pan, H. H., (2017). Glottalization of Taiwan Min checked tones. *Journal of the International Phonetic Association*, *47(1),* 37–63.

Phonological Categories in perception and production: the link and individual variability

Kuniko Nielsen

Oakland University (USA)
nielsen@oakland.edu

Fine phonetic details are used by listeners in processing speech and affect listeners' subsequent speech productions [1, 2] suggesting a close link between speech perception and production. However, the link between phonological categories in speech perception and production is still largely unknown. While previous studies found correlations between speech perception and production [e.g., 3-6], many studies also failed to find such correlations [e.g., 7-11], revealing the complex nature of the perception-production link.

The current study explores the link between perception and production by examining the correlation between individual speakers' categorical boundaries in perception and production through VOT. To capture a holistic picture of phonological categories, different variables that reflect production categories (e.g., mean, minimum, and maximum VOT for /p/ and /b/, and the midpoint of the gap between the two categories) are examined in both isolated and connected speech. Individual variability of perceptual categorical boundaries is also examined: while VOT production is shown to be highly variable across speakers yet structured within speakers [e.g., 12], individual variability of VOT in perception, such as the categorical boundary for voicing contrast, is largely unknown. Previous studies showed that perceptual cues vary across listeners [13, 14], so we would expect to find individual variability in categorical boundaries as well.

Thirty native speakers of American English participated in an online experiment which examines the correlation between perceptual categorical boundaries for the /p/-/b/ contrast and production variables of isolated and connected speech. A native speaker recorded the tokens bear and pear, and a 9-step *bear-pear* continuum was created (on a VOT scale from 12ms to 52ms) as identification task stimuli. Each token was presented twice in random order. Production stimuli for the isolated speech and connected speech were 36 monosyllabic words with initial stops and a short passage from the book "Peppa Pig: Family Trip" (116 words, 20 bilabial stops), respectively. During the production task, participants were randomly presented with the test words on the computer screen (each token was presented twice), and were asked to read them aloud at a comfortable pace. The short passage was presented subsequently, and the participants were asked to read it at a comfortable pace.

The results showed that perceptual boundaries in VOT vary widely across speakers (19.5ms - 47ms), and so did the slope. As expected, the mean production values in VOT also showed a wide variability (/p/: 45ms - 124ms, /b/:-80ms - 29ms). Mixed-effects modeling and additional linear regression analysis showed no significant correlation between /b/-/p/ perceptual categorical boundary and any of the production variables in isolated speech (p>0.1) (Fig. 1), while a significant correlation was found between the categorical boundary and mean /p/ VOT in connected speech (adjusted $R^2 = 0.41$, p<0.001) (Fig. 2). There was no significant correlation between isolated and connected speech nor between categorical boundary and mean /b/ VOT.

Our results showed a significant correlation between (perceptual) categorical boundary and mean /p/ VOT in connected speech, which suggests that the representation of phonemic categories is likely to include fine phonetic details. Our results also showed that the type of production task matters in examining the perception-production link: isolated speech was more hyperarticulated and variable in our data, potentially obscuring the link. One limitation of the current data is that the perception data comes from only one continuum (i.e., *bear-pear*). To address the lack of fine-grained perception data, additional data is currently being collected for a second experiment which includes more stimuli for the categorical perception task. The data will be analyzed using a Bayesian mixed effects regression model.

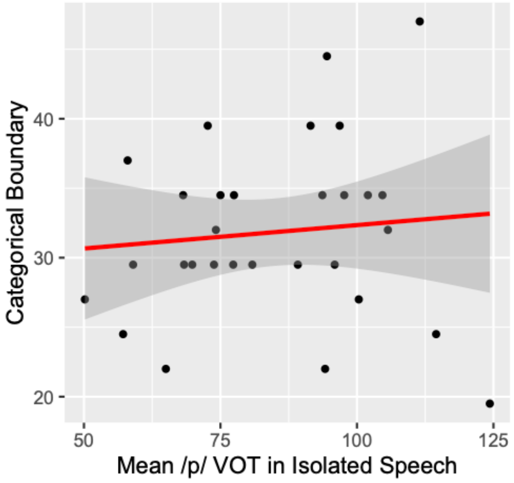


Fig.1 Mean /p/ VOT values and /b/-/p/ categorical boundary for each participant in isolated speech

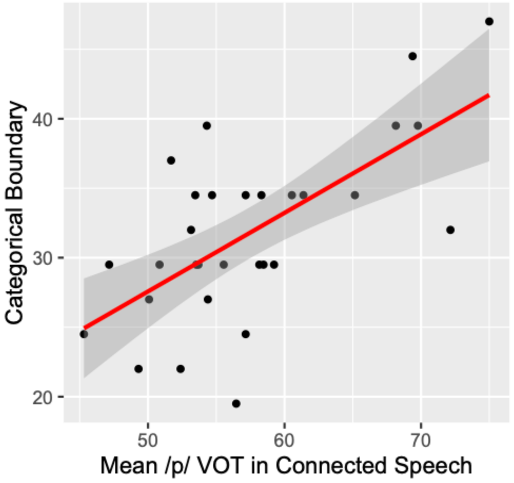


Fig.2 Mean /p/ VOT values and /b/-/p/ categorical boundary for each participant in connected speech

[1] McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. Journal of memory and language, 60(1), 65-91.

[2] Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. Psychological review, 105(2), 251.

[3] Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. The Journal of the Acoustical Society of America, 116(4), 2338-2344.

[4] Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. Language, 94(4), 931-968.

[5] Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. Language, Cognition and Neuroscience, 34(6), 769-786.

[6] Pinget, A. F., Kager, R., & Van de Velde, H. (2020). Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. Language and Speech, 63(3), 660-685.

[7] Bailey, P. J., & Haggard, M. P. (1973). Perception and production: Some correlations on voicing of an initial stop. Language and speech, 16(3), 189-195.

[8] Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. The Journal of the Acoustical Society of America, 113(5), 2850-2860.

[9] Kraljic, T., Samuel, A.G., and Brennan, S.E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. Psychological Science. 19:332–338.

[10] Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. The Journal of the Acoustical Society of America, 132(2), EL95-EL101.

[11] Cheng, H. S., Niziolek, C. A., Buchwald, A., & McAllister, T. (2021). Examining the Relationship Between Speech Perception, Production Distinctness, and Production Variability. Frontiers in Human Neuroscience, 15.

[12] Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. Journal of Phonetics, 61, 30-47.

[13] Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. Journal of phonetics, 28(4), 377-396.

[14] Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. Journal of Phonetics, 59, 40-57.

F0 enhancement in younger speakers of Standard Seoul Korean in internal and initial phrase position

Michaela Watkins[1]
[1]*University of Amsterdam (The Netherlands)*
m.m.watkins@uva.nl

Introduction

Standard Seoul Korean (SSK) has a unique three-way voiceless laryngeal stop contrast in word-initial position. For speakers born after 1965, there appears to be F0 enhancement coupled with a reduction of VOT in aspirated stops [1, 2]. This suggests that F0 is replacing VOT as the contrasting cue for the lenis-aspirated stop distinction. Despite these findings, some questions remain. Firstly, what happens to the third stop type, fortis, is unclear. Secondly, few studies have investigated other positions phrase-internally, with the majority testing only phrase-initial tokens. Finally, research has tended towards hyper-speech analysis [1, 2, 3] that may not reflect actual patterns in 'natural' speech.

The data observed here are partial results of an ongoing larger study of the *Korean Corpus of Spontaneous Speech* [4]. The larger study focuses on comparisons across generations of F0 production, with the results here the findings for male and female speakers of the youngest group. This group should have a strong enhancement effect of F0 when distinguishing between stop types, as all were born after 1965. It was expected that, particularly for lenis and aspirated stops, there should be a clear F0 distinction between the two with aspirated increasing, and lenis remaining low. VOT was expected to be overlapping between lenis and aspirated stops, suggesting that VOT is losing status as the contrasting factor between the two.

Method

The results presented here are for participants aged 15-19 (N = 10). F0 was measured in PRAAT [5] using a script that measured the mean F0 value per 50ms intervals throughout the vowel. VOT was measured manually through visual inspection of the spectrogram. Tokens were found using the script *Search, View & Save* [4] that allows selection of the token within in the phrase by gender and age group. A statistical analysis was performed using the *lme4* package [6] in *RStudio* [7]. A model was created for VOT and F0 respectively, with contrast coding applied to stop type (-a, +f, -l at -1/3, +2/3, -1/3 and +1/2, 0, -1/2), gender (+F-M at +1/2,-1/2) and sentence position (initial vs medial at +1/2, -1/2). An interaction was modeled between stop type, gender, and sentence position. Sentence position*StopType was coded as the random slope, with participant as a random factor. Visualisations were created using the *ggplot2* package [8].

Statistically, F0 demonstrated a distinction between lenis and aspirated stops with aspirated being higher than lenis by on average 28Hz (p = >0.01). Sentence position also was observed to be significant, with stops in initial position being on average 8Hz higher than medial stops (p = 0.02). Sentence position did not interact with gender or any stop type significantly. VOT demonstrated a significant effect for fortis stops, with an average of 45ms shorter than other stop types (p = >0.01). No significant effect was observed between lenis and aspirated stops. An interaction effect was observed between sentence position and aspirated stops, with initial aspirated stops being on average 19.7ms (p = 0.02) shorter in initial position compared with internally. Female speakers were also seen to bear a shorter aspirated stop than male speakers, at 18ms shorter (p = 0.04).

These results suggest that the lenis-aspirated distinction is maintained by F0, and that this is stronger in initial position compared to phrase-internal. Fortis remains distinguished by VOT, with no suggestion of a difference dependent of phrase position. That aspirated stops are shorter in initial position suggests a reduction in the length of aspirated stops, thus assimilating further with lenis, particularly for younger female speakers. Prosodic effects may be the cause of sentence internal lengthening, requiring further investigation.

Graphical results suggest that aspirated stops remain higher in F0 and longer in VOT compared to other stops across phrase positions. In medial position, however, lenis does not remain stable but rather shows an overlap with fortis, especially in F0. This was not observed statistically due to the nature of the contrast coding (as fortis was compared rather than lenis). How these are distinguished in medial position is not clear from initial evaluation, and further implies that F0 enhancement is unequal across the phrase. The reasons for this could possibly be due to another cue not tested here providing the distinction, or the prosodic structure of SSK impacting F0 enhancement. Additionally, it is important to consider the small sample size, and to evaluate how this changes upon introducing more participants. Female speakers show an inconsistent F0 in their fortis stops, which could be either due to underpowering of the data or possibly due to bimodality; fortis has no need to alter its F0 and relies only on VOT. This could imply the development of a bimodal contrast system, possibly as an alternative to full tonogenesis. As younger female speakers are often given as the drivers of linguistic change [9], this warrants further investigation with a larger pool of participants.

In summary, these results imply that F0 enhancement is occurring, more so in initial than internal position. Aspirated stops appear to be stable with a consistently high F0 and long VOT in both positions. VOT remains significant, suggesting that it is not redundant but possibly a secondary cue. The indication that there is some degree of overlap between fortis and lenis stops in medial position requires further investigation into other cues (e.g. voice quality) and more in-depth analyses of the prosodic structure of SSK. A larger pool of female participants would also be beneficial to study, to assess if a bimodal contrast system has formed fully in younger speakers. Finally, to truly evaluate if F0 is replacing VOT as the primary cue, a perception experiment is required across different phrase positions.

References

[1] Kang, K-H. & Guion, S.G. (2008). Clear speech production of Korean stops. *JASA*, 124:3909-3917.

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

[3] 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. *Journal of Phonetics* 66:120-144.

[4] Yun, W., Yoon, K., Park, S., Lee, J., Cho, S., Kang, D., Kim, J. (2015). The Korean Corpus of Spontaneous Speech. *Phonetics and Speech Sciences*. The Korean Society of Speech Sciences

[5] Boersma, P., and Weenink, P. (2023). PRAAT: Doing phonetics by computer Version 6.4.06 [computer program], from http://www.Praat.org/

[6] Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1-48.

[7] R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[8] Wickham., H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

[9] Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors*. Oxford: Blackwell

How does prosody distinguish Wh-statement from wh-question in Shanghai Chinese

Bijun Ling

Tongji University (China)
lingbijun@tongji.edu.cn

It has been widely acknowledged that wh-phrases in many languages (e.g. Chinese, Japanese, Korean) are ambiguous between interrogative and indefinite interpretations. Although wh-words can have different interpretations, the interpretation of wh-words in each sentence is in general unambiguous, as the different interpretations are connected to different licensors. Take Standard Chinese for example, in the absence of overt licensors, a wh-word like jǐ is typically interpreted as an interrogative word ('how many') and the sentence is a wh-question, as (1a); the indefinite interpretation of the wh-word jǐ ('several, many') has been shown to be licensed by sentences containing negation (1b), conditionals (1c), epistemic modalities (1d) or by yes-no questions (1e) (Yang et al., 2020) .

(1a) tā mǎi-le jǐ běn shū?
he buy-ASP how many m.w. book
How many books did he buy?

(1b) tā méi mǎi jǐ běn shū.
he not buy several m.w. book
He didn't buy many books.

(1c) rúguǒ tā mǎi-le jǐ běn shū, wǒ huì hěn kāixī.
if he buy-ASP several m.w. book, I will very happy
If he bought several books, I will be very happy.

(1d) tā hǎoxiàng mǎi-le jǐ běn shū.
he seem buy-ASP several m.w. book
He seems bought several books.

(1e) tā mǎi-le jǐ běn shū ma?
he buy-ASP several m.w. book Q-particle
Did he buy several books?

Though most cases of wh-words are unambiguous, there are a few instances where the wh-word is in fact ambiguous between a declarative and question interpretation, as illustrated in (2).

(2) Zhāng Sān mǎi-le jǐ běn shū gěi Lǐsì
Zhang San buy-ASP *how many/several* m.w. book for Lisi

(2a) How many books did Zhangsan buy for Lisi? [wh-question]

(2b) Zhangsan bought several books for Lisi. [wh-statement]

With ambiguous sentences like (2), the question arises how can the wh-question be distinguished from the wh-statement? Previous studies have demonstrated that prosody interacts with wh-phrases in languages like Japanese (Ishihara, 2007), Korean (Jun & Oh, 1996), German (Truckenbrodt, 2013) and Standard Chinese (Yang et al., 2020) in that wh-interrogatives manifest phonetic prominence whereas wh-indefinites do not. Furthermore, wh-questions in Japanese and Korean are also characterized by a post-wh-word de-phrasing, namely, a deletion of accentual phrasings following the wh-word, but there was no sign in Standard Chinese or Germany. In order to further our understanding of syntax-phonology interface, the paper aims to investigate how Shanghai Chinese speakers refer to prosodic cues in differentiating the ambiguities between jǐ-interrogative and jǐ-indefinite.

Shanghai Chinese, a Wu dialect, has five citation tones and they undergo sandhi changes when syllables are combined into words or phrases, as illustrated in Table 1. Furthermore, Selkirk & Shen (1990) proposed three types of prosodic units in Shanghai Chinese (prosodic word, prosodic phrase and intonational phrase) and the mapping rules between syntax and prosodic units.

The unique prosodic features make Shanghai Chinese an interesting case for the study of the semantics-prosody interface and syntax-phonology interface. Our research questions are as follows:

(1)When statements are string identical to questions, how do speakers use prosodic cues to disambiguate wh-questions from wh-statements in Shanghai Chinese? Specifically, does the process modify the prosodic phrasing like in Japanese and Korean or maintain the same prosodic structure like in Standard Chinese ?

(2) When there are overt licensors and the interpretation of wh-words is unambiguous, do speakers still use different prosodic cues to represent wh-questions and wh-statements? Essentially, wh-word interpretations are disambiguated only depending on syntactic licensors or on prosody as well?

Table 1: the value of citation tones and sandhi tones (using Chao's five-level numerical scale, which divides a speaker's pitch range into five scales with 5 indicating the highest and 1 the lowest).

Register	**Duration**		
	long [CV(N)]		short [CVʔ]
high	T1[HL]	T2[MH]	T4[$\underline{\text{H}}$]
low		T3[LH]	T5[$\underline{\text{LM}}$]
	falling	rising	
Contour			

Citation tone		**Sandhi tone**	
		T+X	**T+X+X**
T1	53	55+31	55+33+31
T2	34	33+44	33+55+31
T4	$\underline{55}$	$\underline{22}$+44	$\underline{33}$+55+31
T3	13	33+44	22+55+31
T5	$\underline{12}$	$\underline{11}$+13	$\underline{11}$+22+13

Based on our production data, wh-statements are lower in F0 and smaller in F0 range than wh-questions at the wh-word and there is a F0 range compression in the post-wh-word region in wh-questions. An implication of this study shows that wh-words are foci in wh-questions but cannot be foci in wh-declaratives. Therefore, we conclude that the F-feature that is treated as lexically inherent to wh-words in Truckenbrodt (2013) is however claimed to be unspecified in Shanghai Chinese. Furthermore, wh-word interpretations are disambiguated mainly depending on syntactic licensing conditions, whereas prosody might serve a subsidiary role.

References:

[1] Yang, Y., Gryllia, S., & Cheng, L. L. (2020). Wh-question or wh-declarative? Prosody makes the difference. *Speech Communication*, 118, 21-32.

[2] Ishihara, Shinichiro. (2007). Major phrase, focus intonation, multiple spell-out. *The Linguistic Review*, 24.2-3: 137-167.

[3] Jun, Sun-ah & Mira Oh. (1996). A prosodic analysis of three types of wh-phrases in Korean. *Language and Speech*, 39.1: 37-61.

[4] Truckenbrodt, Hubert. (2013). An analysis of prosodic F-effects in interrogatives: prosody, syntax and semantics. *Lingua*, 124: 131-175.

[5] Selkirk, E., & Shen, T., (1990). Prosodic domains in Shanghai Chinese. in Inkelas, S., and Zec, D. (eds.) *The phonology-syntax connection (*pp. 313–337*)*. Chicago: University of Chicago Press .

Differential effects of prosodic boundary on glottalization of word-initial vowels in Korean: A preliminary report

Sungwok Hwang[1], Sahyang Kim[2], and Taehong Cho[1]

[1]*Hanyang Institute for Phonetics & Cognitive Sciences of Language, Hanyang University (Korea),*
[2]*Hongik University (Korea)*
sungwokhwang@gmail.com, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Voice quality is often described in terms of vocal fold approximation along a continuum [1], with voiceless sounds [ʔ] and [h] positioned at opposite ends. It is widely recognized that voice quality, including glottalization ([ʔ]), is influenced by prosodic structure, such as larger prosodic junctures or prominence [2,3,4], highlighting the phonetics-prosody interface. For instance, word-initial vowels tend to exhibit more glottalization at the onset of an Intonational Phrase (IP) compared to within the IP, serving as markers for prosodic boundaries. However, in English, glottalization is often associated with prominence rather than boundary marking [4]. This raises questions about the specific role of glottalization in marking prosodic structure, considering the interplay between boundary and prominence marking in different languages. Additionally, glottalization of word-initial vowels can also serve to avoid vowel hiatus across word boundaries [cf. 5], which may not be directly related to prominence or higher prosodic junctures. Our study focuses on Korean, where prominence marking is closely tied to boundary marking as an edge-prominence language [6], aiming to explore how glottalization, along with the temporal variation (preboundary lengthening, pause), operates within the phonetics-prosody interface. Furthermore, we investigate the influence of syntactic structure on phonetic implementation to examine if and how the phonetics-prosody interface can be further modulated by syntactic structure, particularly in the context of syntactic disambiguation.

In this study, we conducted an acoustic experiment with a group of fourteen native Seoul Korean speakers. The participants read sentences containing vowel-initial target words, including proper nouns (N1, N2, N3, such as /ali/, /aʧi/, /ami/), conjunction suffix /-hako/ (‘-and’), and conjunction /animjʌn/ ('or') in syntactically ambiguous coordinate structures, as described in Table 1. These sentences were produced under various focus conditions, allowing for enriched contexts that facilitate the observation of interactions among phonetics, prosody, and syntax. (For simplicity, the focus-related effects are not reported in this study.) We measured H1*-H2* (degree of glottalization) and HNR (noise-related measure) of word-initial vowels for N2 and N3, as well as for the conjunction /animjʌn/. Additionally, we analyzed the duration of the syllable preceding the prosodic juncture, as well as the pause duration. The determination of prosodic boundaries (IP or Word (Wd)) was made collectively by all three authors.

Table 1. An example of test sentences with different syntactic contexts. Proper nouns (/ali/, /aʧi/, /ami/) could be placed in all three locations: Noun1 (N1), Noun2 (N2), Noun3 (N3).

	Q: musɨn ilija	“What’s happening?”
Early Closure	A: a (N1-hako) (**N2 animjʌn N3**)-ka ontɛ	“Well, (N1-and) (N2 or N3) are coming.”
	Q: musɨn ilija	“What’s happening?”
Late Closure	A: a (N1-hako **N2**) **animjʌn** (**N3**)-ka ontɛ	“Well, (N1-and N2) or (N3) are coming.”

Our prosodic analysis revealed a consistent mapping between syntactic juncture and prosodic boundary, as depicted in Fig.1a. Specifically, for Early Closure, an IP boundary consistently aligned with the syntactic juncture with Phrasing Type 1 ([N1-and] # [N2 or N3]), and for Late Closure with Phrasing Type 2 or Phrasing Type 3 (e.g., ([N1-and N2] # or **(#)** [N3]) where '(#)’ denotes an optional boundary). Importantly, however, our results also demonstrated differential effects of the same type of prosodic boundary on the phonetic implementation, relative to the syntactic structure. Regarding temporal variation, the temporal expansion (preboundary lengthening + pause) was more prominent for IP boundaries in Early Closure parsing compared to Late Closure parsing. It was greater for the Phrasing Type 1 (i.e., [N1-and] **#** [N2 or N3], Fig.1b) than for the Phrasing Type 2/3, (i.e., [N1-and N2] **#** or (#) [N3], Fig.1c). Note also that the temporal expansion was larger for the critical IP juncture for Late Closure (i.e., after N2), compared to the optional IP one (i.e., after ‘or’). In terms of glottalization, the presence of an IP boundary did not consistently induce glottalization. For the IP juncture after [N1-and] in Phrasing Type 1, which exhibited the most robust temporal expansion, no boundary-related glottalization was observed (Fig. 1b). Conversely, the IP junctures after N2 (Phrasing Type 2) or ‘or’ (Phrasing Type 3) showed increased glottalization of the initial vowel at the IP-initial position compared to the IP-medial position (Fig. 1c-d).

The main findings of this study highlight a consistent mapping between syntax and prosody in syntactic disambiguation, as discussed in previous literature [e.g., 7]. Notably, our results reveal that glottalization is not always used to mark a higher prosodic (IP) juncture, despite the expected association between prosodic phrasing and prominence marking in Korean as an edge-prominence language [6]. Instead, glottalization appears to be utilized more for avoiding vowel hiatus within phrases, offsetting its use as a marker of higher prosodic junctures in some cases. On a related note, our findings indicate that the phonetics-prosody interface, representing the phonetic implementation of prosodic structure categorically defined by the intonational phonology of the language, is finely tuned by syntactic structural information. We observe varying degrees of glottalization and temporal expansion, possibly in a reverse direction. The robust temporal expansion of an early IP serves as a clear cue for critical syntactic junctures (Early Closure), while glottalization plays a minimal role in this context. However, glottalization becomes more prominent when the temporal cue is less robust. We propose that voice quality is modulated by system-driven factors, particularly the motor system, which considers the relative contributions of available suprasegmental and segmental cues in signaling prosodic structure.

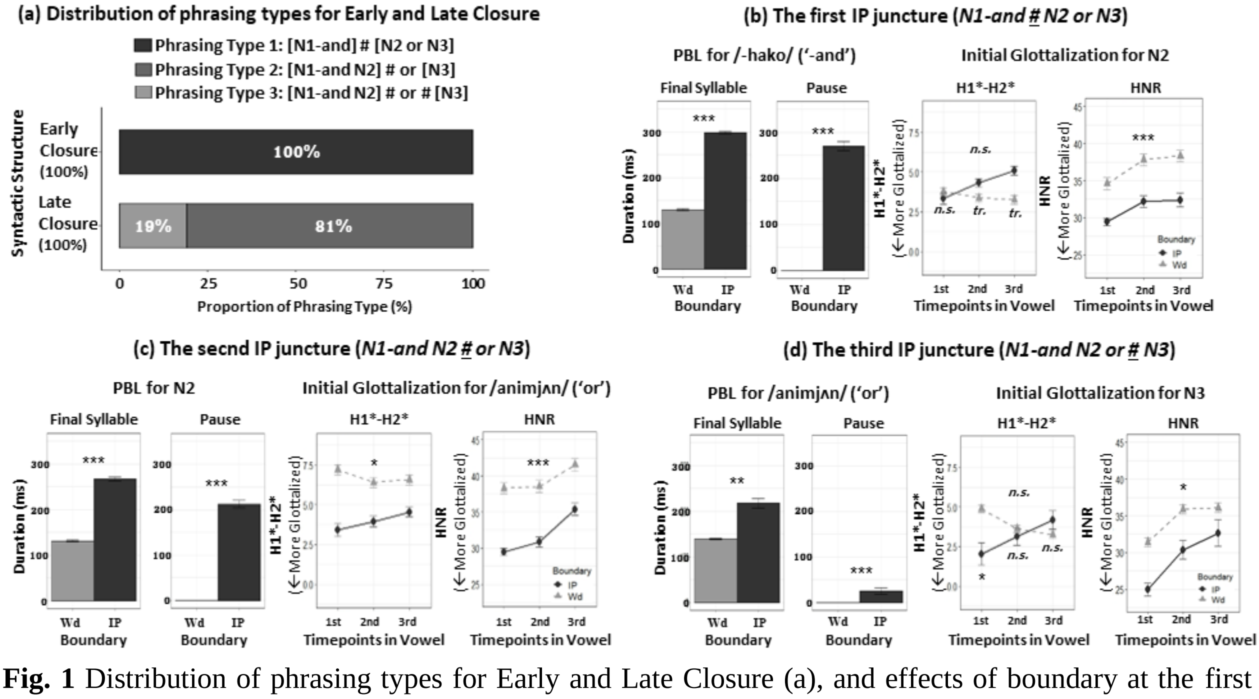


Fig. 1 Distribution of phrasing types for Early and Late Closure (a), and effects of boundary at the first (b), second (c), and third juncture (d) on final syllable and pause duration, and initial vowel’s glottalization. *n.s.*, *p*>0.1; *, *p*<0.05; **, *p*<0.01; ***, *p*<0.001. Error bars refer to standard errors.

References

[1] Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*, University of Chicago.

[2] Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24(4), 423-444.

[3] Mitterer, H., Kim, S., & Cho, T. (2019). The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language* 108, 1-19.

[4] Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of phonetics* 94: 101155.

[5] Davidson, L. (2021). Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian. *Journal of Phonetics* 88, 101075.

[6] Jun, S.-A. “Korean Intonational Phonology and Prosodic Transcription,” in *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford University Press, 2005.

[7] Elfner, E. 2018. The syntax-prosody interface: Current theoretical approaches and outstanding questions. *Linguistics Vanguard*, 4(1), 1-14.

The Effect of L1 Tones and L2 Pitch Accent on Lexical Access

Yusheng Mu[1]

[1]*Shanghai International Studies University (China)*
muyusheng@shisu.edu.cn

For bilingual speakers, acoustic features of words in L1 can influence how their translations are perceived in L2 [1, 2]. Few previous studies have explored the effect that suprasegmental information may have on L2 lexical processing. Shook and Marian (2016) found that Mandarin/English bilinguals performing a translation task would look toward the translation shu4 on the screen more quickly if they heard the English word tree spoken with a falling pitch (matching shu4) than if the pitch mismatches the tone [2]. But their study did not analyse words with different tones separately. The present study investigates whether suprasegmental information has a positive effect on lexical access of all four tones, or only some of them.

To examine whether suprasegmental information in Chinese affects lexical access in English, 80 monosyllabic English words are chosen as materials, whose Chinese meaning can be expressed in one character. These are evenly divided into 4 groups according to the tones of the characters (i.e., 20 English-Chinese pairs under each tone group). All the English words are recorded by a male Mandarin–English bilingual with one of the four lexical tones in Mandarin. Each word is recorded twice, with the tone contour matching or mismatching the tone of corresponding Chinese character. The recordings are then played to 27 native Chinese speakers (15 males and 12 females), who are asked to choose a corresponding Chinese character each time they hear an English word. The target character for each word is presented together with a distractor, a segmentally and suprasegmentally different Chinese character, which has a different meaning. They appear at equal distances from the left and right sides of the screen. Participants' response time is defined as the time lapse between the end of the audio and the identification of the target character and is measured in milliseconds.

The results are analysed separately for each of the four tone groups. As Figure 1 shows, the average response time for Tone 1 and Tone 4 in matching trials is shorter than that in mismatching trials, but the same cannot be seen for Tone 2 and Tone 3. Results of Tones 1 and 4 reveal that suprasegmental information has a facilitatory effect of lexical access. Compared with the result of Shook and Marian (2016), when tones are discussed separately, Tones 2 and 3 show different results. The difference in the results between groups 1, 4 and 2, 3 needs further investigation [2].

A conjecture is put forward that it is not the four tones in Mandarin that promotes the lexical access, but the pitch accent of the English spoken by native speakers of Mandarin [3]. This is explored in a follow-up experiment in which 8 Chinese learners of English are asked to read an English passage of 820 words. The syllables on which the pitch accents are located are selected for analysis, in which F0 contours of the accents are fitted into linear models which can be expressed by the equation $y = kx + b$. For each pitch accent, the data for parameters k and b are plotted in a scatter plot and clustered into 4 categories using Agglomerative Clustering algorithm, as shown in Figure 2.

The cluster in purple has the most points, followed by the yellow, the blue, and the green one. This reveals that high level tones (k close to 0 and high b) and falling tones (negative k) appear more frequently than the others in the pitch accent of the English spoken by native speakers of Mandarin, thus providing evidence that it is the pitch accent in the English spoken by native Chinese speakers that promotes the lexical access, not the four tones in Mandarin per se.

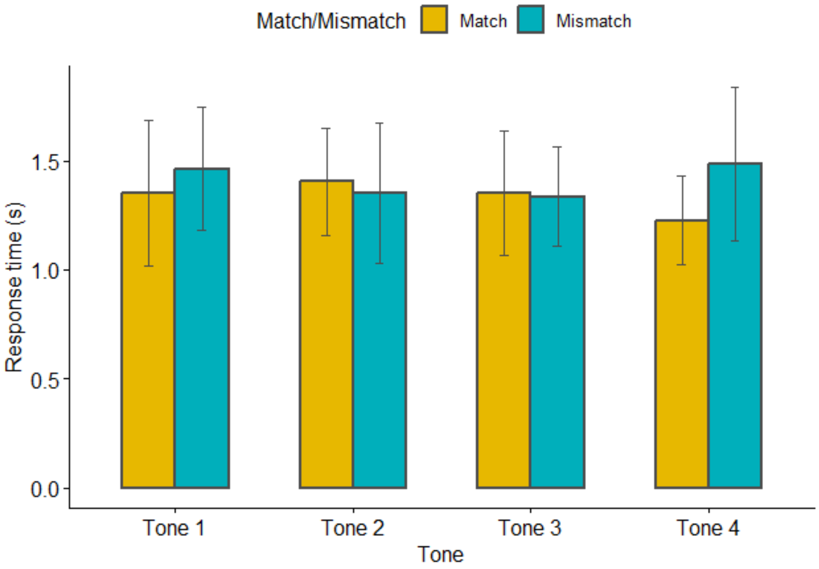


Fig.1 Average response time under match/mismatch conditions in each tone group

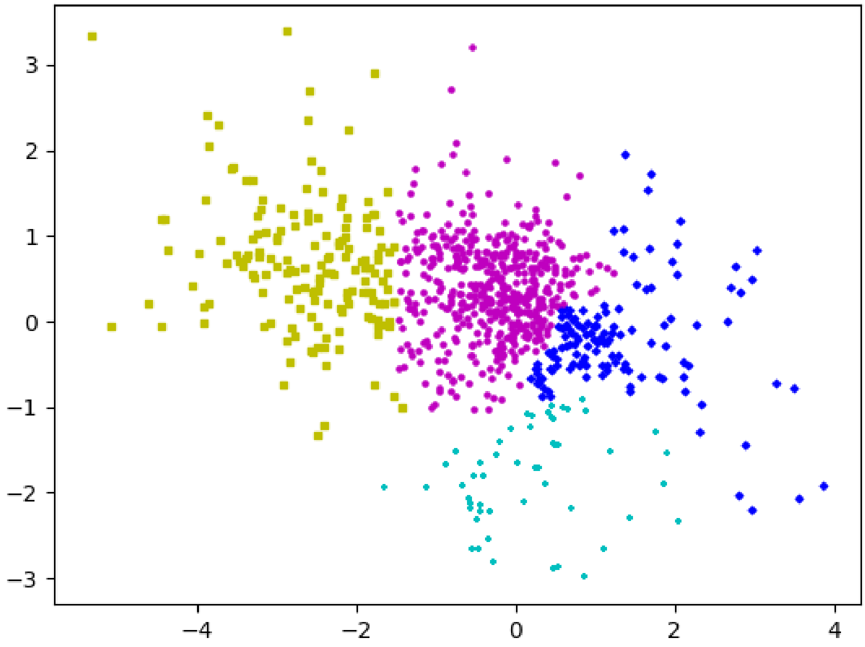


Fig.2 Clustering scatter plot of pitch accents

References

[1] 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.

[2] Shook, A., & Marian, V. (2016). The influence of native-language tones on lexical access in the second language. *Acoustical Society of America*, *139*(6), 3102-3109.

[3] Zahner-Ritter1, K., Zhao, T., Einfeldt, M., & Braun, B., (2022). How experience with tone in the native language affects the L2 acquisition of pitch accents. *Frontiers in Psychology*, *13*, 1-24.

Development of pitch cues in tone discrimination: evidence from Cantonese

Zhenting Liu[1] & Regine Lai[2]

[1]The Chinese University of Hong Kong (Hong Kong), [2]The Chinese University of Hong Kong (Hong Kong)
ztliu@link.cuhk.edu.hk, ryklai@cuhk.edu.hk

Previous research has shown that pitch cues including average pitch height, pitch contour, and pitch at critical points are important in tone perception [1,2], and listeners with different language background have different weightings of pitch cues. Specifically, average pitch height is an important pitch cue across languages [2,3], and non-tone language speakers attend more to pitch height than tone language speakers [4,5]. However, previous studies did not address the question concerning development of pitch cues in early years of life, i.e., (i) whether there is a language-general weighting of pitch cues prior to language exposure and (ii) when they start to change with language exposure. Although many studies on discrimination of lexical tones in infants have revealed different developmental patterns for tone and non-tone language learning infants [6-8], the relative role of pitch cues in tone discrimination is still not clear given the fact that multiple pitch cues are covarying with each other, and that tasks involving identification or judgment on multiple tone pairs could be difficult for infants. Therefore, the current study aims to investigate the role of 4 pitch cues, i.e., average height (AH), contour, onset pitch and offset pitch in tone discrimination in Cantonese-learning infants between the ages of 6m and 14m using manipulated tone pairs. Cantonese adults have been shown to weigh pitch height more importantly than pitch direction [3], and pitch offset more importantly than pitch onset in tone perception [2]. It is though yet to know whether Cantonese children behave the same as their adult counterparts. In order to answer this question, an experiment was conducted as follows.

To tease apart the relative role of each cue in the current study, we used one pair of sub-phonemic tone contrasts within each cue condition (Fig.1). Tone contrasts used in the *Contour* condition are two level tones with 7 Hz differences at every 500 ms interval. Tone contrasts in *AH*, *Onset* and *Offset* conditions included one rising tone and one falling tone sharing the same AH, onset and offset respectively (slope = ±14 Hz/s). This manipulation ensured one particular cue was kept constant in each condition. If that cue is important for perception, it would be relatively difficult for participants to perceive tone differences. Cantonese learning children at 6m (N = 72, 18/condition), 9m (N = 72), 12m (N = 72) and 14m (N = 80) were tested using a discrimination task with these four pairs of pitch trajectories superimposed on the syllable [ma]. A habituation-based visual fixation paradigm was adopted.

The results were analyzed using linear mixed-effects (LME) model. A significant main effect of Trial (Habituated vs. Novel) and Age, as well as 3 significant interactions: Trial × Condition, Condition × Age, and Trial × Condition × Age (Table 1). A separate LME model was conducted for different age groups, and only a significant Trial × Condition interaction was found in 9m and 12m group (Table 2). Post-hoc paired-sample t-tests comparing mean looking time revealed that although children at 6 and 14 months of age can discriminate tone contrasts in all conditions, 9-month-old children cannot discriminate tone contrasts in AH condition, and 12-month-old group cannot discriminate tone contrasts in AH and Offset condition (Fig.2).

The results suggest that Cantonese learning infants at 12 months of age begin to rely on AH and Offset as important cues for tone perception like adults, whereas there is an earlier tendency to weigh AH more importantly than other cues, providing the possibility that AH might be inherently given more weight in tone discrimination than other cues. Successful discrimination of tone contrasts in any condition in 6m and 14m group demonstrates that Cantonese learning infants are able to integrate other pitch cues in the absence of any pitch cue in tone discrimination. This result is similar to the U-shape development pattern found in non-tonal language learning infants in previous studies [6,9], suggesting a generally higher sensitivity to acoustic differences before 6 and after 12 months of age. And a drop in discriminatory ability in the ages between are possibly due to perceptual reorganization affected by native phonetic properties. Incapability of inferring the relative importance of cues when infants can equally discriminate all contrasts is a limitation of the current design. However, a further study testing discrimination of these tone pairs in infants learning other languages would be helpful to see whether AH is naturally the most important cue in tone perception regardless of language background.

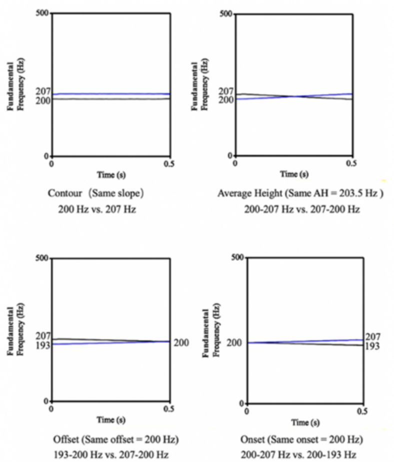


Fig.1 Acoustic details of tone contrasts in 4 cue conditions

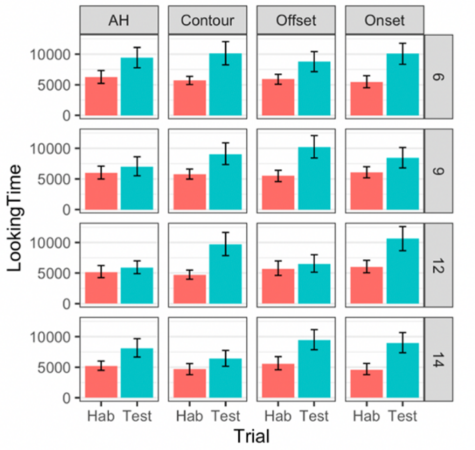


Fig.2 Mean looking time of habituation trials and novel test trials

Table 1 Results of LME assessing the effects of Trial, Age and Condition

	F	*df*	*p*
Trial	**143.79**	**872**	**<0.001*****
Condition	0.87	280	0.465
Age	**3.75**	**280**	**0.011****
Trial×Condition	**2.63**	**872**	**0.049****
Trial	0.73	872	0.531
Condition×Age	**1.80**	**280**	**0.068***

Table 2 Results of LME conducted for separate age group

	F	*df*	*p*
6 months			
Trial	**42.49**	**212**	**<0.001*****
Condition	0.26	68	0.853
Trial×Condition	1.14	212	0.333
9 months			
Trial	**27.92**	**212**	**<0.001*****
Condition	0.64	68	0.593
Trial×Condition	**2.35**	**212**	**0.074***
12 months			
Trial	**29.67**	**212**	**<0.001*****
Condition	**2.84**	**68**	**0.044****
Trial×Condition	**4.28**	**212**	**0.006*****
14 months			
Trial	**45.76**	**236**	**<0.001*****
Condition	**2.24**	**76**	**0.091***
Trial×Condition	1.29	236	0.278

Reference

[1] Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*(1), 1–33.

[2] Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*(1), 104–117.

[3] Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149–175.

[4] Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Citation: The Journal of the Acoustical Society of America*, *128*, 456.

[5] Liu, L., Lai, R., Singh, L., Kalashnikova, M., Wong, P. C. M., Kasisopa, B., Chen, A., Onsuwan, C., & Burnham, D. (2022). The tone atlas of perceptual discriminability and perceptual distance: Four tone languages and five language groups. *Brain and Language*, *229*, 105106.

[6] Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385–394.

[7] Singh, L., Fu, C. S. L., Seet, X. H., Tong, A. P. Y., Wang, J. L., & Best, C. T. (2018). Developmental change in tone perception in Mandarin monolingual, English monolingual, and Mandarin–English bilingual infants: Divergences between monolingual and bilingual learners. *Journal of Experimental Child Psychology*, *173*, 59–77.

[8] Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*(2), 123–139.

[9] Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual reorganization of lexical tones: effects of age and experimental procedure. *Frontiers in Psychology*, *9*, 477.

Final lengthening at Tone Sandhi Group boundaries in Taiwan Southern Min: Boundary strength and surprisal

Sheng-Fu Wang
Institute of Linguistics, Academia Sinica (Taiwan)
sftwang@gate.sinica.edu.tw

This study examined syllable duration as a cue to the right edge of Tone Sandhi Groups (TSGs) in Taiwan Southern Min (TSM). There are two research questions. First, what is the durational pattern before TSG breaks? Specifically, does the presence of pre-boundary lengthening at TSG breaks rely on the overlap between TSG breaks and boundaries of prosodic units such as intermediate phrases (ip) and intonational phrases (IP)?

Tone sandhi in TSM is a phonological process where all syllables except for the rightmost one in a TSG, which correspond to syntactic constituents [1] categorically switch their tones. The boundaries of TSG are defined based on the presence of syllables without a categorical tonal change, i.e., in their citation tones. For example, in /hak$^{4\rightarrow 2}$ siŋ^{55}# aj$^{21\rightarrow 53}$ t^{h}ak$^{4\rightarrow 2}$ ts^{h}eʔ^{2}#/ "Students love/have to read books", the TSGs /hak$^{4\rightarrow 2}$ siŋ^{55}/ "students" and /aj$^{21\rightarrow 53}$ t^{h}ak$^{4\rightarrow 2}$ ts^{h}eʔ^{2}/ "love/have to read books" are identified based on the presence of /siŋ^{55}/ and /ts^{h}eʔ^{2}/ in the citation tones. This definition is different from that of ip and IP breaks in TSM, which are defined and identified based on the gradient cues such as lengthening and pitch range [2]. For the present study, we examined whether TSG breaks, defined without reference to prosodic cues, exhibit similar pre-boundary lengthening patterns as previously found in TSM [3].

The second research is how the correlation between high surprisal and longer duration changes in different prosodic conditions. The correlation between a linguistic unit's predictability (i.e., lexical frequency, conditional probability) and its acoustic cues has been well-established [4, 5, 6] We aimed to test the hypothesis that the prosodic structure may constraint the direct relationship between acoustic cues and information [7, 8] by examining the correlation between syllable duration and surprisal at different prosodic conditions. The measurements include unigram surprisal $-log\ P(word)$ representing lexical frequencies and forward/backward bigram surprisal $-log\ P(word|context)$ representing local contextual probability [5, 6].

Speech data were extracted from an 8-hour spontaneous speech corpus [2] with materials from 16 speakers, which contained annotations of TSGs, ips, and IPs. The data that went into analysis contained 10934 TSGs made up of 46414 syllables (35545 words). To control for the influence of prosodic conditions, we investigated syllables at TSG boundary matching with an intonational phrase boundary (TSG+IP), an intermediate phrase boundary (TSG+ip), and neither (TSG-only). Surprisal were obtained from a language model trained with the SRILM toolkit [9] with modified Kesner-Ney smoothing [10] on a written corpus containing 4.7 Million words [11].

Results showed penultimate and final lengthening at all three types of TSG boundaries ($p < .0001$ for all pairs of comparisons). In other words, TSG boundaries exhibited pre-boundary lengthening even without overlapping with larger iP and IP breaks. The overlap mostly resulted in an incremental lengthening of the final syllable: TSG+IP boundaries have the longest final syllable, followed by TSG+ip boundaries and TSG-only boundaries ($p < .001$).

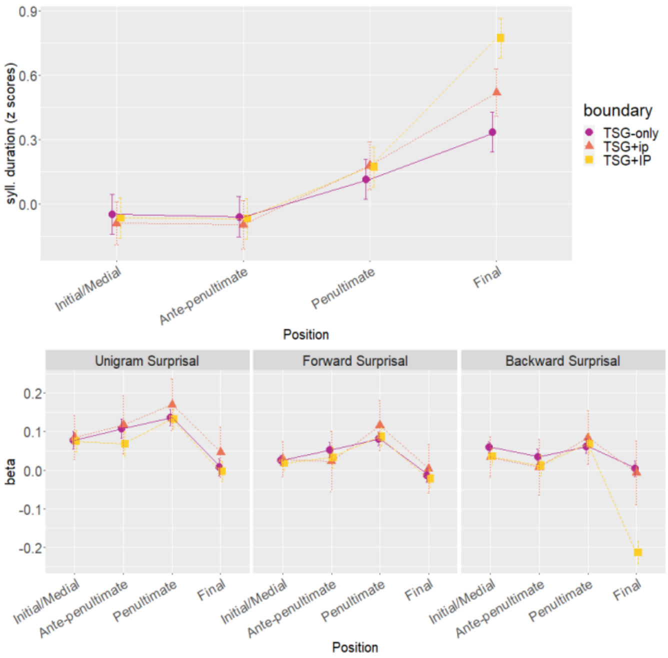


Figure 1. Upper panel: syllable duration as a function of position (x axis) & boundary type (color/shape). Lower panel: effect of surprisal as a function of position (x axis) & boundary type (color/shape)

As for the effect of surprisal as a function of prosodic conditions, we found that unigram surprisal was significant at all conditions except for the final position across three boundary types (p < .05). Forward bigram surprisal was significant for all but the final position for TSG-only breaks (p < .001) and only significant for the penultimate position TSG+ip and TSG+IP (p < 001). Finally, backward bigram surprisal had a positive effect for TSG-only breaks at all positions except for the final (p < .01), for TSG+IP breaks at initial/medial and penultimate positions (p < .01), and notably, a negative effect at the final position of TSG+IP breaks (p < .0001), which potentially suggest syllables in words with low likelihood at the utterance-final position were lengthened.

To sum up, TSG breaks overlapping with different levels of prosodic breaks share similar preboundary lengthening in terms of the domain, with the presence of a larger prosodic mainly resulting in a larger size of final lengthening. As for the interaction with surprisal effects, the positive correlation between high surprisal and longer duration is found to be neutralized at the final position, which is consistent with the view that prosodic marking of boundaries may constrain predictability effects. The neutralizing trend was also found to be consistent regardless of boundary strength other than between backward surprisal and TSG+IP boundaries.

References: [1] Lin, J.-W. 1994. Lexical government and tone group formation in Xiamen Chinese. *Phonology*. [2] Wang, S.-F. & Fon J. 2013. A Taiwan Southern Min spontaneous speech corpus for discourse prosody. *Tools and Resources for the Analysis of Speech Prosody*. [3] Wang, S.-F. & Fon J. 2015. Syllable duration and discourse organization at intonational phrase boundaries in Taiwan Southern Min. *ICPhS*. [4] Jurafsky, D., Bell, A., Gregory, M. & Raymond, W. D.. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. *Typological Studies in Language*. [5] Seyfarth, S. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*. [6] Tang, Kevin & Jason A Shaw. 2021. Prosody leaks into the memories of words. *Cognition*. [7] Aylett, M. & A. Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*. [8] Turk, A. 2010. Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *LabPhon*. [9] Stolcke, A., Zheng J., Wang W. & Abrash A. 2011. SRILM at sixteen: Update and outlook. *IEEE automatic speech recognition and understanding workshop*. [10] Chen, S. F. & Goodman, J. 1999. An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*. [11] Iunn, U.-G. 2005. Taiwanese corpus collection and corpus based syllable/word frequency counts for written Taiwanese. *ROC NSC 93-2213-E-122-001*.

Allotonic variants do not prime each other: evidence from long-lag priming

Stephen Politzer-Ahles[1], Yixin Cui[2]

[1]*University of Kansas (USA),* [2]*Hong Kong Polytechnic University (Hong Kong)*
sjpa@ku.edu, yixin.cui@connect.polyu.hk

Linguistic forms, like words and morphemes, may be realized in different ways depending on the context in which they are produced. For example, in Mandarin, a tonal language, syllables that canonically have Low tone (also called "tone 3") are instead pronounced with Rising tone (also called "tone 2") in certain phonological contexts. Do listeners use their knowledge of this phonological alternation during word recognition? In other words, do listeners hearing a Low-tone syllable in Mandarin also activate its Rising-tone variant, and vice versa?

Politzer-Ahles and colleagues [1] addressed this question using long-lag priming. In the long-lag priming paradigm, participants typically respond to a word faster if they have seen the same word, or a morphologically-related word, earlier in the experiment; for example, people can respond to *hunt* faster if they have seen *hunt* or *hunter* earlier than if they have not. Politzer-Ahles and colleagues [1] had Mandarin listeners make lexical decisions to auditorily-presented one-syllable Mandarin morphemes, like *shi*^{L} (使, meaning to make or let [someone to do something]), which were preceded earlier in the experiment by either identical primes (*shi*^{L}), allotonic primes (e.g. *shi*^{R}, the Rising-tone variant of *shi*^{L}), or unrelated primes (e.g. *hua*^{F}, a syllable with completely different segments and with Falling tone). They found priming for the identical condition, but not the allotonic condition, suggesting that hearing one form (e.g. *shi*^{L}) does not cause comprehenders to activate its allotonic variants (e.g. *shi*^{R}).

A limitation of that study, however, is that Mandarin syllables tend to be highly homophonous: for example, *shi*^{L} is the pronunciation of many different morphemes (such as 史 "history", 驶 "drive", and 屎 "poop"). Politzer-Ahles and colleagues [1] speculate that maybe the reason they didn't observe priming between allotonic variants is because such priming depends on the comprehender being able to uniquely identify a particular morpheme to activate. In other words, maybe *shi*^{L} doesn't activate its variant *shi*^{R} if the listener doesn't know which morpheme to activate.

The present study addresses that limitation by testing participants in a paradigm in which each prime unambiguously corresponds to one morpheme. We conducted a long-lag priming experiment with the same conditions as Politzer-Ahles et al. [1]: targets (like *shi*^{L}) were preceded earlier in the experiment by either an identical prime (*shi*^{L}) or an allotonic prime (*shi*^{R}); targets without either an identical or allotonic prime were unprimed. Unlike that experiment, though, we presented the primes visually, in Chinese characters, and participants' task was to name each character aloud. The experiment thus consisted of two blocks: a "priming" block in which participants saw Chinese characters and named them, and a later "target" block in which they performed lexical decisions to auditorily-presented targets. Thus, the identical prime for the target *shi*^{L} was the character 使 (pronounced *shi*^{L}) and the allotonic prime for that same target was the character 十 (pronounced *shi*^{R}). We avoided using any characters that have multiple pronunciations. The experiment consisted of 71 critical items, divided across three lists in a Latin square design. Both the prime and target blocks included additional filler/distractor trials consisting of words with other tones to help mask the purpose of the experiment, and the target block also included nonword foils.

60 native Mandarin speakers participated in the experiment. Priming effects for the identity and allotone conditions (relative to the unprimed baseline) are shown in Figure 1, and the effects were statistically evaluated using a Bayesian mixed model implemented in the {brm} package of R. As suggested by the figure, there was significant identity priming (22 ms; Bayesian 95% credible interval: [6, 38]) but no significant allotone priming (1 ms; Bayesian 95% credible interval: [-17, 15]).

These results provide a further confirmation that there is not long-lag priming between phonologically related variants of the same morpheme in Mandarin (e.g., *shi*^{L} does not prime *shi*^{R}), while ruling out the possibility that previous experiments' failure to find this priming was due to participants' not being able to uniquely identify one meaning from a homophone. The use of written primes in the present study should guarantee that participants can uniquely identify a morpheme for activation during the priming phase of the experiment, but this still did not facilitate auditory lexical decisions during the target phase.

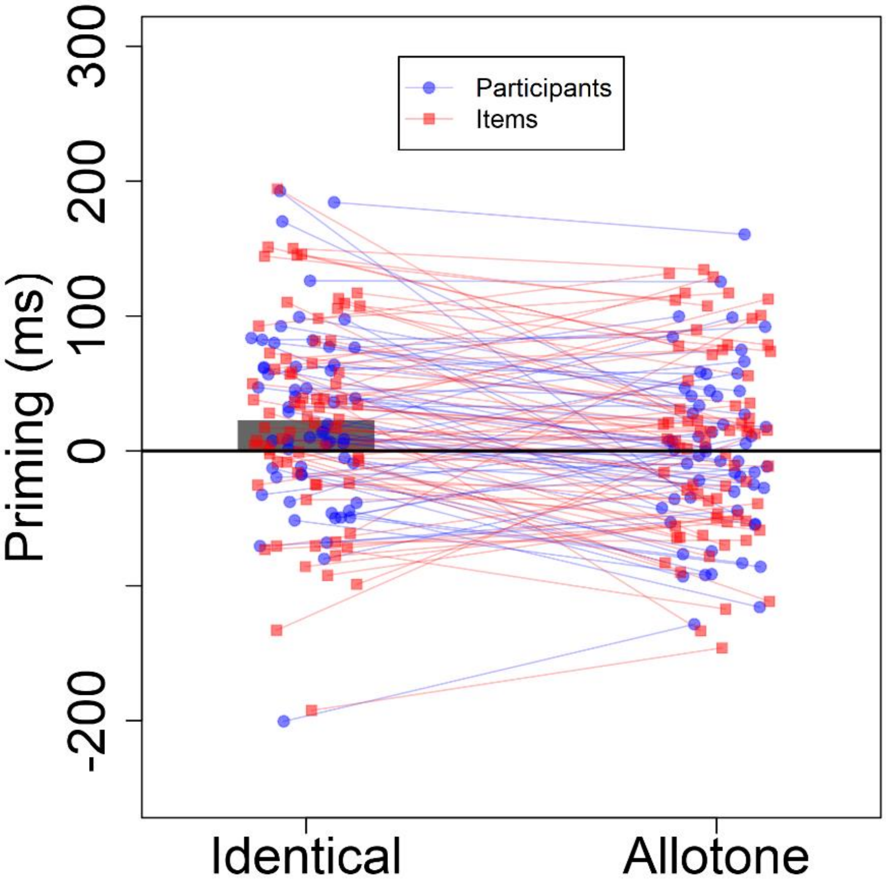


Fig.1 Priming effects for identity priming (left) and allotone priming (right), each expressed as the difference between the unrelated condition and the corresponding related condition. Blue circles represent priming effects for individual participants, red squares represent priming effects for individual items, and the gray bar represents the average priming effect.

References

[1] Politzer-Ahles, S., Pan, L., Lin, J., & Lee, K. (in press). Long-lag identity priming in the absence of long-lag morphological priming: evidence from Mandarin tone alternation. *Glossa: Psycholinguistics*.

Exemplar Effects in the Perception and Production of Advanced and Intermediate L2 Korean Wh-Question Intonation

Bonnie J. Fox[1]

[1]*University of Hawaiʻi at Mānoa University (USA)*
foxbonni@hawaii.edu

This research study takes a hybrid exemplar model (Pierrehumbert, 2002; 2016; Tenpenny, 1995) of L2 intonation and test its predictions on Korean L2 advanced and intermediate proficiency users' perception and production of a three-way distinction of the intonation of Korean wh-phrases. Thus, this experiment seeks to answer the following research questions: 1) How accurately do high and low proficiency L2 Korean users interpret three types of wh-phrases? 2) What prosodic cues do high and low proficiency L2 users utilize to interpret three types of wh-phrases? 3) How accurately do high and low proficiency L2 Korean users produce three types of wh-phrases? 4) What prosodic features do high and low proficiency L2 users utilize to produce three types of wh-phrases?

In Korean, wh-phrases are ambiguous because wh-phrases can be interpreted both as wh-question words and as indefinite pronouns. This leads to a three-way ambiguity phenomenon where a phrase such as "nwuka hakkyo-ey kasse-yo" (lit. 'who school-to go-past-polite') can be interpreted as a wh-question ('Who went to school?') with intermediate phrase focus on the wh-question word, a yes/no-question ('Did someone go to school?') with no focus but generally ending with a rising tone, or a statement ('Someone went to school.') with no distinct focus and generally no rising boundary tone. Prior work looking at L2 Korean acquisition of wh-phrases is sparse. Results from Choi (2009) and Gil, Marsden & Park (2020) looking at L2 perception of Korean wh-phrases indicate identification of the correct interpretation of the three-way distinction is uniquely difficult for most L2 speakers of Korean. While wh-question interpretation was comparable to natives for advanced speakers, they were misinterpreted as declaratives by high-intermediate speakers, and yes-no wh-phrases was significantly worse than natives at all levels, but almost unidentifiable for below advanced speakers, often being confused for wh-questions. Results for declarative interpretation were not reported. Jun & Oh (2000) conducted a study on the acquisition of L2 Korean question intonation of 4 speakers not specifically focusing on the three-way distinction and determined prosodic grouping improved with proficiency level with less inter-sentential Intonational Phrases (IPs), but surface tone realization of Accentual Phrases (APs) and long boundary tones remained difficult for all L2 speakers.

Based on the foundations of a hybrid exemplar model (Pierrehumbert, 2002; 2016; Tenpenny, 1995) we can expect that the amount of input L2 users have had with wh-phrase intonation to play a significant role in the acquisition of the three-way distinction. Lower proficiency users will have had little exposure to the phenomenon at all and are therefore more likely to substitute in either native intonation patterns or the more familiar Korean wh-question intonation patterns to interpret yes-no and declarative wh-phrases. Higher proficiency users, while having more exposure to the three types, will still have had significantly more exposure to the wh-question interpretations followed by yes-no patterns, leading to a predicted reliance on those wh-question patterns and a dearth of success in the declarative case.

A total of 30 participants (10 L2 advanced and intermediate proficiency speakers and 10 native Korean speakers) were asked to complete a perception task (selecting an appropriate continuation for a given utterance), and a production task (reading wh-phrases from a conversation with surrounding contexts). Experimental materials were hosted on Gorilla Experiment Builder (www.gorilla.sc) (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2018), allowing for remote data collection. After completing a headset check (Woods et al., 2017), participants completed a simple decision task where they listened to target utterances recorded by three Korean native speakers ambiguous between a statement (12 stimuli), a wh-question (12 stimuli), and a yes-no question (12 stimuli) along with filler stimuli, and then choose the appropriate continuing response as fast as they could. Then, in a simple elicitation task, participants were asked to read a series of randomized simple conversational texts aloud, playing both roles. This task has a total of 5 critical conversations with 3 target wh-phrases per conversation. Finally participants were asked to complete a short oral interview similar to the ACTFL OPIc (Oral Proficiency Interview – computer) to determine their speaking proficiency level.

In the perception task, the advanced group correctly identified over 80% of all wh-phrases. Wh-questions were the most correctly identified, while the yes-no questions and the statements were equally misidentified. The intermediate group identified only 53% of all wh-phrases correctly, but accurately identified most of the wh-questions correctly but misidentified the majority of yes-no and statements as wh-questions. The differences between the two L2 groups here indicate that the amount of time spent in the L2 provides strong advantages to the acquisition of less common intonation patterns. Additionally, comparing the Reaction Times between the groups, while it may be expected that faster RTs should coincide with higher accuracy, this was not the case, and the intermediate group was faster. This can be explained if we understand that the intermediate group is simply unaware of or unable to process the multiple ways of interpreting the phrases based on intonation patterns. They were fast, but inaccurate in their responses, while the advanced group was slower but more accurate.

In the production task, across all four conversations, the wh-question showed the most similar pitch tracks to the native speakers, with focus marking seen on the wh-question word each time and followed by a rising boundary tone. For the yes/no questions, only the intermediate speakers tend to misattribute focus marking on the wh-phrase instead of the subsequent verb. However, even advanced speakers showed difficulty in producing fully accurate intonation patterns especially when combined with a higher processing load from more difficult combinations of segmentals. When looking at the production of wh-statements, all participants unexpectedly showed a similar tune production to the native speaker for each conversation. It is interesting that the production of this distinction seemed more accurate from the lower-level speaker than their perception, indicating that perhaps it might be so that production precedes perception in the case of acquiring this distinction accurately.

References

[1] Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.) Laboratory Phonology, 6. Mouton de Gruyter.

[2] Pierrehumbert, J. B. (2016). Phonological representation: beyond abstract versus episodic. Annual Review of Linguistics, 2, 33–52.

[3] Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. Psychonomic Bulletin & Review, 2(3), 339-363.

[4] Choi, M. H. (2009). The Acquisition of wh-in-situ constructions in second language acquisition. [Doctoral dissertation, Georgetown University].

[5] Gil, K.-h., Marsden, H., & Park, S.-y. (2020, September 7-20). Beyond L1-L2 morphological similarities: L2 Korean WH lexical ambiguity [Paper Presentation]. 28th Japanese/Korean Linguistics Conference, Lancaster, UK.

[6] Jun, S.-a., & Oh, M. (2000). Acquisition of 2nd language intonation. Proceedings of International Conference on Spoken Language Processing, 4, 76-79.

Quantifying Phonetic Informativity: An Information Theoretic Approach

James Whang

Seoul National University (Korea)
jamesw@snu.ac.kr

For successful speech perception, a listener must assign higher weights to phonetic cues that are crucial for categorical contrasts than to non-crucial ones. The problem, however, is that phonetic cues are often redundant in natural speech [1], making weighting a non-trivial task. It is expected then that there are individual differences in how specific cues are weighted which in turn can lead to language change over time. Indeed such diachronic 'mis'-weighting is considered a major driver of language change [2]. In information theoretic [3] terms, redundant cues have low informativity. Research in segmental [4] and lexical [5] domains show that it is precisely such low information units that are targeted for reduction. The current study presents the results of a working model that extends this information theoretic approach to the subsegmental level.

The phonetic informativity quantification model presented here uses surprisal to quantify redundancy and entropy to quantify overall information of a given feature x.

Surprisal: $$-\log_2 \Pr(x|Context) \quad (1)$$
Entropy (H): $$\sum \Pr(x|Context) * -\log_2 \Pr(x|Context) \quad (2)$$

Given the formulation above, zero surprisal for x would mean that it is completely redundant in the context of one or more other cues. Zero entropy would mean that x is completely uninformative and has no function in the language. Both surprisal and entropy have no theoretical upper bound, but the higher the value, the higher the functional load in the language.

As a test case, the current study focuses on the informativity of vowel cues in Japanese. Vowel features are used as stand-ins for acoustic cues for simplicity's sake in the current study. All calculations are based on CSJ-RDB, a 500K-word subset of the Corpus of Spontaneous Japanese [6]. The corpus contains a total of 679,123 vowels. The model first converts each vowel to a user-defined, n-sized set of phonetic features, with each feature's frequency being equal to the sum of the vowel frequencies that contained the feature. Second, surprisal (redundancy) is calculated for each feature with all possible subsets of the remaining features as context, resulting in 2^{n-1} contexts. For example, given a vowel with feature set $A = \{a, b, c, d\}$, surprisal for feature a is calculated with all possible subsets of $A\backslash a$ as context (including {} and $\{b, c, d\}$), then again for b with all possible subsets of $A\backslash b$, etc. Lastly, entropy (informativity) is calculated based on the surprisal values.

For the current analysis, the set consisted of the following eight features with their respective parameters shown in []:

V → {**height** [*high, mid, low*], **backness** [*front, central, back*], **roundedness** [*rounded, unrounded*], **length** [*short, long*], **peripherality** [*centralized, peripheral*], Δ**height** [*level, rising, falling*], Δ**backness** [*stable, fronting, backing*], Δ**roundedess** [*constant, rounding, unrounding*]}

Although the CSJ-RDB annotations only contained short and long monophthongs (/i, e, a, o, u, ii, ee, aa, oo, uu/), the three delta features were included under the assumption that the set of possible features must be the same across all languages. All Japanese vowels, therefore, were {…, *level, stable, constant*} but this would not be case for languages that have vowels with significant movement (e.g., English /aʊ/ → {…, *rising, backing, rounding*}).

Three key results are reported here in the interests of space. First as expected all delta features were completely redundant in Japanese, with zero surprisal in all contexts and consequently zero entropy. The lack of informativity for these features suggests that Japanese listeners are insensitive to vowel movement. Second, the feature [*high*] had the highest entropy (97.80), suggesting that

Japanese listeners have heightened sensitivity to high vowels. Lastly, [*long*] had the second highest entropy (88.58). In conjunction with [*peripheral*] having zero entropy, the results predict that Japanese listeners should be more sensitive to length manipulations than peripherality manipulations. All three predictions are supported by previous experimental studies [7,8,9].

The current study presents a simple working model for quantifying the informativity of phonetic cues/features in a given language. The use of Information Theory places the model in a long line of research that sought to understand gradient predictability effects in phonetics and phonology [4,5,8,10]. It also shows the theory's incredible flexibility to handle linguistic representations of varying sizes and granularity. Future work includes further developing the model to process not just simple sets of features but also sequences of cue/feature vectors to quantify segment-internal timing relations [11], as well as extending the model to fine-tune cross-linguistic perception frameworks [12,13], which typically rely on cue-weighting differences to explain perceptual errors.

References

[1] Clements, G. N. (2009). The role of features in phonological inventories. *Contemporary views on architecture and representations in phonological theory*, 19-68.

[2] Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

[3] Shannon, C. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

[4] Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6(2). 243–278.

[5] Hall, K. C., Hume, E., Jaeger, F. T., & Wedel, A. (2016). The message shapes phonology. Ms. UBC, University of Canterbury, University of Rochester and University of Arizona.

[6] Maekawa, K., and Kikuchi, H. (2005). Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In *Voicing in Japanese*, ed. Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara. Mouton de Gruyter.

[7] Strange. W, Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Soceity of America, 109,* 1691-1704.

[8] Whang, J. (2019). Effects of phonotactic predictability on sensitivity to phonetic detail. *Laboratory Phonology*, *10*(1):8.

[9] Strange, W., Hisagi, M., Akahane-Yamada, R., Kubo, R. (2011). Cross-language perceptual similarity predicts categorial discrimination of American vowels by naïve Japanese listeners. *The Journal of the Acoustical Society of America* 130: EL226–31.

[10] Cherry, E. C., Halle, M., & Jakobson, R. (1953). Toward the logical description of languages in their phonemic aspect. *Language*, 34-46.

[11] Inkelas, S., & Shih, S. S. (2017, May). Looking into segments. In *Proceedings of the Annual Meetings on Phonology* (Vol. 4).

[12] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 233–77.

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

Perception of Korean Coda Consonants /p, t, k/ by Chinese Learners of Korean in Taiwan

Shu-Wei Yang[1], Jung-Yueh Tu[2]

[1]*National Chengchi University(Taiwan),* [2]*National Chengchi University(Taiwan)*
weiphone@nccu.edu.tw, jytu@nccu.edu.tw

A previous study has investigated the production patterns of Korean obstruent coda consonants by Chinese learners of Korean in Taiwan and found that the "supposedly unreleased" Korean obstruent coda consonants were mostly pronounced with a release by Chinese learners [1]. To further explore the perceptual confusion of Korean obstruents /p, t, k/ in the coda position among Chinese L2 learners of Korean, this study examines the effects of preceding vowel contexts and learning experience on the perception of obstruent coda consonants. It is hypothesized that both preceding vowel contexts and learning experience will affect Chinese learners' perception of Korean obstruent coda consonants [2, 3, 4].

In this study, a perception experiment was conducted with Chinese learners of Korean, who were divided into two groups based on their learning experience: beginners and intermediates. There were 30 participants (9 males, 21 females; mean age: approximately 20; age range: 19-22) in this study, 17 at the beginning level and 13 at the intermediate level. The participants completed an identification task, in which they were presented with stimuli containing three Korean obstruent coda consonants /p, t, k/ and one filler /l/ with three preceding vowels /a, i, u/. The stimuli were recorded by two Seoul Korean speakers, one male and one female. This study used a Kruskal-Wallis one-way analysis of variance to analyze the correctness rate of each coda consonant and each coda consonant with different preceding vowels, and a Wilcoxon rank sum test to analyze the effect of learning experience.

The results showed that Chinese learners perceived /p/ better than /t/ (p<.001), and /t/ better than /k/ (p<.001) (as shown in Figure 1). The relatively high correctness rate of /p/ perception is in line with the high accuracy of /p/ production found in the previous study [1].

In addition, the effect of the preceding vowel context differed between /p/, /t/, and /k/ (see Figure 2). For the preceding vowel /a/, there was no statistically significant difference in the correctness rate of /ap/ and /at/ (p=.71), while both /ap/ and /at/ had higher correctness rates than /ak/ (p<.001). For the preceding vowel /i/, the correctness rate of /ip/ was higher than /it/ (p<.001) and /it/ was higher than /ik/ (p<.05). For the preceding vowel /u/, there was no statistically significant difference in the correctness rate of /up/ and /ut/ (p=.08), while both /up/ and /ut/ had higher correctness rates than /uk/ (p<.001). That is, the preceding vowel /i/ made it easier to perceive /p/ compared to the other preceding vowels. The pattern of errors in the perception of Korean coda consonants was also related to the preceding vowels. The coda consonant /p/ tended to be perceived as /t/ when preceded by a low vowel (/a/), and as /k/ and /t/ when preceded by a high vowel (/u, i/). The coda consonant /t/ was mostly misperceived as /k/ when preceded by a back vowel (/a, u/), and as /p/ when preceded by a front vowel (/i/), similar to the production error shown in the previous study [1]. The coda consonant /k/ was misperceived as /p/ when preceded by a back vowel (/a, u/), and as /t/ when preceded by a front vowel (/i/). This suggests that the perception of the coda /p/ is influenced by the height of the preceding vowel, whereas the perception of the coda /t/ and /k/ is influenced by the backness of the preceding vowel.

Finally, the learning experience did not show any effect as there was no significant difference between the correctness rates of beginning-level learners and intermediate-level learners (shown in Figure 3). (/p/: p=.59; /t/: p=.26; /k/: p=.59). The current findings did not show the effect of learning experience, as can be probably attributed to several factors: (1) L1 background: Chinese has only two coda consonants — /n/ and /ŋ/. The lack of /p, t, k/ in coda position in Chinese may make it not easy to perceive those in Korean [5]. (2) Insufficient inputs: the quantity and quality of L2 inputs greatly influence L2 speech learning. The participants in this study learned Korean with a non-Korean teacher and in a non-Korean language environment [6]. (3) The relative unimportance of these coda consonants in perception: since people mostly perceive words through contexts, and the contexts will certainly help L2 learners to understand the meaning of the given sentence. That is, whether L2 learners correctly perceive Korean coda consonants may not greatly affect the understanding of the whole sentence. However, this study did not investigate the influence of Taiwanese Southern Min (widely spoken in Taiwan), which has /p/, /t/, /k/ coda consonants. Whether the Taiwanese Southern Min has an impact on the perception of Korean coda consonants needs further research.

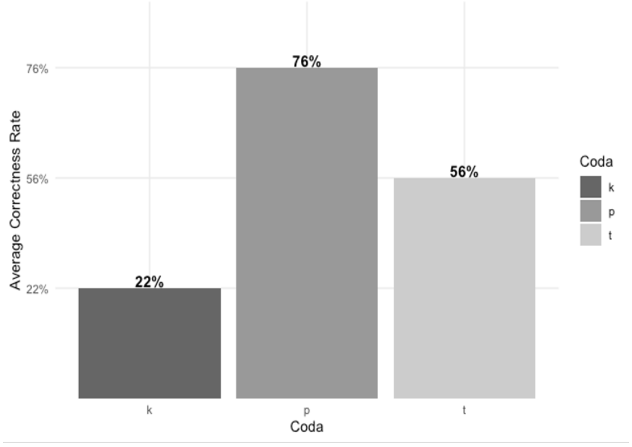


Fig. 1 Average correctness rate of Korean coda consonant /p, t, k/

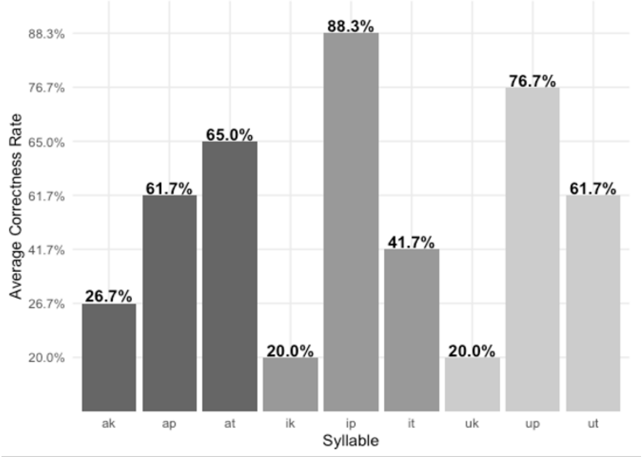


Fig. 2 Average correctness rate of Korean coda consonant /p, t, k/ with different preceding vowels

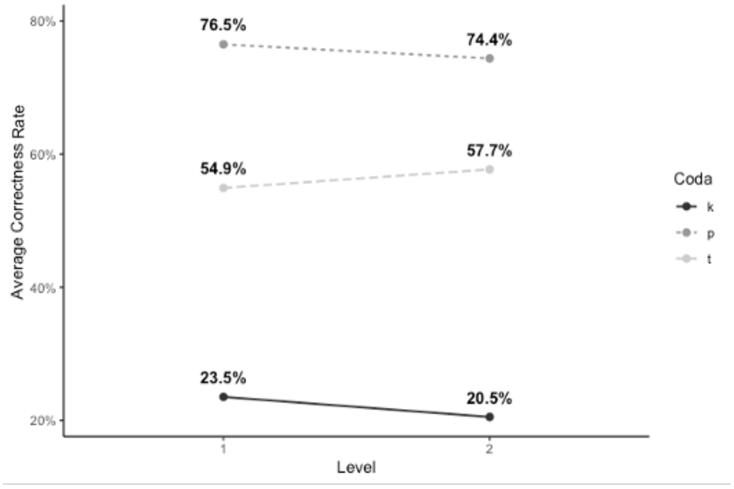


Fig. 3 Average correctness rate of Korean coda consonant /p, t, k/ by different levels of Korean proficiency
(1: Beginning; 2: Intermediate)

References

[1] Chang, R. R. (2015). A Study on Korean Pronunciation Teaching for Taiwan Learners-focused on Final Consonant-, MA dissertation, Keimyung University.

[2] Oh, M. R. (2002). Place Perception in Korean Consonants, *Speech Sciences*, 9(4), 131-142.

[3] Cardoso, W. (2011). The development of coda Perception in Second, language phonology: A variationist perspective. *Second Language Research*, *27*(4), 433–465.

[4] Kim, J. Y. (2016). Perception of Korean coda consonants by Chinese learners of Korean: A one-year longitudinal study. Phonetics and Speech Sciences, 8(4), 79-87.

[5] Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, O.-S. & Munro, M. J. (eds.), Language experience in second language speech learning: In honor of James Emil Flege, 31-52. Amsterdam: John Benjamins Publishing Company.

[6] Flege, J., & Bohn, O. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), Second Language Speech Learning: Theoretical and Empirical Progress, 3-83. Cambridge: Cambridge University Press.

[7] This paper is a revised and supplemented content published by the "2022 Learning Site for Foreign Researchers of Korean Language Project" conducted at the service cost of the National Institute of Korean Language.

Detecting the Accentual Phrase boundaries in Seoul Korean using tonal and segmental cues

Seung Suk Lee

University of Massachusetts Amherst (USA)
seungsuklee@umass.edu

The most widely adopted model of Seoul Korean intonational phonology (K-ToBI, [1, 2, 3]), proposes that Accentual Phrases (APs)–prosodic constituents larger than Phonological Words (PWds)–can be identified via characteristic tonal events at their junctures [1]. It also proposes that APs are demarcated by a particular kind of perceived juncture, which is transcribed with the Break Index (BI) level of 2 in the model [2]. For example, (1) can be interpreted as (1a) or (1b) depending on where this juncture is located. However, the parsing of APs via non-tonal BI cues has been less emphasized in the literature compared to the tonal marking [4]. Moreover, less is known about what acoustic properties, tonal or non-tonal, might underlie the percept of AP juncture.

Experimental work has shown that listeners are sensitive to the tonal fall from the AP final H tone and the AP initial L tone and use it as a cue for segmentation [5, 6, 7]. APs can also start with an H tone if they start with an Aspirated or Fortis obstruent but in this paper, we focus exclusively on the L-initial APs. The fall associated with the AP-level juncture was shown to be larger and steeper than the fall that may happen within the AP [7, 8]. In (1a), the tonal fall from [ɾim] to [man] (across AP boundaries) should be larger than the fall from [gu] and [tʃʰʌn] (within the same AP), for example. Previous work has also found the AP serves as the domain of segmental allophony of Lenis obstruents: Lenis is voiceless AP-initially but optionally voiced AP-medially [1], which indicates that it can signal whether it is in the AP left edge. The /k/ (in bold) in /ku/ is voiceless when the syllable is AP initial in (1b) but voiced in (1a). Experimental work showed that Korean listeners were sensitive to the allophonic realization of Lenis in the absence of tonal cues [9]. In other words, the AP juncture can be instantiated both tonally and non-tonally and listeners are sensitive to the tonal change between syllables and the segmental details of Lenis in finding prosodic constituents, we don't know how these cues are distributed and how robustly they separate the AP initial syllables ([$_{\text{AP}}\sigma$) from the AP non-initial syllables ([$_{\text{AP}}\ldots\sigma$) in spontaneous speech data.

In this paper, I investigated how PWd initial syllables ([$_{\text{PWd}}\sigma$) and PWd non-initial syllables ([$_{\text{PWd}}\ldots\sigma$) are distributed in the acoustic cue space defined by the tonal and segmental cues, in two speech corpora of Seoul Korean. The first corpus was taken from [10] and it was transcribed in K-ToBI, but it did not have enough Lenis tokens to investigate segmental cues. The second corpus was taken from a spontaneous speech corpus [11] that was larger in size, but it was not transcribed in K-ToBI. Two (one teenage female and one male in his forties) out of forty speakers from the second corpus were investigated as a first pilot. The tonal cue was parameterized as the maximal change in F0 from the previous syllable, normalized as the fraction of range for each utterance. This value was negative when the tonal change was falling from previous syllable. The finding that the tonal fall is larger at the AP juncture than within the AP suggests that the distribution for [$_{\text{PWd}}\sigma$ would be negatively larger than the distribution for [$_{\text{PWd}}\ldots\sigma$ because only [$_{\text{PWd}}\sigma$ can be AP initial under the strict layer hypothesis [12]. The segmental cue was parameterized by combining three common lenition measurements via Principal Components Analysis [13]: percentage of voiced interval [14, 15], the difference between the maximum and minimum rate of change in intensity [16], and the speech-rate-normalized closure duration [9]. It was expected that [$_{\text{PWd}}\sigma$ would have a larger value for this cue compared to [$_{\text{PWd}}\ldots\sigma$, for the same reason as above, since only [$_{\text{PWd}}\sigma$ can be AP initial. The separability of the cue was evaluated by measuring the overlapping area between the kernel density estimate curve of [$_{\text{PWd}}\sigma$ and [$_{\text{PWd}}\ldots\sigma$ which were normalized independently (the hatched areas in figures). The fact that the second corpus is not transcribed gives us a good testing ground whether the acoustic cues can be used to separate [$_{\text{PWd}}\sigma$ from [$_{\text{PWd}}\ldots\sigma$.

The tonal cue separated [$_{\text{PWd}}\sigma$ from [$_{\text{PWd}}\ldots\sigma$ in the expected way in the first dataset (Fig 2): the distribution of [$_{\text{PWd}}\sigma$ was further left compared to that of [$_{\text{PWd}}\ldots\sigma$. This was confirmed in Fig 1, which shows the distribution of [$_{\text{AP}}\sigma$ and [$_{\text{AP}}\ldots\sigma$, which were transcribed in [10]. The distributions were remarkably similar between the two figures since 69% of the PWd formed an AP on their own. However, this was not replicated in the second dataset (Fig 3) which might be since it was not the case in the larger dataset that most [PWdσ were also [APσ. [3] showed that contrastive focusing can cause dephrasing which would increase the proportion of [PWdσ that are not [APσ. It also indicated that the tonal cue is realized with extreme variation and overlap between the categories in a spontaneous speech [4]. On the other hand, [PWdσ was better separated from [PWd…σ on the segmental cue (Fig 4), which shows that the Lenis onsets in [PWdσ were more often realized with larger values than the Lenis onsets in [PWd…σ. This study showcased the concern addressed in [4] that while the prosodic constituents are often defined and claimed to be segmented with respect to their tonal markings, it is not entirely straightforward how this can be done, especially in a larger corpus containing more realistic speech. It also showed that the segmental cues such as the one presented here measuring the allophonic variation of Lenis might be more robust in separating the prosodic categories. Further studies are needed to explore what other robust segmental or tonal cues exist that might allow better detection of AP boundaries.

(1) /pʰa.ɾan.kɨ.ɾim.man.**k**u.tʃʰʌn.wʌn.i.ja/ - parentheses indicate PWd boundaries
a. [(pʰa.ɾan) (gɨ.ɾim)AP] (BI: 2) [AP(man) (**g**u.tʃʰʌn) (wʌn i.ja)]
Gloss: blue painting 19000 *won* be
Translation: The blue painting is 19000 *won* (about $15).
b. [(pʰa.ɾan) (gɨ.ɾim - man)AP] (BI: 2) [AP(**k**u.tʃʰʌn) (wʌn i.ja)]
Gloss: blue painting - only 9000 *won* be
Translation: Only the blue painting (but not other paintings) is 9000 *won* (about $7).

Blue = [AP, Gray = [AP…σ Blue = [PWdσ, Gray = [PWd…σ

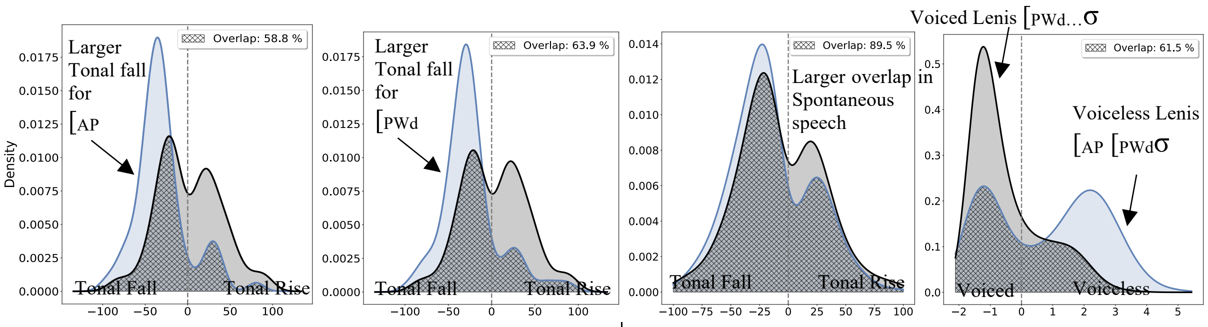


Fig. 1 Fig. 2 Fig. 3 Fig. 4 Lenis PCA Cue

Dataset 1 Dataset 2

Tonal cue: Maximum difference in F0 from the previous syllable in percent First Component of PCA

References

[1] Jun, S. A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, *15*(2), 189-226.

[2] Jun, S. A. (2000). K-ToBI (Korean ToBI) labelling conventions. *Speech Sciences*, 7(1), 143-170.

[3] Jun, S. A. (2007). The intermediate phrase in Korean: Evidence from sentence processing. *Tones and tunes: Experimental studies in word and sentence prosody*, 143-170.

[4] Jun, S. A. (2022). The ToBI transcription system: Conventions, strengths and challenges. *Prosodic theory and practice*, 151-181.

[5] Kim, S., & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *The Journal of the Acoustical Society of America*, *125*(5), 3373-3386.

[6] Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Studies in Second Language Acquisition*, *34*(3), 415-444.

[7] Tremblay, A., Cho, T., Kim, S., & Shin, S. (2019). Phonetic and phonological effects of tonal information in the segmentation of Korean speech: An artificial-language segmentation study. *Applied Psycholinguistics*, *40*(5), 1221-1240.

[8] Lee, J. Y., & Lee, H. Y. (2013). Korean Intonation Patterns from the Viewpoint of F 0 Percentage Change. *Phonetics and Speech Sciences*, *5*(1), 123-130.

[9] Yoo, Kayeon (2020). *The production and perception of domain-initial strengthening in Seoul, Busan, and Ulsan Korean*. Ph.D. Thesis, University of Cambridge.

[10] Jun, S. A., Lee, S. H., Kim, K., & Lee, Y. J. (2000). Labeler agreement in transcribing Korean intonation with K-ToBI. In *INTERSPEECH* (pp. 211-214).

[11] Yun, Weonhee, Kyuchul Yoon, Sunwoo Park, Juhee Lee, Sungmoon Cho, Ducksoo Kang, Koonhyuk Byun, Hyeseung Hahn & Jungsun Kim (2015). The Korean corpus of spontaneous speech. *Phonetics and Speech Sciences* 7:2, 103-109.

[12] Selkirk, E. O. (1986). *Phonology and syntax: the relationship between sound and structure*. MIT press.

[13] Dalcher, C. V. (2007). Statistical methods for quantitative analysis of multiple lenition components. *ICPhS*.

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

[15] Davidson, L. (2018). Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association*, *48*(3), 331-356.

[16] Kingston, J. (2008). Lenition. In *3rd Conference on Laboratory Approaches to Spanish Phonology* (pp. 1-31). Cascadilla Proceedings Project.

Is there an optimal window size for a moving window analysis of pitch entrainment?

May Pik Yu Chan[1], Meredith Tamminga[1]

[1]*University of Pennsylvania (USA)*
pikyu@sas.upenn.edu, tamminga@ling.upenn.edu

1 Introduction. Conversational entrainment broadly refers to the adaptation of speech characteristics between interlocutors to become more (or less) similar to one another, which shapes the dynamics of speech variation and may be a mechanism for the propagation of sound change. [1] reviews the ever-growing range of work on this phenomenon and propose a systematic entrainment typology. In their terms, our current study focuses on **static global synchrony**, which refers to whether interlocutors fluctuate together in variability of speech features (here, pitch) “across any time scale greater than adjacent turns” [1].

We adopt a moving window approach to static global synchrony, following e.g. [2, 3]. As [1] points out, past work has measured static global synchrony over windows of widely varying sizes; however, there has been little systematic inquiry on what constitutes an ideal window size for analysis. On one hand, window sizes that are too small may be insensitive to entrainment behaviour that requires time beyond a turn domain to appear. On the other hand, window sizes that are too large may get too close to the overall average of the conversation and wash out effects of observable synchrony. Methodologically, we might therefore wish to find the minimum window size that maximizes consistency in the conclusions drawn. Thus, we systematically compare window sizes to ask whether different methodological decisions would lead to consistent conclusions about which dyads exhibit weaker or stronger pitch synchrony. However, the question of window size optimization is not merely methodological; it also speaks to questions about the true time scale on which entrainment behavior takes place and whether that time scale might vary across individuals, dyads, or interactions.

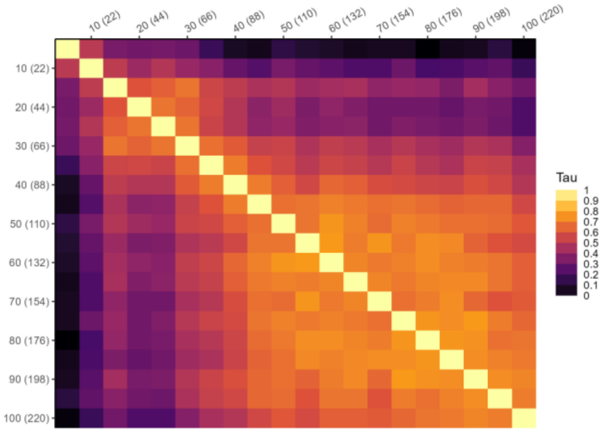
2 Methods. We analyzed recordings from fifty four same-gender (47 female, 7 male) existing friendship pairs (108 participants) from Philadelphia that spoke English as a native language. The pairs of friends engaged in dyadic conversations lasting 30 minutes in a laboratory room without a researcher present. Recordings were transcribed and forced-aligned using FAVE [4]. F0 information was extracted on a frame by frame basis from Praat based on the autocorrelation method.

The HYBRID method as outlined in [2] was adopted, which is an utterance sensitive approach to the time aligned moving average method of synchronic convergence [3]. In this method, the summarized speech characteristic (in our case, median F0) of both speakers within a given window length until the end of the speakers' utterance is measured for each step. We test 20 different settings ranging from a step size spanning 5s to 100s, each 5 seconds apart, with the window length being 2.2 times the length of the step size, which is comparable to [2]’s settings of a window length of 110s, and a step size that spans 50s. In other words, for every step size of a given length (e.g. 5 seconds), we took the median F0 from a series of overlapping windows of a given length (e.g. 22 seconds), but extending to the end of the utterance(s) that began within the window boundary. Here we label step/window settings using the step size in seconds. In this study, utterances were separated by silent portions or interruptions (e.g. laughter, noise) of a speakers' speech as defined by the forced aligner. Upon converting the series of median F0 values for each speaker into semitones with a baseline of 100 Hz, a Spearman's ρ value was calculated for each dyad, for every given window length setting. This was used as a proxy for the overall magnitude of synchrony across the conversation for each dyad.

We then compared the rankings of dyad synchrony (the Spearman's ρ) for each given window length setting with every other step size setting using Kendall's τ. These results tell us whether the rankings of high vs low synchrony dyads are consistent across different window length settings.

3 Results. Results of the matrix of Kendall's τ values are shown in Figure 1. Brighter yellow/orange regions show that the pair of window length settings resulted in similar dyadic rankings of synchrony. In other words, those pairs of window length settings led to more consistent results regarding which dyads synchronized more. Darker purple regions, conversely, indicate that the pair of window lengths resulted in different conclusions about which dyads synchronize most.

Figure 1 shows that smaller window lengths result in dark purple regions on the heatmap, suggesting that dyad synchrony rankings are quite unstable at small window lengths (especially below 25s). After around step size 45s however, the general τ values increase, suggesting a relatively stable ranking between conversations at higher step sizes. For example, when comparing 40s to 100s steps (τ = 0.48), there is less stability in the dyad rankings than comparing 45s to 100s steps (τ = 0.56).



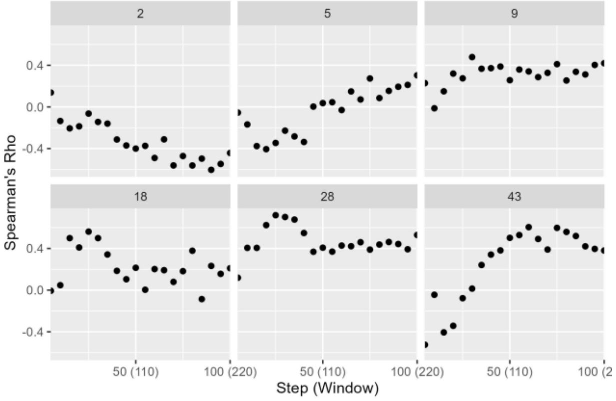


Fig.1 Heatmap visualizing Kendall's τ values of dyad synchrony rankings across different step size / window length settings.

Fig.2 Spearman's ρ value by window length setting for some sample dyads.

4 Discussion and Conclusion. In this study we explored the relationship between different window length settings in an utterance sensitive moving window approach to synchronic conversational entrainment. We found that there is a range of smaller window lengths that lead to relatively unstable rankings between dyads, but once a threshold is passed, the relative rankings of dyads became more comparable at wider window lengths. A theoretical question of why the smaller window lengths have more unstable rankings remains. We outline three possibilities: First, it is possible that the interlocutors are not that fine grained in their levels of entrainment, meaning that the effect of attending to interlocutors' speech characteristics may come with temporal delay or be spread over longer time spans. Therefore, too small of an analysis window may result in too much noise. Second, it may be that there is true micro-temporal entrainment on a dyadic level, but that the optimal window size for detecting this rapid synchrony is contingent on the dyad. In other words, we may be unable to find a one-size-fits-all setting because each pair of speakers entrains to each other on different time windows. This seems possible given Figure 2, which shows the ρ value by window length setting for a sample of dyads. Some dyads appear to have a preferential setting that maximizes ρ, while others have more gradient or jittered relationships between settings. A third possibility is that entrainment takes place not on a time-sensitive window, but rather is contingent on discourse-specific units, such as turns, topics or other conversational events. In other words, entrainment might happen at a window length that is dynamic within the conversation, and cannot be prespecified without attention to the content of the interaction.

Taken together, our results suggest that conclusions drawn about dyadic synchrony are sensitive to analysis methods, meaning researchers need to give careful consideration to reasonable settings based on their dataset. Based on our dataset, we might suggest a window length of 99 seconds spaced at 45 second steps as a practical starting point for exploration. Future work comparing different datasets, methods, and speech features may help validate whether this recommendation could be applied widely or if it is context-sensitive.

References

[1] Wynn, C. J., & Borrie, S. A. (2022). Classifying conversational entrainment of speech behavior: An expanded framework and review. *Journal of Phonetics, 94*, 101173.

[2] Bonin, F., Looze, C.D., Ghosh, S., Gilmartin, E., Vogel, C., Polychroniou, A., Salamin, H., Vinciarelli, A., Campbell, N. (2013). Investigating fine temporal dynamics of prosodic and lexical accommodation. *Proc. Interspeech 2013,* (pp. 539-543). Lyon, France.

[3] Kousidis, S., Dorran, D., Wang, Y., Vaughan, B., Cullen, C., Campbell, D., McDonnell, C. & Coyle, E. (2008). Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues. In *Ninth Annual Conference of the International Speech Communication Association (Interspeech 2008)*, (pp. 1692-1695). Brisbane, Australia.

[4] Rosenfelder, I., Fruehwald, J., Evanini, K., and Yuan, J. (2011). FAVE (Forced Alignment and Vowel Extraction) Program Suite. http://fave.ling.upenn.edu.

Production and Perception of Merging Tones in Dalian Mandarin

Yan Dong[1]

[11]*Dalian University of Technology (China)*
sophiayandong@gmail.com

The topic of sound change has captivated researchers in various subfields of linguistics, including phonetics, phonology, and sociolinguistics, due to their interest in understanding the forces driving the process (Hay et al, 2006; Yu, 2007). The current study focuses on a specific aspect of sound change: tone merger, particularly in its initial stages. In these early stages of tone merger, it is common for speakers to consistently differentiate the merging tones in their speech production, but struggle to distinguish them in perception (Mok and Wong, 2010; Yiu, 2009), or alternatively, to have trouble differentiating the merging tones in production but not in perception (Yiu, 2009).

Recently, there have been proposals of a merger between two falling tones in Dalian Mandarin (Gao, 2007; Liu, 2012). However, there is disagreement over which acoustic features remain distinct in the merging tones and the status of the merger, whether it is ongoing or complete. Previous research on Dalian has not documented the acoustic properties of the production stimuli, leaving the perception cues and the connection between production and perception unknown (Song, 1963). According to a study by Liu (2012), the key distinction between the two merging tones is their length. The research posits that they are perceived as completely merged. However, the study has limitations in its methods, as only a 34-year-old female Dalianese who left the city when she was 23 was used as both the speaker and the listener. This might have impacted her determination of a complete merger. Moreover, other factors, such as word frequency, age and gender of speakers, which may also have an effect on the merging process, has not been investigated. The present study intends to investigate the production and perception of the merging tones and the effects of word frequency, age and gender of speakers in the process.

A production experiment and a perception experiment (a forced-choice character identification test) were carried out with 11 speakers. The results revealed that, in terms of production, the finalf0 and contour shape had a significant impact. Tone4 was found to have a high finalf0 and a concave shape, while Tone1 had a low finalf0 and a convex shape. Additionally, word frequency was found to have a significant effect on production. As for duration, the differences between Tone1 and Tone4 were smaller for high frequency words than for low frequency words when it came to final f0 and concave. Age and gender were also found to have significant effects, with young and female speakers producing less distinct Tone1 and Tone4 tokens compared to older and male speakers.

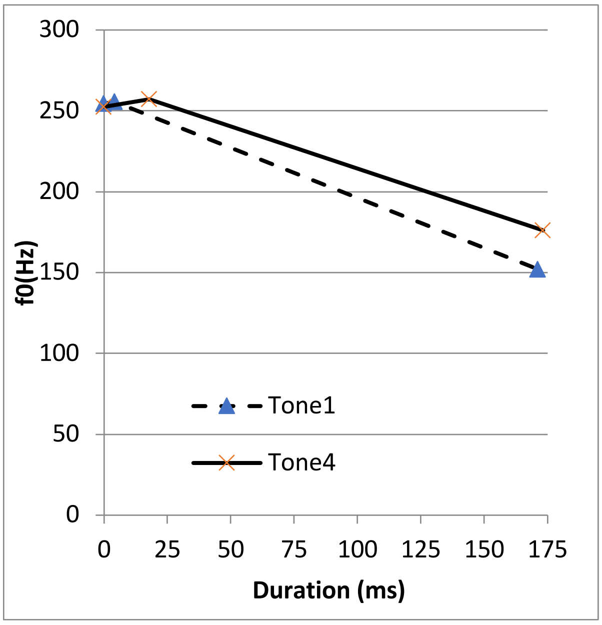
In perception, the accuracy rate across both tones was found to be 77%, which suggests that this is not a complete merger as claimed by Liu (2012). The finalf0 and contour shape acoustic cues identified in the production experiment were the cues used by listeners in perception, as revealed by the generalized linear mixed effect model. Listeners were found to be more accurate when listening to speakers whose production of Tone1 and Tone4 was more distinct, as opposed to those with highly overlapping productions in the two-dimensional space of finalf0 and contour shape. Word frequency was also found to have a significant impact, with listeners perceiving high frequency words more accurately than low frequency words. This may be due to the speaker selection, a different type of frequency such as syllable and tone combination frequency, or the effect of lexical access, which may override the weakness of acoustic cues in high frequency words. There were also variations between speakers and listeners. Some speakers had highly overlapping productions of Tone1 and Tone4, while some listeners had near chance accuracy regardless of the speaker, which indicates that this is a merger in progress.

In sum, the falling tones merger in Dalian is still a work in progress. Finalf0 and contour shape are the cues utilized by both speakers and listeners, and factors such as word frequency, as well as the age and gender of the speakers, have an impact on both production and perception.

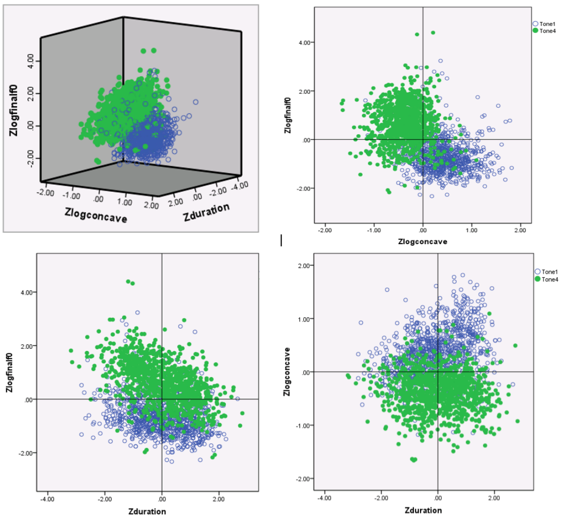
(1) Test material in production and perception experiment

Word frequency group	Word	IPA transcription	Tone	Word frequency ranking	Character frequency ranking	Gloss
High frequency	班	pæn	1	961	884	class
	办	pæn	4	261	367	to handle
	边	pjæn	1	441	316	edge
	变	pjæn	4	264	225	to change
Low frequency	蹉	tshuə	1	NA	5828	error/slip
	锉	tshuə	4	28352	4866	tool for shaping metal
	裆	taŋ	1	19694	4099	crotch of a pair of trousers
	宕	taŋ	4	NA	4231	dissipated

(2) Tone contour for Tone1 and Tone4 pooled across 11 speakers in f0 scale



(3) Zlogconcave and Zduration. (for all plots, blue empty circles= Tone1, green solid circles= Tone4)



References

[1] Hay,J., Warren, P. and Drager,K. (2006). Factors influencing speech perception in the context of a merger- in-progress. *Journal of Phonetics* 34,458-484.

[2] Yu, A. C. L. . (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology*, 24(1), 187-214.

[3] Pik-Ki, P. , Mok, P. , Wai, Y. , & Wong. (2010). Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables. *Speech Prosody* 2010.

[4] YM Yiu. (2009). A preliminary study on the change of rising tones in Hong Kong Cantonese: an experimental study. *Language and Linguistics*, 10(2), 269-291.

[5] Gao Yujuan (2007). *Dalian Fangyan Shengdiao Yanjiu* [A Study on Tones of Dalian Dialect]. Liaoning Normal University Press.

[6] Liu, Te-hsin. (2012). The phonology of incomplete tone merger in Dalian. *Journal of Chinese Linguistics*, 40.2, 362-396.

[7] Song, Xue. (1963). Liaoning Yuyin Shuolue (A sketch of Liaoning Phonology). In *Zhongguo Yuwen* 2:104-114.

Discrimination and Identification of Japanese Quantity Contrasts by Native Cantonese, English, French, and Japanese Listeners

Albert Lee[1], Yasuaki Shinohara[2], Faith Chiu[3] & Tsz Ching Mut[1]

[1]*The Education University of Hong Kong (Hong Kong),* [2]*Waseda University (Japan),* [3]*University of Glasgow (UK)*
albertlee@eduhk.hk, y.shinohara@waseda.jp, faith.chiu@glasgow.ac.uk, stcmut@gmail.com

One line of research in L2 phonological acquisition asks whether the relationship between L1 phonology and acquiring L2 sounds is based on discrete sound categories or on phonetic features. The latter implies that, for example, having a /t/-/d/ contrast in L1 would help learners acquire a new /p/-/b/ contrast in L2, as the primary acoustic cue to both contrasts is voice onset time [1]. However, in answering this question, most previous studies have not achieved to (i) compare more than two languages at a time, (ii) compare languages with similar orthographic depths, or (iii) test participants of proficiency levels comparable to those in other studies. Therefore, in this paper, we compared listeners from **four L1** backgrounds and compared their ability to perceive Japanese quantity contrasts. This paper is part of a larger project in which we also investigated their ability to perceive quantity contrasts in another language.

We selected to consider Japanese, English, Cantonese, and French in this study, as these languages use duration as a quantity cue to different degrees. Japanese has systematic short vs. long differences in both vowels and consonants [2]. Both obstruent consonants (e.g. /kita/ 'came' vs. /kitta/ 'cut') and vowels (e.g. /obasan/ 'aunt' vs. /obaːsan/ 'grandmother') phonologically contrast in quantity, with duration being the primary acoustic cue [2]. English has short vs. long vowels (e.g. *bit* vs. *beat*) although duration is only one of the acoustic cues (alongside vowel quality) [3]. Cantonese has short vs. long vowels but only limited to a small set of pairs (e.g. /ɐi/ vs. /aːi/). French has no phonemic quantity contrasts [4], and is said to be 'quantity insensitive'.

Here we tested the following hypotheses: (H1) Japanese listeners' perception accuracy in both discrimination and identification is the highest; (H2) French listeners' perception accuracy is the lowest. In addition, we are also interested in whether Cantonese or English speakers would perform better, as both languages partially use duration to cue quantity contrasts, but in different ways.

Twenty native listeners of Cantonese, 20 Japanese, 20 English and 15 French were recruited. They had no (history of) hearing or language impairments. They completed an AXB discrimination task and an identification task. Stimuli were 45 Japanese (nonce) words (15 CVCV base real words × 3 quantity conditions: CVCV, CVVCV, CVCCV) generated using VocalTractLab 2.2 [5]. These were produced in three synthetic voices, differing in fundamental frequency (male 110 Hz, male 150 Hz, female 200 Hz), vocal tract length, and voice quality. The actual duration of each segment is based on [6].

Figure 1 (left panel) displays the discrimination accuracy of Japanese quantity contrasts (short vs. long) by Cantonese, English, French and Japanese listeners. A logistic mixed effects model was fitted to the correct or incorrect responses. The fixed factor was participant's L1 and the random effects were participant, token, and presentation order (e.g., long-long-short, short-long-long). Orthogonal contrasts were set for the L1 factor. Results show that Japanese speakers' discrimination accuracy was significantly higher than the other three language groups, $\beta = 0.14$, $SE = 0.07$, $z = 2.03$, $p = .042$. English speakers yielded significantly higher accuracy than Cantonese speakers, $\beta = -0.47$, $SE = 0.16$, $z =-2.94$, $p < .01$.

The right panel displays identification accuracy of Japanese short and long stimuli by the same participants. We fitted another logistic mixed effects model to the correct or incorrect responses. The fixed factors were participants' L1, stimulus quantity (short, long), and their interaction. Orthogonal contrasts were set for those categorical variables. The random effects for participants and stimuli were also included.

Japanese listeners' identification accuracy was significantly higher than the other three groups overall, $\beta = 0.16$, $SE = 0.05$, $z = 3.52$, $p < .001$. There was a significant effect of stimulus quantity (short vs. long), $\beta = -0.48$, $SE = 0.05$, $z = -10.06$, $p < .001$, and the quantity effect was significantly smaller for Japanese speakers than the other group, $\beta = -0.03$, $SE = 0.01$, $z = -2.25$, $p = .024$. In addition, the quantity effect was significantly larger for English group than for Cantonese group, $\beta = -0.09$, $SE = 0.03$, $z = -3.30$, $p < .001$.

We found that Japanese listeners outperformed Cantonese, English, and French listeners in both discrimination and identification, supporting H1. On the other hand, French listeners were not found to perform worse than other groups in any of the tasks, thus refuting H2. Cantonese listeners performed worse than English listeners in discrimination but not in identification.

That Japanese listeners did not do as well on identifying long sounds may be attributed to the fact that in this study we used synthetic stimuli, in which non-durational cues to quantity were held controlled. In addition to duration, native Japanese speakers also rely on pitch movement as a quantity cue [7] when listening to natural speech.

What is puzzling is why the French listeners performed much better than expected, despite the fact that their L1 is often deemed 'quantity-insensitive' [4]. It is unclear what their good performance in the present study can be attributed to. Further investigation is needed.

For the Cantonese listeners, their partial use of duration to mark vowel quantity contrasts (in only a small subset of vowels) in their L1 may have helped them discriminate non-native quantity contrasts (i.e. above chance accuracy). However, it is unclear why they performed less well than English listeners, to whom duration is only one of multiple acoustic cues to vowel quantity. Their performance in identification, however, was not significantly different.

Although we have selected to consider multiple L1 backgrounds and compared listeners' perception accuracy in non-native word stimuli, we found that only Japanese listeners unambiguously outperformed the others. The relative performance of Cantonese and English (partial quantity distinctions) as well as French ('quantity-insensitive') in different tasks does not seem to be easily attributable to their respective use of duration as a quantity cue (*contra* [8]). Although recent experimental findings (looking at two languages at a time) have improved our understanding of L2 quantity acquisition, the present direct comparison of *four* language backgrounds has shown that the picture is far from clear. A production study with these four listener groups is currently underway to shed further light on this.

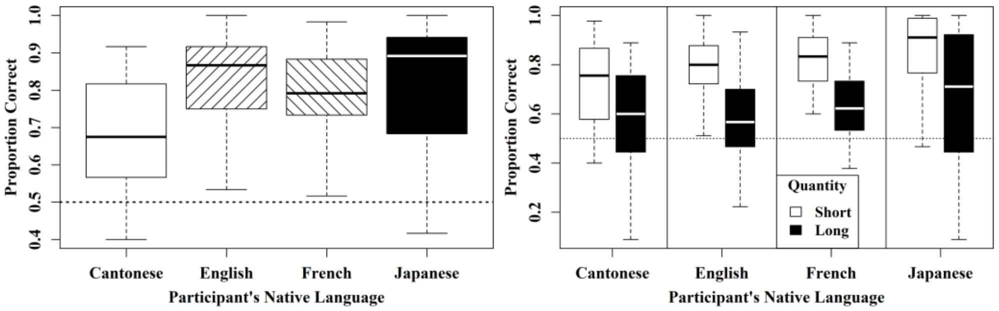


Figure 1. Discrimination (left) and identification (right) accuracy of Japanese vowel and consonant quantity contrasts by different L1 groups. The horizontal dashed lines represent chance levels.

References

[1] J. E. Flege and R. F. Port, "Cross-language phonetic interference: Arabic to English," *Lang. Speech*, vol. 24,no. 2, pp. 125–146, 1981.

[2] Y. Hirata, "Effects of speaking rate on the vowel length distinction in Japanese," *J. Phon.*, vol. 32, no. 4, pp. 565–589, 2004.

[3] A. S. House, "On vowel duration in English," *J. Acoust. Soc. Am.*, vol. 33, no. 9, pp. 1174–1178, 1961.

[4] P. A. Hallé, R. Ridouane, and C. T. Best, "Differential difficulties in perception of Tashlhiyt Berber consonant quantity contrasts by native Tashlhiyt listeners vs. Berber-naïve French listeners," *Front. Psychol.*, vol. 7, no. 209, pp. 1–16, 2016.

[5] P. Birkholz, "Modeling consonant-vowel coarticulation for articulatory speech synthesis," *PLoS One*, vol. 8, no. e60603, pp. 1–17, 2013.

[6] A. Lee and P. K. P. Mok, "Acquisition of Japanese quantity contrasts by L1 Cantonese speakers," *Second Lang. Res.*, vol. 34, no. 4, pp. 419–448, 2018.

[7] I. Takiguchi, H. Takeyasu, and M. Giriko, "Effects of a dynamic F0 on the perceived vowel duration in Japanese," 2010.

[8] R. McAllister, J. E. Flege, and T. Piske, "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian," *J. Phon.*, vol. 30, no. 2, pp. 229–258, 2002.

Neutralization of vowel length contrast in Hong Kong Cantonese checked syllables

Mei-ying Ki

The Chinese University of Hong Kong (Hong Kong)
kimeiying@link.cuhk.edu.hk

Historically, Cantonese checked syllables (i.e., those with coda -p, -t, or -k) used to have only high and low register tones, and the high tone further split into two based on vowel length: T1 (high-level tone, 55) for short vowels, and T3 (mid-level tone, 33) for long vowels (see Fig.1). Although most checked syllables still follow this pattern, there are some exceptions: some T1 checked syllables have long vowels, and some T3 checked syllables have short vowels. This shows that the matching between vowel length and tone is changing. Wong [1] observed the same phenomenon and attempted to explain it from a diachronic view of sound change and the daily usage of the exceptional words. However, no scholars have examined the phenomenon acoustically. This paper aims to explain the phenomenon and discuss its implication by examining the phonetic realization of the exceptions by focusing on the more common exception of "long vowel + T1".

A list of exceptions was identified and minimal pairs were created with other legitimate syllables. Table 1 provides an example (type B) where the long vowel [a] is matched with T1, which is an exception. In addition to the exceptions, syllables of types A, C, and D were included in the recording materials for comparison. All target syllables formed multisyllabic words or phrases to provide the context. Speakers were asked to read aloud the multisyllabic words or phrases, and then the target syllables. 33 native Hong Kong Cantonese speakers (16 males; 17 females) ranging from their twenties to sixties were recruited for the production task.

No significant difference was found in F0 between type A (mean = 185 Hz) and type B (mean = 186 Hz) syllables. However, regarding vowel duration, it was observed that a type B syllable (mean = 141 ms, e.g. [pak^{55}]) was shorter than a type C one (mean = 161 ms, e.g. [pak^{33}]). Although the difference in duration between type B and type C syllables could be attributed to their respective lexical tones (as T1 is typically shorter than T3, which may cause type B (T1) to be shorter than type C (T3)), it is notable that younger speakers tended to produce type B syllables with a vowel duration closer to type C than older speakers did. In other words, the younger speakers had longer vowel durations for type B syllables compared to older speakers. This suggests that even though type B syllables exceptionally have long vowels, older speakers still tended to shorten the vowel duration as a measure of compensation to preserve the matching between vowel length and tone.

The above results indicate that a phonological change is underway, wherein the matching between vowel length and tone is weakening. It is also noteworthy that some type A syllables (e.g. [hɐk^{55}], 'black') can have their vowel prolonged to become a type B syllable (e.g. [hak^{55}], 'black'), without affecting the word meaning or usage. This trend is more prevalent among younger speakers. These findings suggest that the vowel length contrast in Hong Kong Cantonese checked syllables is being neutralized. However, the vowel length contrast remains stable in other linguistic environments (e.g. [sam^{55}] 'three' and [sɐm^{55}] 'heart' still present an obvious and stable contrast before the -m coda). While vowel length does not contrast in most Chinese dialects, the checked syllables may hint the starting point of vowel length neutralization in Cantonese.

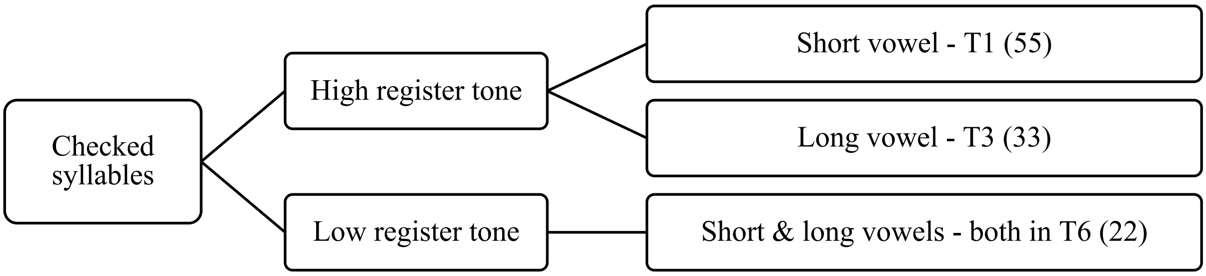


Fig.1 The matching between vowel length and tone of Cantonese checked syllables

Type A	[pɐk^{55}]	Legitimate (high register tone + short vowel + T1)	‘north’
Type B	[pak^{55}]	Exception (high register tone + **long vowel** + T1)	‘pop’
Type C	[pak^{33}]	Legitimate (high register tone + long vowel + T3)	‘uncle’
Type D	[pak^{22}]	Legitimate (low register tone + long vowel + T6)	‘white’

Table 1 Syllables that form minimal pairs with [pak^{55}]

References

[1] Wong, T. S. (2010). The Phonological Rule between Tone D1b and Vowel Length in Cantonese Revisited. In Yì Xíanghuī & Líu Cūnhàn (eds.), *Proceedings of the 14th International Conference on Yue Dialects*. Guilin, China: Journal of Guilin Normal College.

Rounded or unrounded? An examination of high vowels in Taiwan Mandarin

Chenhao Chiu[1], Po-Hsuan Huang[2]

[1, 2]*National Taiwan University*
chenhaochiu@ntu.edu.tw, r09142003@ntu.edu.tw

Taiwan Mandarin contrasts three high vowels with frontness ([i] and [y] being front) and roundedness ([y] and [u] being rounded) [1,2]. With the assignments of [round] and [back], it is suggested that sounds with the same feature value share the same articulatory gestures. However, whether or not these two sounds are identical in terms of their lip postures is not determined or specified by the assignment of their feature values. The +/− values only provide dichotomic categories between sounds; either all or nothing, but nothing in between. In particular, the [round] feature appears to provide rather simplistic interpretations of the lip postures for these high vowels as this feature is associated more with a narrowed aperture than with a specific lip posture. For example, both endolabial (e.g., [u] or [o]) and exolabial (e.g., [y] or [ø]) vowels bear the [round] features, but they contrast with each other in terms of aperture posture and the degrees of protrusion [3,4]. Insofar, whether postural differences are articulatorily achieved to distinguish the Taiwan Mandarin high vowels from one another remains unanswered. The current study investigates if there is any postural difference among the high vowels in Taiwan Mandarin and whether or not these subtle differences are consistently realized in production.

Eighteen native speakers of Taiwan Mandarin (9 F, mean = 23.44) participated in the experiment. The experiment involved a self-paced reading task. Critical stimuli consisted of the three high vowels in Taiwan Mandarin (i.e., [i], [y], and [u]), produced both in isolation (i.e., monosyllabic) and embedded in disyllabic words. Monosyllabic words were matched with a high level tone (Tone 1): [i1] (‘*one*’), [u1] (‘*house*’), and [y1] (‘*mud*’). Disyllabic words all carried a low-dipping tone (Tone 3), following the same adjective with a falling tone (Tone 4): [ta4 i3] (‘*big ant’*), [ta4 u3] (‘*fifth-grader in college*’), and [ta4 y3] (‘*heavy rain*’). The examinations of lip postures from the designated high vowels focused on *aperture postures* and *lip protrusion*. Aperture postures included horizontal distance, vertical distance, axial ratio, and aperture area; lip protrusion included the degrees of protrusion measured from both the upper and lower lips. These six measurements were then submitted to linear mixed models; with each measurement as dependent variable, and VOWEL ([i], [y], [u]) and WORD (isolation vs. disyllabic word) as fixed effects. Random slopes for VOWEL and WORD and random intercept for participant were also included.

The results showed that vowel [y] is postured significantly different from [i] in terms of aperture distances (both horizontal and vertical), axial ratio, aperture area, and lip protrusion (all p <.01), with [y] being associated with shorter aperture distances, lager axial ratio, smaller aperture area, and more protrusion. On the other hand, monosyllabic [y] only contrasted with [u] in horizontal distance, with the latter being shorter ($\beta = -0.34$, p =.01). No other differences with regards to aperture posture and lip protrusion were reported between [y] and [u] (all p >.05). Result figures are presented in Figures 1 ~ 6.

The current study compared perioral postures for the three high vowels in Taiwan Mandarin. While lip aperture characterizes the high front vowel [i], the difference between [u] and [y] did not reside in the degree of roundedness or protrusion, but rather in the postural difference at the corners of the mouth, which defines the horizontal distance in the aperture. The results show that [u] and [y], though both traditionally labeled as rounded, contrast with each other in horizontal distance between the mouth corners, yielding a more circular round posture for [u] and a more laterally compressed posture for [y]. Collectively, our results suggest that high vowels in Taiwan Mandarin are better distinguished along aperture area and lip posture. These observed postural differences for the three high vowels in Taiwan Mandarin ought to be available to the speakers and therefore may serve its function for perceptual identification, which would call for future research.

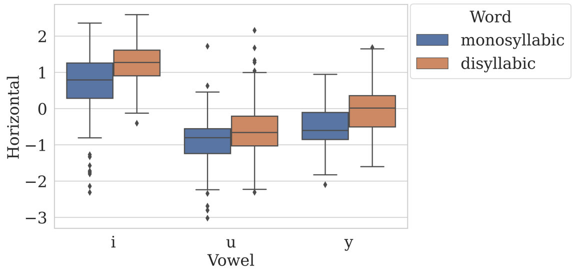


Fig. 1: Horizontal distance (z-scored) by vowels and conditions

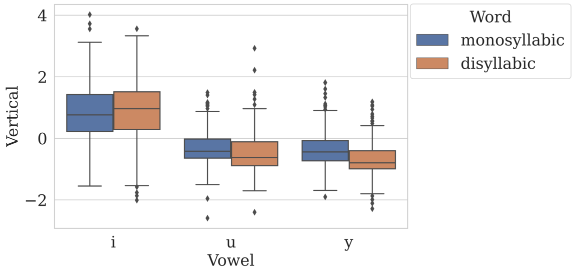


Fig. 2: Vertical distance (z-scored) by vowels and conditions

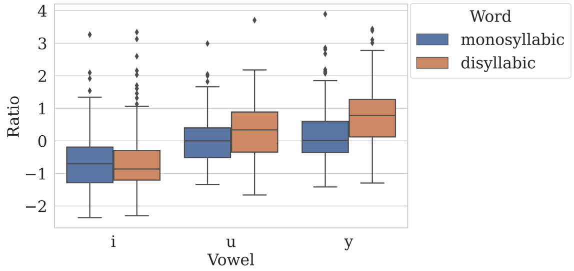


Fig. 3 Axial ratios (z-scored) by vowels and conditions

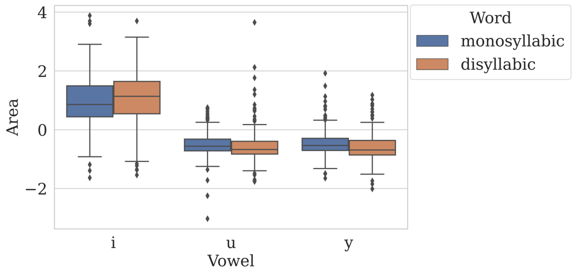


Fig. 4 Lip aperture (z-scored) by vowels and conditions

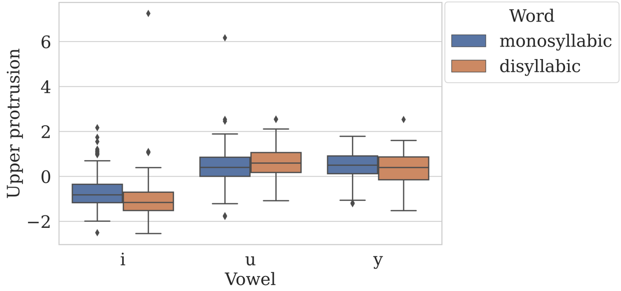


Fig. 5 Upper lip protrusion (z-scored) across vowels and conditions

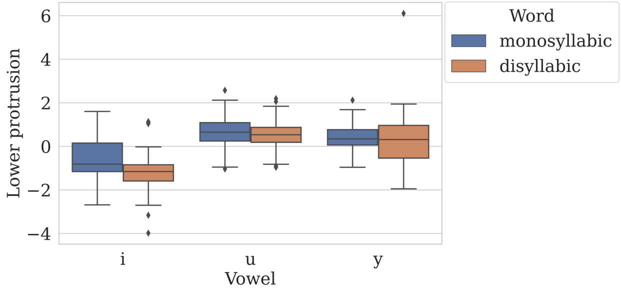


Fig. 6 Lower lip protrusion (z-scored) across vowels and conditions

References

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

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

[3] Linker, W. (1982). Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study. *UCLA Working Papers in Phonetics* 56.

[4] Catford, J. C. (1988). *A Practical Introduction to Phonetics*. Oxford: Clarendon Press.

Probabilistic accessibility of words and vowel phonetic details of L1 and L2 speakers

Jonny Jungyun Kim[1] & Mijung Lee[2]

[1]Pusan national University (Korea), [2]Independent researcher
jonnykkim@gmail.com, jude3000@naver.com

Previous studies [1, 2] showed that vowel formants are more drastically contrasted in low-frequency words with many neighborhood words (probabilistically hard words, henceforth), compared to frequent words with few neighborhood words (easy words). This effect is broadly explained by the communicative function of improving intelligibility of words with perceptual difficulty [3], and word-specifically shaped exemplar-based phonetic representations via lifetime experiences with probabilistic accessibility of words [4]. However, the spectral expansion in hard words may be accompanied by temporal reduction [2, 5], arguably because a hard word's high density also possibly means frequent articulatory activation of its segmental sequences through the use of many phonologically similar words. We re-examined these hypotheses (i.e., spectral expansion and temporal reduction of vowels in hard words) with L1 and L2 speakers in utterance-final position where articulatory declension was expected. We also tested if the effects interacted with the presence or absence of the speaker's communicative attention driven to the target word.

6 native speakers of American English and 6 high-proficiency Korean learners of English participated in a speech production experiment. Our lexical set in Table 1 was taken from [2], in which 6 different vowel categories /i, ɪ, æ, ɑ, o, u/ were repeatedly measured across 15 easy and 15 hard monosyllabic words. Each participant was recorded in two experimental blocks, differing in the level of speaker attention to the target word. In the 'unattended' block, participants read naturally the sentence "I _______ say the word, [TARGET].", filling in the blank with an adverb of frequency of their choice that best-matched their own usage frequency (among *seldom*, *sometimes*, *usually*, or *often*). By doing so, we intended to draw attention to the word's frequency, eliciting a narrow focus on the adverb and relatively reduced articulatory attention to the target word. Following the unattended block, the 'attended' block was conducted. The target appeared in a different carrier sentence, "This is the word, [TARGET].", drawing attention to the target word itself. Lexical items were randomly presented for each speaker, with 4 repetitions per block. Thus, a total of 2,880 tokens (30 words × 2 attentions × 4 repetitions × 12 participants) were obtained.

As shown in Fig. 1, while the native group showed clear separation of spectral distributions across vowel categories, the L2 group showed subsequent overlaps between /i/ and /ɪ/, and between back vowels, both of which are typically found in Korean L2-ers' speech. Importantly, the native group showed an 26.8% increase in the hexagonal area created by each of the six vowels' mean x- (F2) and y- (F1) values (both z-scored) on the reversed coordinate plane, when the vowel was contained in a hard word (3.12 z^2), compared to an easy word (2.46 z^2). In line with previous studies [1, 2, 5], back vowels induced less clear expansion than front vowels, particularly for /o/ and /ɑ/, possibly due to restrictions for retracting the tongue root. On the other hand, the L2 group barely exhibited such a trend, with a 1.5% increase from easy words (2.67 $z^{2)}$ to hard words (2.71 z^2) while their vowels were somewhat generally fronted in hard words. Manipulation of speaker attention, however, did not add any significant effect on the spectral pattern.

As for the effects on relative vowel length (vowel length divided by word length) shown in Fig. 2, the native group had a significant main effect of lexical difficulty (p<.05) in a conservative linear mixed effects analysis (with by-participant random intercept and slopes for difficulty and attention, and by-item intercept and slope for group), indicating that vowels were shorter in hard words than in easy words. Vowel length also varied as a function of speaker attention. Vowels were generally longer in the attended condition (p<.001), and the difficulty effect was greater in magnitude in the attended condition (p<.001). The L2 group, however, showed none of these effects.

Our results suggest that native speakers fine-tune the vowel form in word memory in accordance with probabilistic accessibility of words in line with [1, 2, 5], even utterance-finally. Further, vowel shortening of high-density hard words may also be part of the phonetic feature that is enhanced in contextually-driven hyperarticulation [3]. The lack of both types of effects in our L2 data highlights high-proficiency L2-ers' insensitivity to communicative cooperation suited for word accessibility and/or weaker low-level connections among phonologically similar words; alternatively, it may sheerly arise from sparse phonological links extractable from relatively small-sized lexicon, all of which are related to experience-based phonological shaping in interaction with lexical use [4].

Table 1. Lexical stimuli

Vowel	/i/	/ɪ/	/æ/	/ɑ/	/o/	/u/
EASY word	*peace teeth*	*give ship thing*	*gas Jack path*	*job shop wash watch*	*both vote*	*food*
HARD word	*bead weed*	*hick kin kit*	*hack hash pat*	*cod cot knob wad*	*goat moat*	*hoop*

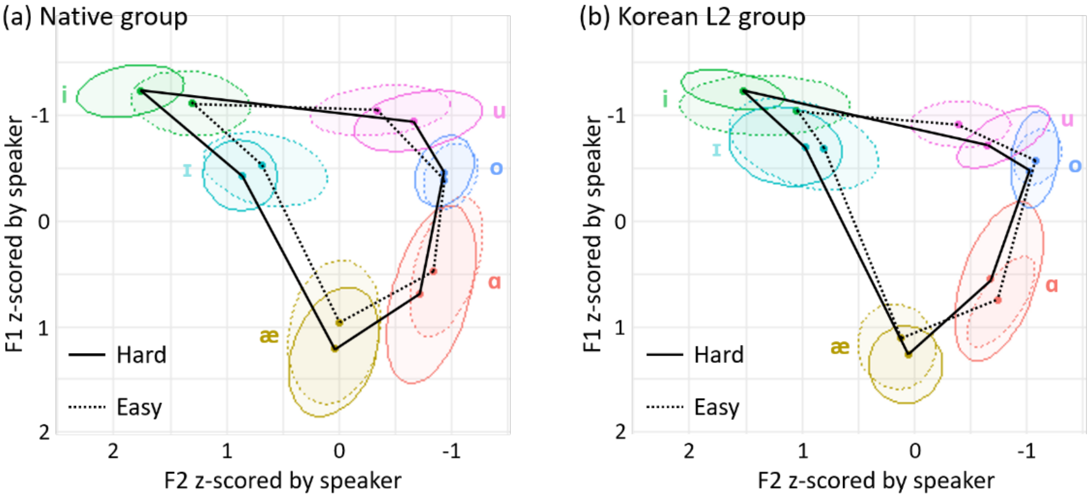
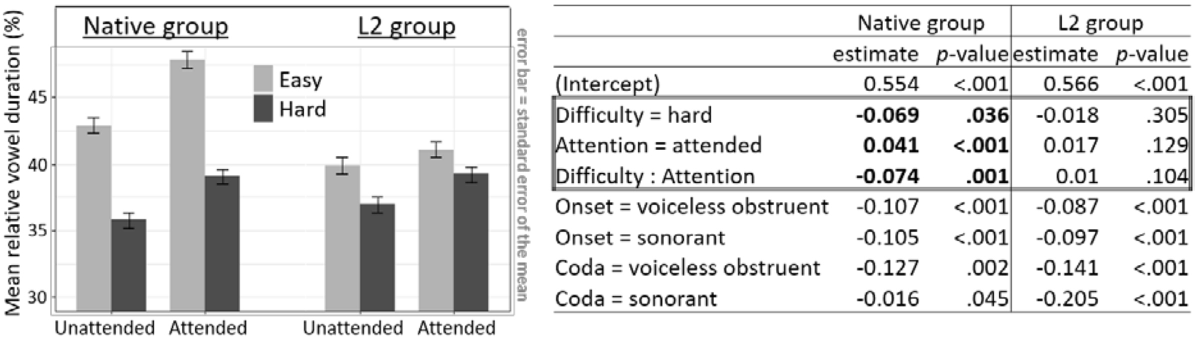


Fig. 1. Vowel plots conditioned by lexical difficulty: The ovals in six different colors were created using the stat_ellipse function (with level = 0.5) in the ggplot2 package in R to delimit the central distribution of each vowel category.



	Native group		L2 group	
	estimate	*p*-value	estimate	*p*-value
(Intercept)	0.554	<.001	0.566	<.001
Difficulty = hard	**-0.069**	**.036**	-0.018	.305
Attention = attended	**0.041**	**<.001**	0.017	.129
Difficulty : Attention	**-0.074**	**.001**	0.01	.104
Onset = voiceless obstruent	-0.107	<.001	-0.087	<.001
Onset = sonorant	-0.105	<.001	-0.097	<.001
Coda = voiceless obstruent	-0.127	.002	-0.141	<.001
Coda = sonorant	-0.016	.045	-0.205	<.001

Fig. 2. Results for vowel length: Relative vowel duration (vowel/word, in %) is predicted by difficulty, attention, and their interaction for each group in the bar-plot and in the lmer model. The model was fit separately to each group with random effects structure best supported by the data.

References

[1] Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, R. Temple, M. E. Beckman, & J. Kingston (eds.), *Papers in Laboratory Phonology VI: Phonetic Interpretation* (pp. 75-87). Cambridge, UK: Cambridge University Press.

[2] Munson, B., & Solomon, N. P. (2004). The effects of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1048-1058.

[3] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling*. (pp. 403-439). Dordrecht, Netherlands: Kluwer.

[4] Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.), *Laboratory phonology VII*, (pp. 101-139). Berlin: Mouton de Gruyter.

[5] Kilanski, K. J. (2009). *The effects of token frequency and phonological neighborhood density on native and non-native english speech production* (Doctoral dissertation). University of Washington, Seattle, WA.

Dialectal Variation of the Effect of Prosodic Prominence on Diphthong Reduction in Taiwan Mandarin – using /aɪ/ as an example

Chieh-Ching Chen[1] & Janice Fon[2]

[1]National Taiwan University (Taiwan), [2]National Taiwan University (Taiwan)
r10142003@ntu.edu.tw, jfon@ntu.edu.tw

Vowel reduction is frequently observed in spontaneous speech [1]. Previous studies show that prosodic prominence greatly influences the reduction of vowels in English [2], Dutch [3], and Swedish [4]. As prosodic strength is manifested through tonal realization in Mandarin [5], not vowel reduction itself, which is more commonly found in stress-timed languages like English, it would be interesting to see whether and how vowel reduction is affected by stress. Moreover, previous studies on vowel reduction mainly focused on monophthongs, whose reduction was manifested through formant undershoot and shorter duration [2, 3, 4]. For diphthongs, which differ from monophthongs by having an additional vowel target [6], little is said about whether and how reduction should take place. This study thus examines how the reduction of the diphthong /aɪ/ is realized in Taiwan Mandarin and how stress influences the reduction pattern. Besides, since segmental variations exist among Northern and Southern dialects of Taiwan Mandarin [7] and gender difference was observed in vowel reduction [8], we analyzed male and female speakers of the two dialects separately.

Eight hours of monologue recordings of 16 Mandarin-Min bilinguals were chosen from the Taiwan Mandarin-Min Spontaneous Speech Bilingual Corpus [9]. The speakers were evenly divided into two dialectal groups, in which half of them were male, and half were female. We found 5,127 tokens of /aɪ/ in total and labeled their phonetic realizations accordingly. Three levels of stress, S1-S3, were labeled according to Pan-Mandarin Tone and Break Indices (M-ToBI) [5]. S1 is the lowest level of stress and is used when a tone has completely lost its tonal shape. S2 is the next higher level. It is used when the tone still retains its distinctive contour, even though some of its tonal specifications have been lost. S3 is the highest level of stress. It is used to label tones that are realized with a full-fledged contour. In spontaneous speech, S2 is the most common and could be considered the default stress level [10].

Results showed that only 48.8% tokens of /aɪ/ were retained. 46.9% were monophthongized and 4.3% were reduced to other diphthongs. Monophthongization is mainly in the form of merging rather than deletion, as [e] occupied 65.4% of the reduced monophthongs, and 19.3% were centralized as [ə]. Only 15.3% were reduced to [a], in which the second vowel target was dropped. Logistic regression on reduction rate and linear mixed effects models on duration and formant frequencies were performed to examine the effect of stress, gender, and dialect in vowel reduction. Diphthongs with lower levels of stress were found to have a higher reduction rate. We did not find the effect of S1 in terms of duration and spectral results, but S2 is consistently shorter in duration with formant undershoot on both vowel targets compared with S3. Besides, the reduction of diphthongs likely lacks a negative connotation since no gender difference in reduction rate, duration, or spectral results was found [11]. Moreover, speakers of the northern and southern dialects realized [aɪ] differently and adopted different strategies in the formant undershoot. Generally, female southerners had lower and less fronted tongue position for [a] and more fronted [ɪ] in the realization of retained /aɪ/, whereas southern male speakers had higher and less fronted [a] compared with their northern counterparts. As to the realization of prosodic prominence, it was found that in unstressed [aɪ], the tongue position for [a] was higher, while [ɪ] was lower and less fronted in both dialects. Additionally, [a] was less fronted for female northern speakers, but more fronted for male southern speakers when [aɪ] was unstressed.

In this study, we found that diphthongs were more reduced as the two vowel targets merged and the reduction pattern did not differ across genders. Also, we found that diphthongs realized with more reduced tonal shapes tended to have higher reduction rates, shorter duration, and more centralized spectral results. Finally, dialectal variation was found to be manifested in the spectral results of stressed and unstressed [aɪ].

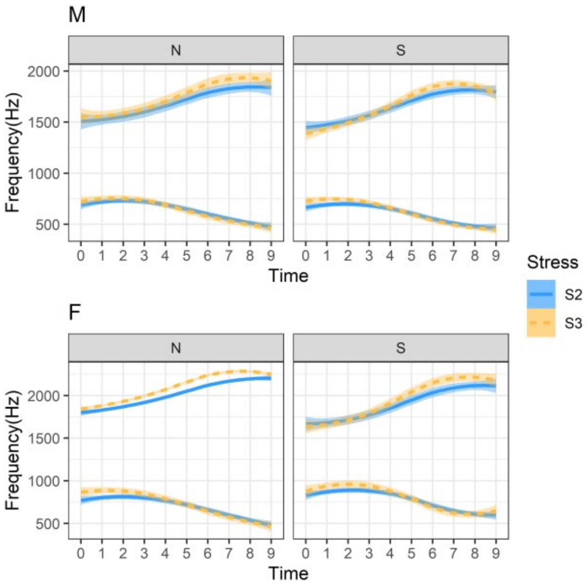


Fig.1 Spectral results of retained [aɪ] across S2 and S3 among males (M) and females (F) of the northern (N) and southern (S) dialects.

References

[1] Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language*, *22*(2), 171-184.

[2] Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical society of America*, *90*(4), 1816-1827.

[3] Van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech communication*, *12*(1), 1-23.

[4] Lindblom, B. (1963). Spectrographic study of vowel reduction. *The journal of the Acoustical society of America*, *35*(11), 1773-1781.

[5] Peng, S. H., Chan, M. K., Tseng, C. Y., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a Pan-Mandarin system for prosodic transcription. *Prosodic typology: The phonology of intonation and phrasing*, 230-270.

[6] Gussenhoven, C., & Jacobs, H. (2017). *Understanding phonology*. Routledge.

[7] Fon, J., Hung, J. M., Huang, Y. H., & Hsu, H. J. (2011). Dialectal variations on syllable-final nasal mergers in Taiwan Mandarin. *Language and Linguistics*, *12*(2), 273-311.

[8] Meunier, C., & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, *39*(3), 271-278.

[9] Fon, J. (2004). A preliminary construction of Taiwan Southern Min spontaneous speech corpus. National Science Council technical report [NSC-92-2411-H-003-050-].

[10] 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. In *Speech Prosody 2010-Fifth International Conference*.

[11] Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language variation and change*, *2*(2), 205-254.

Tonal alignment with articulatory gestures in South Kyungsang Korean

Hyunjung Joo[1], Sahyang Kim[2,4] & Taehong Cho[3,4]

[1]*Rutgers University (USA),* [2]*Hongik University (Korea),* [3]*Hanyang University (Korea),* [4]*Hanyang Institute of Phonetics and Cognitive Science*

hyunjung.joo@rutgers.edu, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Studies in Intonational Phonology have shown that tones (f_0 movements) are consistently aligned with segmental landmarks, known as *segmental anchoring* [1]. But they also showed that specific timing of tonal alignment with segments may differ across languages, especially depending on whether tonal pattern is determined lexically or post-lexically. Interestingly, several studies have examined this tone-segment alignment using Articulatory Phonology (AP), where not only segments but also tones can be considered as an articulatory gesture with an abstract target [2]. In AP, onset consonant and vowel gestures are generally known to have an in-phase coupling relationship. But the studies showed that the effect of tonal alignment on the timing of segmental gestures may differ depending on the intonational system of the language [3,4]. For instance, in some post-lexical pitch accent languages, Catalan and Viennese German, the alignment of H tone gesture in rising pitch accent does not adjust the timing of onset CV gestures, regardless of whether H tone is aligned with onset CV, simultaneously or sequentially [3]. In a lexical tone language, Mandarin Chinese, however, the alignment of H tone modifies the timing of onset CV gestures [4]. That is, in CV with H tone, onset C gesture is shifted leftward, and H tone gesture is shifted rightward, and in the middle of these two gestures, V gesture is aligned (i.e., like a C-center effect in English CCV).

Unlike these languages, South Kyungsang Korean (SKK) is a lexical pitch accent language, where tonal patterns of particular words are determined lexically (e.g., /pam/: H tone, *rice* vs. LH tone, *chestnut*), while the rest of the tonal pattern is defined post-lexically [5]. However, previous studies have only looked at tone-segment alignment patterns in the post-lexical pitch accent and the lexical tone languages, but not in the lexical pitch accent languages. Therefore, the present study will explore how a tone gesture for H and LH is aligned with segmental gestures in CVC and CVN in SKK. We will also examine how specific tonal alignment on segments differs depending on tonal type (H/LH) and coda type (CV$\underline{\text{C}}$/CV$\underline{\text{N}}$).

In the experiment, 11 SKK speakers in their 20's (6F, 5M) participated. Three monosyllabic target words varied in terms of lexical pitch accent (H vs. LH) and coda (obstruent, CVC vs. sonorant, CVN): /pap/ (H, *rice*), /pam/ (H, *night*), and /pam/ (LH, *chestnut*). They were included in the phrase-medial position of a carrier sentence and were provided a mini discourse situation where the target word is assigned a contrastive focus (Table 1). In the recording session, participants heard a pre-recorded question by a male SKK speaker through a loudspeaker and answered with a carrier sentence presented on a screen. A total of 660 sentences were recorded: 3 words x 20 repetitions x 11 speakers.

An Electromagnetic Articulography was used to measure oral constriction gestures (Lip Aperture for /p, m/, Tongue Body movement for /a/). Acoustic data were also obtained to measure tone gestures (f_0 movement). The ONSET and TARGET of constriction gestures were defined as a point of 20% of peak or trough in velocity profile. T- ONSET was at a point where f_0 initiates rising for H tone, and T-TARGET was at a point where f_0 reaches its peak (f_0 maximum).

Results showed that onset consonant and vowel gestures started simultaneously (Fig.1), in line with the notion of an in-phase relationship between onset CV gestures [2]. As for the VC timing, coda C gesture came sequentially after vowel gesture attained its target, showing a specific VC timing pattern in SKK in conjunction with the notion of an anti-phase coupling relationship between VC gestures [2]. Interestingly, H tone gesture showed an in-phase coupling relationship with coda C gesture, simultaneously showing an anti-phase coupling relationship with onset CV gestures. This shows that the effects of tonal alignment on segmental timing in SKK are different from a C-center like effect in Mandarin (lexical tone language) [4], but similar to those found in Catalan (post-lexical pitch accent language) [3], in a way that onset CV gestures remain in-phase coupled with each other, while H tone gesture is aligned much later after the CV gestures. Crucially, fine phonetic details of tone gesture in SKK differed depending on coda and tonal type. As for the coda type (Fig1.a-b), H tone gesture reached its target differently in CVC and CVN depending on coda's sonorancy. That is, the tonal target was attained much earlier in CVC than in CVN with respect to the coda's target. As for the tonal type (Fig1.b-c), H tone gesture was realized differently with a simplex tone (H) vs. a complex tone (LH) in the same CVN contexts. H tone gesture started after vowel gesture, but it was shifted much to the right for LH than for H, both T-onset and T-target being timed much later relative to V-target in LH than in H. Also, H tone gesture was longer for LH than for H, in line with the expectation that a complex tone requires longer duration to realize its target compared to a simplex tone.

To conclude, our study showed that how a tone gesture is aligned with segmental gestures in a lexical pitch accent language, SKK, by comparing differences and similarities observed in the lexical tone and the post-lexical pitch accent languages. We also showed that fine-phonetic details of the tone-segment alignment differ depending on coda and tonal type. Further studies are needed to generalize these patterns to other languages.

Table 1. Test sentences. Target words are underlined, and the focused words (a contrastive focus) in the prompt sentences (Q) and the carrier sentences (A) are in bold.

Word	Tone	Meaning	Test sentences
/pap/	H	‘rice’	Q: [jobʌn tanʌnɯn ʌnni**kuk**dwi-ɛ-da non-na]? *Do (I) put the word behind sister’s* ***soup*** *this time?* A: [ani. ʌnni**pap**dwi-ɛ]. *No. Put (it) behind sister’s* ***rice****.*
/pam/	H	‘night’	Q: [jobʌn tanʌnɯn ʌnni**pam**dwi-ɛ-da non-na]? *Do (I) put the word behind sister’s* ***chestnut*** *this time?* A: [ani. ʌnni**pam**dwi-ɛ]. *No. Put (it) behind sister’s* ***night****.*
/pam/	LH	‘chestnut’	Q: [jobʌn tanʌnɯn ʌnni**pam**dwi-ɛ-da non-na]? *Do (I) put the word behind sister’s* ***night*** *this time?* A: [ani. ʌnni**pam**dwi-ɛ]. *No. Put (it) behind sister’s* ***chestnut****.*

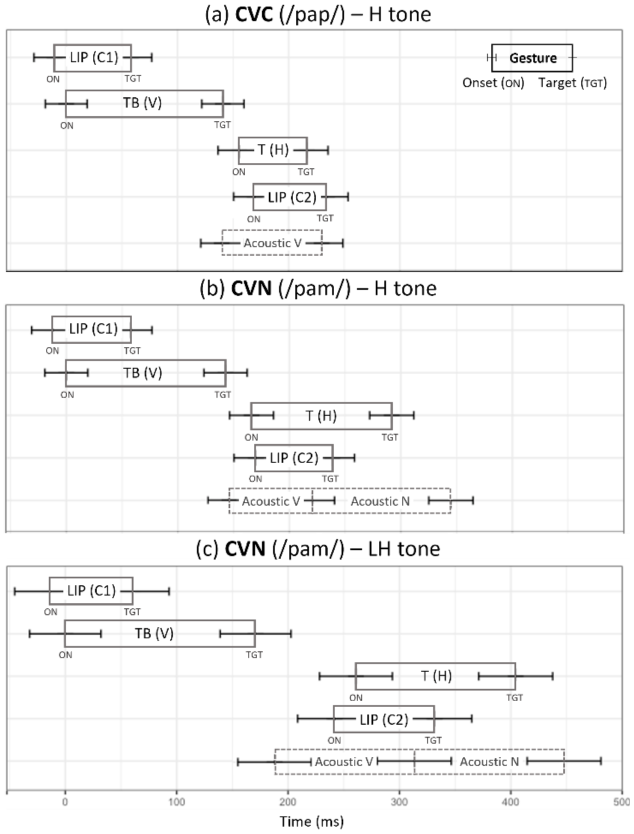


Fig1. Timings of C1-closing (Lip Aperture) gesture, V-gesture (Tongue Body lowering), C2-closing (Lip Aperture), and Tone (High, f_0 movement) gesture, depending on (a-b) coda type (CVC/CVN) and (b-c) tonal type (H/LH).

References

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

[2] Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 201-251.

[3] Mücke, D., Nam, H., Hermes, A., & Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. *Consonant clusters and structural complexity*, 26, 205.

[4] Gao, M. (2008). *Tonal alignment in Mandarin Chinese: An articulatory phonology account*. Doctoral dissertation. Yale University, New Haven.

[5] Kim, J., & Jun, S. A. (2009). Prosodic structure and focus prosody of South Kyungsang Korean. *Language Research*, 45.1, 43-66.

Gender-related variation of nasality and sound change of denasalization driven by prosodic boundaries in Seoul Korean: A preliminary report

Jungah Lee[1], Sahyang Kim[2], Taehong Cho[1]

[1]Hanyang Institute for Phonetics & Cognitive Sciences of Language, Hanyang University (Korea), [2]Hongik University (Korea)
jlee14@hanyang.ac.kr; sahyang@hongik.ac.kr; tcho@hanyang.ac.kr

Degree of consonantal nasality and its influence on the vowel's coarticulatory nasalization can be modified by the prosodic structure in a language specific way [1]. In Korean, the nasality of the nasal consonant is extremely reduced in a phrase-initial position [1], unlike other languages such as Chinese, French and English. The extreme reduction of nasality is considered as 'denasalization,' an on-going sound change in Seoul Korean [2], which seems to originate from a boundary-induced reduction of nasality in phrase-initial position [1].

The current study investigates the prosodically driven, on-going sound change of denasalization in Seoul Korean, examining effects of gender as a sociolinguistic factor and of speech rate. First, given that females play a leading role in sound change [3], we examine whether younger females also show more innovative speech form in denasalization. In addition, given that speech production can be influenced by speech rates [2], we aim to study effects of two different speech rates (normal speech vs. fast speech) on degree of nasalization. It has been reported that acoustic cues tend to be systematically different depending on speech rates [2]. In the present study, therefore, we aim to test how the nasality of the nasal consonant and the degree of vowel nasalization is conditioned by prosodic boundary and how the position-sensitive denasalization process can be different between the female versus the male speech in two different speech rates (normal vs. fast) in Seoul Korean.

Twenty speakers (10 females) in their 20's from Seoul read a passage that contains the bisyllabic target words with nasal consonants (/mami/, /mima/). The passage was originally constructed to induce these words to be produced in different prosodic boundary contexts. To elicit an IP boundary context, the target names were preceded by an adverbial phrase, which aided the speaker in inserting a phrase boundary as in **Table 1** (a); and to elicit an IP-medial context, each target word was used as the second part of a two-word compound noun as in **Table 1** (b). The first and the second syllables were analyzed separately. Nasal duration (N-duration) of nasal consonant (/m/) and degree of vowel nasalization (V-nasalization, 25%, 50%, and 75% in the vowel for relative timepoints) were measured. We also measured A1-P0 at absolute time points (20ms, 40ms, and 60ms) from the vowel onset, but the results are not included due to the space limit.

The results are summarized in Figs.1 (N-duration) and 2 (V-nasalization). As for the N-duration (Fig.1), the speech rate effect was significant. When the speakers were asked to speak faster, the N-duration became significantly shorter in general. In addition, the Boundary effect (i.e., shorter N-duration for the IP-initial than for the IP-medial position) was significant for the first syllable in both speech rates. There was no significant effect of Gender on N-duration. The N-duration in the second syllable showed similar results: the Boundary effect was significant but there was no Gender effect in both speech rates (Fig.1b). Unlike N-duration, the speech rate effect was not significant in V-nasalization. However, the effects of Boundary and Gender on V-nasalization were consistent regardless of speech rate. For the first syllable, in both speech rates, the female speakers nasalized the vowel much more than male speakers did in the IP-initial context (Fig.2a). Notably, they nasalized the vowel to a larger extent, as much as they did in the IP-medial context. In contrast, the male speakers substantially reduced V-nasalization in the IP-initial than in the IP-medial position (Fig.2a). For the second syllable, there was no effect of either Boundary or Gender on V-nasalization (Fig.2b).

To sum up, the speech rate effect was found only for N-duration, but not for V-nasalization. The gender effect was not found with N-duration, but the female speakers persistently nasalized the vowels more in IP-initial position than the males did. Interestingly, contrary to the previous findings [3], the results indicate that the on-going sound change is being led by males rather than females. We interpret that denasalization might be associated with negative social meanings in females' speech in Seoul Korean.

Table 1. Example sentences
Target words are underlined and '#' refers to a prosodic boundary.

Conditions	**Target-bearing sentences**
(a) #=IP	*[# mimanεman kanɯn salamtɯkwa ʌtʃεna # maminεman kanɯn salamtɯlo nanwiʌssʌ]* '# It was divided into two groups who only go to the bakery 'mami' and bakery 'mima' ...미마네만 가는 사람들과 언제나 마미네만 가는 사람들로 나뉘었어.
(b) #=Wd	*[yaŋpʰa#mimapaŋ, kjεpʰi#mimap*aŋ, yaŋpʰa#mamipaŋ, kjεpʰi#mamip*aŋ ilʌnsikɯlo ilɯmɯl putyεttæ]* 'They named as onion #mima bread, Cinamon #mima bread, Onion #mami bread, Cinnamon #mami bread. 양파#미마빵, 계피#미마빵, 양파#마미빵, 계피#마미빵 이런 식으로 이름을 붙였대.

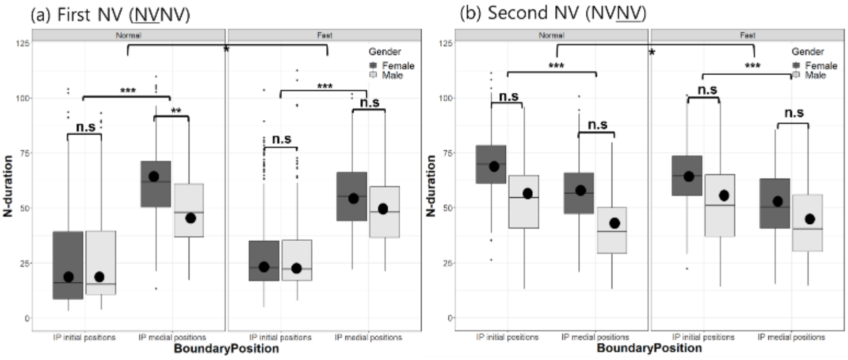


Fig 1. N-duration of first (a) and second (b) syllables across Boundary Positions in normal and fast speech (black dot represents mean values)

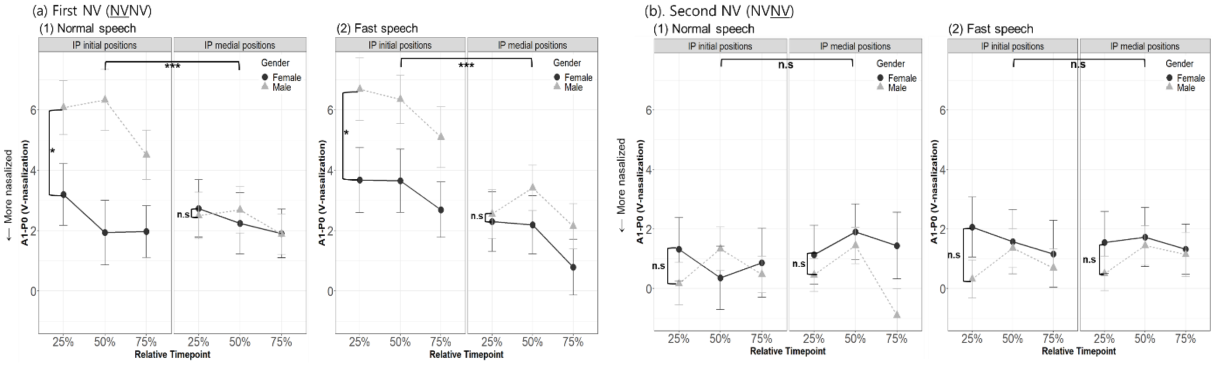


Fig 2. V-nasalization of first (a) and second (b) syllables across Boundary Positions in normal and fast speech

References

[1] 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.

[2] Yoo, K., & Nolan, F. (2020). Sampling the progression of domain-initial denasalization in Seoul Korean.

[3] Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of phonetics*, *34*(4), 500-515.

[4] Kelso, J. A., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, *14*(1), 29-59.

Prosodic Realization of Accented and Unaccented Postpositions in Japanese

Le Xuan Chan[1], Rina Furusawa[2], Rin Tsujita[2] & Seunghun J. Lee[2,3]

[1]National University of Singapore (Singapore), [2]International Christian University (Japan), [3]IIT Guwahati (India)
lxlinguistics@gmail.com, furusawar72@gmail.com, applepie846.clock@gmail.com, seunghun@icu.ac.jp

Background This paper investigates the prosodic realizations of lexically specified bi- and trimoraic post-nominal particles (postpositions) in Japanese, which are either accented (A), i.e. realized with a HL falling pitch, or unaccented (U), i.e. realized without a falling pitch.

Postpositions have conventionally been thought to form a prosodic unit with its preceding noun [1]. [2,3] supports this claim with detailed descriptions of varying noun-postposition accent combinations, but only utilizes data from pitch accent dictionaries. The study in [4] using the Corpus of Spontaneous Japanese, meanwhile, shows that A postpositions retain their lexical accent following nouns, but only looks at AA sequences. As such, there is a lack of production studies examining all possible A and U noun-postposition sequences.

A compounding factor on the prosody of these postpositions is Boundary Pitch Movement (BPM), which are F0 movements at phrase boundaries that correspond to pragmatic functions [5]. This study is interested in a specific type of BPM: the LHL% rise-fall BPM observed in the casual speech of younger Tokyo Japanese speakers, marking an “explanatory” tone. Crucially, LHL% BPM was not accounted for in [2,3], as intonation of casual speech is not included in pitch accent dictionaries. Corpus data from [4], however, shows that A postpositions subject to LHL% BPM have their accents deleted.

Methodology The present study provides a detailed investigation of the interaction between i) the lexical accent of the postposition, ii) the lexical accent of the preceding noun, and iii) LHL% rise-fall BPM. In addition, we also include U nouns and postpositions in our analysis, which were not studied in [4]. The research questions are:

a. How are A/U postpositions realized following A/U nouns?
b. How does BPM influence the prosodic realization of these postpositions?

The data for this study is taken from a larger set of data of complex DPs elicited from 6 Tokyo Japanese speakers in their 20s (4 male, 2 female). 192 tokens with the following noun-postposition sequences were analyzed: AA, AU, UU, UA (A: accented, U: unaccented). An example of an AA sequence in a carrier sentence is shown in (1).

Results Of 192 tokens, 107 were realized with LHL% BPM, and 85 were realized without BPM. U postpositions were more likely to take on a LHL% BPM (70%) than A postpositions (40%). Noun-postposition accent sequence did not affect this pattern.

In terms of prosodic realization, non-BPM tokens in UU, UA, and AU sequences showed that postpositions form a prosodic unit with its preceding noun, with realizations typical of standard Tokyo Japanese prosody (Table 1). These realizations concur with the patterns described in [2,3]. Two subpatterns, however, were displayed in AA sequences: i) downstepping of the A postposition, indicating a noun-postposition prosodic unit, and ii) pitch boost, where the postposition is given prominence over the noun and takes on a raised F0.

As for tokens with LHL% BPM, lexical accents of both A and U postpositions subject to BPM were deleted, resulting in identical realizations. The rise-fall pitch contour of LHL% BPM was realized on the final mora in both A and U postpositions (Fig. 1). Postposition F0 was always higher than the preceding noun, indicating that postpositions subject to BPM are prosodically dissociated from the preceding noun.

Discussion & Conclusion In all combinations, we see evidence that postpositions not subject to BPM are prosodically grouped with its preceding noun, thus supporting the patterns in [2,3]. The pitch boost observed in certain AA sequences, however, shows that speakers may also choose to highlight the postposition over the noun. Despite the discrepancy in AA sequences, the accents of A postpositions are preserved in all tokens, supporting the findings of [4].

This study also shows that LHL% BPM overrides these noun-postposition accent interactions and renders identical realizations of both A and U postpositions regardless of the preceding noun. Prosodic prominence is also assigned to these postpositions. Additionally, we found that U postpositions take on a LHL% BPM more readily compared to A postpositions. To explain this, we can posit that the realization of lexical pitch fall accents are prioritized in Japanese. As BPM has the effect of deleting the accent, non-BPM realizations are preferred to BPM realizations.

(1) AA sequence (boldface) elicited in a carrier sentence
gakkoo-de *ookii* *shiroi* ***megane*** ***bakari*** *hakkiri* *mieta*
school-LOC big white **glasses** **only** clearly see
‘At school, I clearly saw big white glasses only.’

	$\mathrm{U}_{\mathrm{Particle}}$	**$\mathrm{A}_{\mathrm{Particle}}$**
$\mathrm{U}_{\mathrm{Noun}}$	$\mathrm{U}_{\mathrm{Particle}}$ forms a plateau with $\mathrm{U}_{\mathrm{Noun}}$	$\mathrm{A}_{\mathrm{Particle}}$ realized faithfully without pitch boost
$\mathrm{A}_{\mathrm{Noun}}$	$\mathrm{U}_{\mathrm{Particle}}$ compressed following $\mathrm{A}_{\mathrm{Noun}}$	$\mathrm{A}_{\mathrm{Particle}}$ downstepped after $\mathrm{A}_{\mathrm{Noun}}$ $\mathrm{A}_{\mathrm{Particle}}$ realized with pitch boost

Table 1: Realization of non-BPM tokens by noun-particle sequence

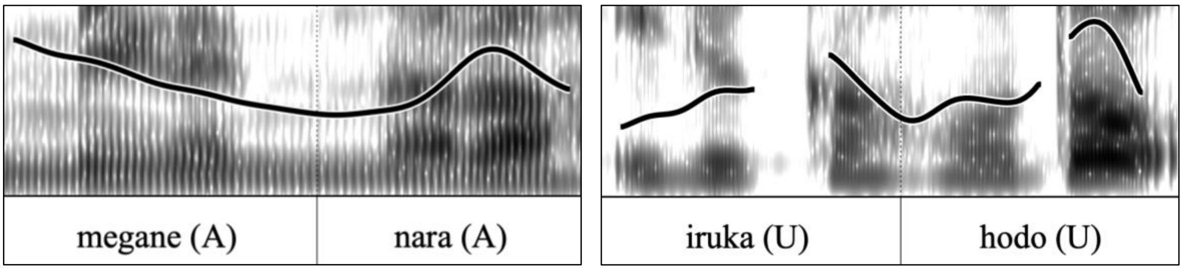


Figure 1: LHL% BPM in AA (left) and UU (right) sequences

References

[1] Igarashi, Y. (2014). Typology of intonational phrasing in Japanese dialects. In Sun-Ah Jun (ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing*. (pp. 464–492). OUP.

[2] Kori, S. (2015). Joshi jodōshi no akusento nitsuite no oboegaki: chokuzen keishiki to no fukugō keitai no kanten kara no bunrui (Reflections on the accents of particles and auxiliary verbs: classifications from the perspective of compound form with the preceding pattern). *Gengo bunka collaborative research project* 2014: 63-74.

[3] Kori, S. (2020). Nihongo no joshi jodōshirui no akusento : ichiran to tsukaiwake, henka no hōkōsei (Accents of particles and auxiliary verbs: list and the direction of its proper use and change). *Gengo bunka collaborative research project* 2019: 13-24.

[4] Maekawa, K., & Igarashi, Y. (2007). Prosodic phrasing of bimoraic accented particles in spontaneous Japanese. In *Proceedings of ICPhS 16,* (pp.1217-1220). Saarbrücken.

[5] Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. (pp.172-200). OUP.

Sound Change and Emergence of Patterns of the Syllable-final Consonants in the Chinese Dialects

Wai-Sum Lee & Eric Zee

Department of Linguistics and Translation, City University of Hong Kong (Hong Kong)
w.s.lee@cityu.edu.hk, ctlzee@friends.cityu.edu.hk

Introduction. The syllable-final consonants that occur in the present-day Chinese dialects include the unreleased voiceless stops [-p -t -k], glottal stop [-ʔ], and nasals [-m -n -ŋ]. They are the reflexes of the historical syllable-final stops, -**p* -**t* -**k*, and nasals, -**m* -**n* -**ŋ*, that occurred in Middle Chinese and Old Chinese ([1,2,3]). In some dialects, -**p* -**t* -**k* -**m* -**n* -**ŋ* are preserved. In other dialects, some of them are retained, resulting in the varied patterns of the syllable-final stops and nasals in different dialects. The purpose of this study is twofold, (i) to identify the changes of the historical -**p* -**t* -**k* -**m* -**n* -**ŋ* in the present-day Chinese dialects through examining the content of the defective patterns, which do not contain all the descendants of -**p* -**t* -**k* -**m* -**n* -**ŋ*, and (ii) to explain the sound changes in the syllable-final consonants that have led to the formation of the defective patterns. The study is based on the occurrence data on the syllable-final stops and nasals in a representative sample of genetically and areally balanced dialects of the 70 subgroups of the 11 Chinese dialect groups, including *Mandarin* (8), *Jin* (6), *Hui* (5), *Gan* (8), *Xiang* (5), *Wu* (8), *Min* (9), *Yue* (8), *Kejia* (9), *Pinghua* (2), and *Tuhua* (2). The numeral in parentheses denotes the number of the subgroups of each of the 11 dialect groups.

Patterns. A total of 15 patterns of the syllable-final stops and nasals in the 70 sample dialects are identified and they may be grouped into five types. <u>Type 1</u> consists of six patterns, [-p -t -k -m -n -ŋ] (15, i.e., occurring in 15 dialects), [-p -t -k -ʔ -m -n -ŋ] (1), [-p -k -m -ŋ] (2), [-p -ʔ -m -n -ŋ] (1), [-t -k -n -ŋ] (2) and [-t -ʔ -n -ŋ] (2). These patterns contain the syllable-final oral/glottal stops and nasals. <u>Type 2</u> consists of three patterns, [-ʔ -ŋ] (9), [-ʔ -n -ŋ] (6) and [-ʔ -m -n] (1), that contain a glottal stop and one or two nasals. <u>Type 3</u> consists of a single pattern, [-ʔ] (3), that contains only a glottal stop. <u>Type 4</u> consists of four patterns, [-n -ŋ] (14), [-ŋ] (9), [-n] (3) and [-m -ŋ] (1), that contain one or two nasals. <u>Type 5</u> consists of a single pattern that contains no or zero syllable-final consonant 'Ø'.

Sound change in syllable-final stops. The historical -**p* -**t* -**k* are preserved in the dialects, in which the <u>Type 1</u> patterns, [-p -t -k -m -n -ŋ] and [-p -t -k -ʔ -m -n -ŋ], occur. The pattern, [-p -t -k -ʔ -m -n -ŋ], occurs in a single *Min* dialect, *Xiamen*. The extra [-ʔ] in the pattern results from the bifurcation of -**p* -**t* -**k* into (i) [-p][-t][-k] when following the back vowels and (ii) [-ʔ] when following the front vowels. The changes in the syllable-final stops in the other <u>Type 1</u> patterns, [-p -k -m -ŋ], [-p -ʔ -m -n -ŋ], [-t -k -n -ŋ] and [-t -ʔ -n -ŋ], include (i) the merging of -**p* into [-k] and [-t], -**t* into [-p], -**t* into [-k], and -**k* into [-t]; and (ii) cases of debucaalization, [-ʔ] < -**p*, [-ʔ] < -**t* and [-ʔ] < -**k*. The changes are bi-directional in terms of place of articulation. Also, the change from -**p* to [-k] does not require -**p* to first change to [-t]; and the change from -**p* or -**t* to [-ʔ] does not require -**p* or -**t* to first change to [-k]; the change in -**p* -**t* -**k* is triggered by the change in the preceding vowel; and the direction of change is conditioned by the vowel type. In <u>Type 2</u> and <u>Type 3</u> patterns, [-ʔ -ŋ], [-ʔ -n -ŋ], [-ʔ -m -n] and [-ʔ], each of -**p* -**t* -**k* has turned into [-ʔ].

Sound change in syllable-final nasals. The historical -**m* -**n* -**ŋ* have remained intact in the <u>Type 1</u> patterns, [-p -t -k -m -n -ŋ], [-p -t -k -ʔ -m -n -ŋ] and [-p -ʔ -m -n -ŋ]. The pattern, [-p -t -k -ʔ -m -n -ŋ], occurs only in *Xiamen*. In the dialect, -**m* -**n* -**ŋ* have remained intact in some rimes, and in other rimes, they are dropped with concomitant change of the preceding vowel into a nasal vowel ([Ṽ] < 'Ø' < -**m* -**n* -**ŋ*). Such changes have not taken place in the syllable-final nasals in the other two patterns, [-p -t -k -m -n -ŋ] and [-p -ʔ -m -n -ŋ]. The occurrence of nasal vowels in *Xiamen* suggests that [-m][-n][-ŋ] are unstable. Despite the similarity on the surface level between [-p -t -k -ʔ -m -n -ŋ] and [-p -t -k -m -n -ŋ] in number and type of the syllable-final nasals, the phonological behaviour of [-m][-n][-ŋ] in the two patterns differs. This is also true between [-p -t -k -ʔ -m -n -ŋ] and [-p -ʔ -m -n -ŋ]. The other <u>Type 1</u> patterns, [-p -k -m -ŋ], [-t -k -n -ŋ], [-t -ʔ -n -ŋ], <u>Type 2</u> patterns, [-ʔ -ŋ], [-ʔ -n -ŋ], [-ʔ -m -n], and <u>Type 4</u> patterns, [-n -ŋ], [-ŋ], [-n], [-m -ŋ], contain one or two syllable-final nasals. The changes in the syllable-final nasals in these patterns are (i) elision of one or two of -**m* -**n* -**ŋ*, (ii) change in place of articulation, including [-n] < -**m*, [-ŋ] < -**m*, [-m] < -**n*, [-ŋ] < -**n*, [-n] < -**ŋ*, but not [-m] < -**ŋ*, (iii) bifurcation, including [-m, -n] < -**m*, [-m, -ŋ] < -**m*, [-n, -ŋ] < -**m*; [-n, -m] < -**n*, [-n, -ŋ] < -**n*, [-m, -ŋ] < -**n*; and [-ŋ, -n] < -**ŋ*, but not [-ŋ, -m] < -**ŋ* and [-m, -n] < -**ŋ*, and (iv) emergence of nasal vowels ([Ṽ] < 'Ø' < -**m* -**n* -**ŋ*).

Theoretical consideration and explanation. **(I)** Chen [4] proposes a theory of diachronic change of -**p* -**t* -**k* and -**m* -**n* -**ŋ* in the Chinese dialects and postulates successive stages of the developmental changes of the historical syllable-final stops and nasals. The stages of change in the syllable-final stops are Stage 1: -**p* -**t* -**k* → Stage 2: [-t, -k] ([-t] < -**p*; -**t* -**k* unchanged) → Stage 3: [-k] ([-k] < -**t*; -**k* unchanged) → Stage 4: [-ʔ] ([-ʔ] < [-k]) → Stage 5: 'Ø' ('Ø' < [-ʔ], i.e., [-ʔ] dropped). The stages of change in the syllable-final nasals are Stage 1: -**m* -**n* -**ŋ* → Stage 2: [-n, -ŋ] ([-n] < -**m*; -**n* -**ŋ* unchanged) → Stage 3: [-ŋ] ([-ŋ] < -**n*; -**ŋ* unchanged) → Stage 4: [Ṽ] (< 'Ø' < [-ŋ], i.e., [-ŋ] dropped and occurrence of the nasal vowel) → Stage 5: V (< [Ṽ], i.e., vowel de-nasalization). Chen's postulation indicates (i) the diachronic changes are unidirectional in respect to the place of articulation, that is, from front to back, [-p] > [-t] > [-k] > [-ʔ] > 'Ø' and [-m] > [-n] > [-ŋ] > [Ṽ] > [V], and (ii) the successive stages are unalterable and unskippable. However, the present study shows that the diachronic changes in -**p* -**t* -**k* and -**m* -**n* -**ŋ* are (a) bi-directional, such as [-p] < -**t*, [-t] < -**p*; [-t] < -**k*, [-k] < -**t*; and [-m] < -**n*, [-n] < -**m*; [-n] < -**ŋ*, [-ŋ] < -**n*, (b) the successive stages are alterable and skippable, and (c) the changes in most cases are conditioned by the pre-consonantal vowel type. There is no a priori phonetic or phonological justification that -**p* or -**t* must first turn into [-k] before changing to [-ʔ]. Articulatorily for -**p*, -**t* or -**k* to turn into [-ʔ], the release of the oral closure suffices. As reported in Iwata, et al. [5,6], the laryngoscopic data reveal that the production of the syllable-final applosives in the Chinese dialects, Fukienese and Cantonese, is glottalized, characterized by a laryngeal constriction with a closed glottis as observed in the production of the glottal stop. The glottalization prevents the vocal folds from vibrating at the vowel offset and creates the phonatory condition for effectively producing the unreleased final stops. There are parallels between the syllable-final stops and nasals. (a) It is unnecessary for -**m* and -**n* to turn into [-ŋ] before the occurrence of [Ṽ], as the changes, [Ṽ] < 'Ø' < -**m* and [Ṽ] < 'Ø' < -**n*, occur in many dialects, (b) the diachronic changes in the historical syllable-final nasals are bi-directional, such as [-m] < -**n*, [-n] < -**m*; [-n] < -**ŋ*, [-ŋ] < -**n*, and (c) the successive stages are alterable and skippable. The diachronic data in the present study thus call into question on the validity of Chen's theory. **(II)** In this study, the syllable-final consonants in descending frequency of occurrence are [-ŋ] (61) > [-n] (56) > [-ʔ] (23) > [-m] (20), [-t] (20), [-k] (20) > [-p] (18). A larger number of occurrence**s** of [-m][-n][-ŋ] than that of [-p][-t][-k] at the same place of articulation may be because the nasal murmur and nasality on the V-to-N transition are more perceptible than the unreleased stops, contributing to nasal identification and nasal place distinction ([7,8]). As for a larger number of occurrence**s** of [-ŋ] than that of [-m -n], it may be because in the dialects [-ŋ] occurs more frequently after the vowels [a ɑ ɔ], which have a higher intensity level ([9,10]), contributing to a more perceptible V-to-N transition. **(III)** The occurrence of [-ʔ] following the loss of [-p][-t][-k] serves the function of preventing the disappearance of checked syllables and checked tones, thus maintaining the contrast between CV, CVN and CVS syllables (N = nasal; S = stop). The same can be said about the function of the nasal vowels that occur following the loss of [-m][-n][-ŋ].

Conclusion. The change in the syllable-final consonants is a link in the chain of sound change. It is triggered by the change in the preceding vowel, and it in turn brings about the changes in syllable type and tone type and the appearance of nasal vowels, which contribute to the shaping of the sound systems in the Chinese dialects.

References

[1] Karlgren, B. (1954). Compendium of Phonetics in Ancient and Archaic Chinese. *Bulletin of the Museum of Far Eastern Antiquities*, No. 22. Stockholm, Sweden: MFEA.

[2] Pulleyblank, E.G. (1977-1978). The final consonants of Old Chinese. *Monumenta Serica, 33*, 180-206.

[3] Pulleyblank, E.G. (1984). *Middle Chinese: A Study in Historical Phonology*. Vancouver, US: UBC Press.

[4] Chen, M. (1973). Cross-dialectal comparison: a case study and some theoretical considerations. *Journal of Chinese Linguistics, 1*(1), 38-63.

[5] Iwata, R., M. Shawashima, H. Hirose, & S. Niimi (1979). Laryngeal adjustments of Fukienese stops - Initial and final applosives. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 13*, 61-81.

[6] Iwata, R., M. Shawashima, & H. Hirose (1981). Laryngeal adjustments for syllable-final stops in Cantonese. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 15*, 45-54.

[7] Racasens, D. (1983). Place cues for nasal consonants with special reference to Catalan. *Journal of the Acoustical Society of America, 73*(4), 1346-1353.

[8] Kurowski, K. & S.E. Blumstein (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *Journal of the Acoustical Society of America, 76*(2), 383-390.

[9] Peterson, G.E. & H.L. Barney (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America, 24*(2), 175-184.

[10] Lehiste, I. & G.E. Peterson (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America, 31*(4), 428-435.

Oral Presentations

Day 2

(Saturday, May 27, 2023)

Challenges of analyzing variability in speech from linguistic and motor control perspectives

D. H. Whalen[1,2,3]

[1]City University of New York (USA), [2]Haskins Laboratories (USA), [3]Yale University (USA)
whalen@haskins.yale.edu

Variability is intrinsic to movement in biological systems, including speech. Although such variability has, in the past, been treated as unwanted noise, there is increasing evidence to indicate that variability has uses as well. When learning a new task, variability can lead to faster learning via "exploration" [1]. Further, lack of typical variability can be classified in extreme cases as a disorder [2]. In speech, variability is seen in virtually every measure taken, and it seems to be unavoidable as well [3]. Speakers appear to match the variability in their environment even when it does not match their intrinsic rate [4]. The fact that variability does appear to be normally distributed [5] suggests that the central tendency is indeed the target for speech sounds. Online compensation might indicate that trajectories are corrected during a syllable's production [6], but a more likely alternative is that speakers have a somatosensory indication of the accuracy of a starting position. Changes in starting position do affect intergestural timing [7], consistent with the availability of such information.

Theories of phonetics have dealt with variability in different ways. To the extent that linguistic phonologies are the beginning of a planning process, they leave any implementation of variability to a phonetic level. The Directions into Velocities of Articulators (DIVA) model has mappings between somatosensation locations and acoustic consequences of such configurations [8]. Although this model can accommodate motor equivalence for producing similar acoustic outputs, it is not at all clear how the targets are selected within the target region during production. Articulatory Phonology [9, 10] is implemented via a dynamical system, but the basic theory generates a single, determinate set of parameters for a given context, resulting in a lack of variance. Variability in phonetic measurements has been handled by adding stochastic noise to the model [11]. However, even though the pattern of results can be matched in this way, stochastic noise does not seem to allow for a differentiation between deliberate (exploratory) variability and inadvertent (true noise) variability.

An approach to useful and harmful variability that has been developed in the motor control literature is the Uncontrolled Manifold Analysis (UCM) [12, 13]. Given multiple repetitions of successful movement trajectories, it is possible to see which variants are benign (lying on the uncontrolled manifold) versus destructive (being part of the controlled manifold, i.e., the path to success). Some attempts have been made to apply this model to speech [14, 15], but the nonlinear relations between articulation and acoustics make it difficult to obtain enough tokens to train an appropriate model (note that in reaching studies, the relations were mostly linear). Further, the size of the target changes the manifold itself, and the manifold is the way in which the action achieves the target. Smaller targets lead to more constrained actions. With larger targets, there will be some correct productions that may nonetheless be considered non-ideal ("kind of a success"). The UCM says nothing about this effect. The other major issue missed by focusing only on the target is that there is information that often exists in the trajectory itself. Listeners make use of contextual variability when they "parse" the speech signal for coarticulatory effects [16]. Thus, definitions of the target and the success are both difficult to define for speech.

Where does this leave us? We need to study larger datasets than has been possible in the past. However, because even finding a true measure of the shape of the variability requires about 200 token of the same production [5], direct experimentation is challenging. Analysis of large corpora necessarily excludes a careful analysis of contextual factors (which may contribute to parsing), in addition to relying on inaccurate measures of vocal tract resonances [17]. It would therefore seem that a combination of modeling, production and perception experiments, elaboration of theories, and corpus work is necessary. The overarching issue is what the target of speech sounds is, and how much we make use of the variability intrinsic to the targets and in individual tokens reaching the targets.

References:

1. Wu, H.G., et al., *Temporal structure of motor variability is dynamically regulated and predicts motor learning ability.* Nature Neuroscience, 2014. **17**(2): p. 312-321.
2. Goldberger, A.L., *Fractal variability versus pathologic periodicity: Complexity loss and stereotypy in disease.* Perspectives in Biology and Medicine, 1997. **40**: p. 543-561.
3. Tilsen, S., *Structured nonstationarity in articulatory timing*, in *Proceedings of the 18th International Congress of Phonetic Sciences*, The Scottish Consortium for ICPhS 2015, Editor. 2015, University of Glasgow: Glasgow. p. 1-5.
4. Tang, D.-L., B. Parrell, and C.A. Niziolek, *Movement variability can be modulated in speech production.* Journal of Neurophysiology, 2022. **128**: p. 1469-1482.
5. Whalen, D.H. and W.-R. Chen, *Variability and central tendencies in speech production.* Frontiers in Communication, 2019. **4**(49): p. 1-9.
6. Niziolek, C.A., S.S. Nagarajan, and J.F. Houde, *What does motor efference copy represent? Evidence from speech production.* Journal of Neuroscience, 2013. **33**: p. 16110-16116.
7. Shaw, J.A. and W.-R. Chen, *Spatially conditioned speech timing: Evidence and implications.* Frontiers in Psychology, 2019. **10**(2726).
8. Guenther, F.H., *A neural network model of speech acquisition and motor equivalent speech production.* Biological Cybernetics, 1994. **72**: p. 43-53.
9. Browman, C.P. and L.M. Goldstein, *Towards an articulatory phonology.* Phonology Yearbook, 1986. **3**: p. 219-252.
10. Iskarous, K. and M. Pouplier, *Advancements of phonetics in the 21st century: A critical appraisal of time and space in Articulatory Phonology.* Journal of Phonetics, 2022. **95**(101195): p. 1-28.
11. Gafos, A.I., et al., *Stochastic time analysis of syllable-referential intervals and simplex onsets.* Journal of Phonetics, 2014. **44**: p. 152-166.
12. Latash, M.L., *Biomechanics as a window into the neural control of movement.* Journal of Human Kinetics, 2016. **52**(1): p. 7-20.
13. Scholz, J.P. and G. Schöner, *The uncontrolled manifold concept: identifying control variables for a functional task.* Experimental Brain Research, 1999. **126**: p. 289-306.
14. Kang, J., *The effect of speaking rate on vowel variability based on the uncontrolled manifold approach and flow-based invertible neural network modeling*. 2021, City University of New York.
15. Szabados, A. and P. Perrier, *Uncontrolled Manifolds in vowel production: Assessment with a biomechanical model of the tongue*, in *Proceedings of Interspeech 2016*, N. Morgan, Editor. 2016. p. 3579-3583.
16. Fowler, C.A. and M. Smith, *Speech perception as "vector analysis": An approach to the problems of segmentation and invariance*, in *Invariance and variability in speech processes*, J.S. Perkell and D.H. Klatt, Editors. 1986, Lawrence Erlbaum Associates: Hillsdale, NJ. p. 123-136.
17. Whalen, D.H., et al., *Formants are easy to measure; resonances, not so much: Lessons from Klatt (1986).* Journal of the Acoustical Society of America, 2022. **152**: p. 933-941.

Temporal coordination of CV: The case of liaison and enchaînement in French

Sejin Oh[1], Alice Yildiz[1], Cecile Fougeron[1], Philipp Buech[1], Anne Hermes[1]

[1]*LPP (UMR7018), CNRS/Sorbonne-Nouvelle (France)*
se-jin.oh@cnrs.fr, alice.yildiz@sorbonne-nouvelle.fr, cecile.fougeron@sorbonne-nouvelle.fr, philipp.buech@sorbonne-nouvelle.fr, anne.hermes@sorbonne-nouvelle.fr

OVERVIEW: Syllable structure is hypothesized to be associated with a characteristic pattern of temporal coordination [1-4]. The coupled oscillator model hypothesizes that CV is coordinated in-phase while VC is coordinated anti-phase with each other. However, a consonant which was an underlyingly coda consonant can also be resyllabified to the following syllable when it is followed by a vowel. Fundamental questions are whether a resyllabified consonant is a 'true' onset, and whether it is shown in the coordination pattern. In French, there are two different types of resyllabification: 1) the resyllabification of an underlying word-final coda (Enchaînement CV); 2) the resyllabification of a word final liaison consonant, i.e., a latent consonant surfacing only when the following word starts with a vowel (Liaison CV). As shown in Table 1, for example, these two resyllabification cases and the case with a true word-initial onset consonant (Onset CV) are distinct underlyingly, but are said to be homophonous in French. However, acoustic studies have shown that the phonetic neutralization between forms like the ones presented in Table 1 is incomplete: liaison and enchaînement consonants can be shorter in acoustic duration, the vowel preceding them can be longer, and/or they can preserve some specific allophonic properties associated with their word-final position [e.g., 5-10]. Furthermore, an articulatory study from two Quebec French speakers also showed that Liaison is articulatorily different from Onset and Enchaînement, (e.g., a smaller magnitude release gesture) with mixed results for different kinds of lexical items [11].

THE TEMPORAL COORDINATION: Shaw et al. [12] hypothesized that for complex segments, such as /p^j/, the onset of G2 is temporally coordinated with the onset of G1, while for segment sequences, such as /pj/, the onset of G2 is temporally coordinated with the offset of G1. These competing coordination relations were explored by investigating how the lag between the onset-to-onset varied with G1 duration. They found that for complex segments (palatalized consonants in Russian, such as /p^j/), variation in duration had little effect on lag. In contrast, for English segment sequences (such as /pj/), as consonant duration increased, so too did the lag between consonant and glide gestures, showing a strong positive correlation. Exploiting this temporal diagnosis, the current study aims to understand the temporal coordination of three different CV types (Onset CV, Enchaînement CV, and Liaison CV). The present study asks the following research questions: 1) Is there a temporal difference among the three different types of CVs and does this difference hold at different speech rates? 2) Do the different types of CVs affect the temporal coordination of gestures in terms of the relationship between C duration and lag?

EXPERIMENT: Three female native speakers of French participated. The materials include the three minimal and near-minimal triplets, where each triplet consisted of an Onset CV, Enchaînement CV, and Liaison CV. The target sequences were produced within a carrier sentence, and each sentence was produced 14 times (7 at normal and 7 at fast speech rates). Sensors, attached to the upper and lower lips, jaw, tongue tip (TT), tongue blade (TB), tongue dorsum (TD), and left/right mastoids were tracked by means of Electromagnetic Articulography (EMA, AG501) and an audio-recording setup. The TT sensor indexed the consonant gesture /t/ and the TD sensor was used to identify the vowel gesture /a/. Articulatory movements were parsed using a custom Python script, which identifies temporal landmarks of gestures with reference to the velocity signal. The four key temporal intervals computed from these articulatory landmarks were (1) consonant duration: C DURATION = $C_{RELEASE}$ – C_{TARGET}; (2) gestural overlap: OVERLAP = $C_{RELEASE}$ – V_{ONSET}; (3) lag between the gestural onsets: ONSET LAG = V_{ONSET} – C_{ONSET}; (4) lag between the gestural targets: TARGET LAG = V_{TARGET} – C_{TARGET}. Also, we analyzed the correlation between C DURATION and TARGET LAG. Also, we analyzed the correlation between C DURATION and TARGET LAG. Bayesian linear mixed models using the *brms* package v2.18.0 [13] in R v4.2.2 [14] were performed for the statistical analysis.

RESULTS: Figure 1 plots the relation between C DURATION (x-axis) and TARGET LAG (y-axis) across CV TYPE. C DURATION is not strongly correlated with TARGET LAG, showing only a slight upward trend. Notably, we observe the same pattern for all CV TYPE. To assess these results, we fit a series of Bayesian regression models to the data. We found evidence that variation in C DURATION impacts TARGET LAG, but the impact is small (β=0.82 [0.50, 1.15]). Crucially, however, we also found evidence that the way that variation in C DURATION impacts TARGET LAG is uniform across different types of CV in French (Cdur_Enchaînement β=-0.25 [-0.68, 0.18]; Cdur_Liaison β=0.40 [-0.09, 0.89]). Table 2 summarizes the results from Bayesian regression models for C DURATION, OVERLAP, ONSET LAG, and TARGET LAG. The statistical analysis reveals that there was no evidence for Enchaînement CV or Liaison CV being different from Onset CV in terms of four temporal intervals: C DURATION, OVERLAP, ONSET LAG, and TARGET LAG. The results indicate that Enchaînement CV and Liaison CV showed the same coordination as Onset CV.

DISCUSSION: The results of the present study provide evidence that resyllabified consonants are also timed in-phase with the following vowel, unlike [11] have found for Quebec French. Obtaining physiological data from EMA allowed us to examine the temporal coordination of gestures involving the articulation of external sandhi in French. Moreover, we make use of the concept of coordination to relate speech kinematics to the syllable structure. However, this resyllabification across words raises many questions—to be resolved—on the process of inter-gestural coordination. In future work, we will further investigate the temporal properties of the preceding vowel, which was also reported to be one of the acoustic characteristics preserving the contrast between the three sequence types. In addition, we will examine whether enchaînement C and liaison C maintain coordination with the preceding vowel.

Onset	**Enchaînement**	**Liaison**
/CV#**CV**/	/CV.**C#V**/	/CV.**CL#V**/
petit tamis /pəti # tami/ [pə.ti.ta.mi]	petite amie /pətit # ami/ [pə.ti.ta.mi]	petit ami /pəti # ami/ [pə.ti.ta.mi]

Table 1: Examples of three types of CV in French.

	Enchaînement – Onset	Liaison – Onset
C DURATION	β=-1.45 [-5.79, 2.89]	β=-1.71 [-6.03, 2.59]
OVERLAP	β=3.85 [4.31, 11.76]	β=-3.85 [-11.63, 4.04]
ONSET LAG	β=-10.51 [-22.50, 1.54]	β=4.99 [-6.09, 16.45]
TARGET LAG	β=-10.55 [-30.03, 8.68]	β=-1.61 [-20.48, 17.61]

Table 2: Summary of Bayesian linear mixed models

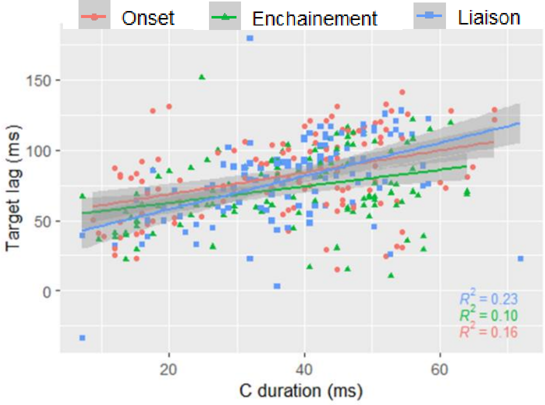


Figure 2: Target lag by C duration across CV

[1] Goldstein, L., Byrd, D., & Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. *In: Arbib M, editor. From action to language: The mirror neuron system. Cambridge: Cambridge University Press; 2006. pp. 215–249.*

[2] Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. 2009. Coupled oscillator planning model of speech timing and syllable structure. In G. Fant, H. Fujisaki & J. Shen (Eds.) *Frontiers in phonetics and speech science: Festschrift for Wu Zongji* (pp. 239–249). Beijing: Commercial Press.

[3] Nam, H., Goldstein, L., & Saltzman, E. 2009. Self-organization of syllable structure: A coupled oscillator model. *Approaches to phonological complexity* (pp. 297-328). Berlin/New York: Mouton de Gruyter.

[4] Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Proceedings of the 4th International Conference on Speech Prosody (Speech Prosody 2008), Campinas, Brazil*, 175-184.

[5] Gaskell, M., Spinelli, E., and Meunier, F. 2002. *Perception of resyllabification in French. Memory and Cognition*, 30, 798–810.

[6] Spinelli, E., McQueen, J., and Cutler, A. 2003. Processing resyllabified words in French. *Journal of Memory and Language*, 48, 233–254.

[7] Rialland, A. 1986. Schwa and syllables in French. *Studies in compensatory lengthening*, *23*, 187.

[8] Fougeron, C., Bagou, O., Stefanuto, M., & Frauenfelder, U. H. 2003. Looking for acoustic cues of resyllabification in French. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPHS)*.

[9] Fougeron, C. 2007. Word boundaries and contrast neutralization in the case of enchaînement in French. *Papers in laboratory phonology IX: Change in phonology*, 609-642.

[10] Nguyen, N., Wauquier-Gravelines, S., Lancia, L., & Tuller, B. 2007. Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications. *In Pilar Prieto. Segmental and Prosodic Issues in Romance Phonology, John Benjamins, pp.3-23, 2007, Current Issues in Linguistic Theory.*

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

[12] Shaw, J. A., Oh, S., Durvasula, K., & Kochetov, A. (2021). Articulatory coordination distinguishes complex segments from segment sequences. *Phonology*, *38*(3), 437-477. doi:10.1017/S0952675721000269.

[13] Bürkner, P.-C. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

[14] R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Speech and Sign: The Whole Human Language

Wendy Sandler
The University of Haifa

Traditional historical and comparative linguistics emphasized differences across spoken languages and language families. Chomskyan generative linguistics caused a paradigm shift by emphasizing universal properties, thought to be innate, and minimizing the role of the body, dubbed “externalization” [1].

Both approaches are called into question by natural sign languages. In the absence of auditory input, humans inevitably create an alternative sign language system, a system exhibiting impressive formal universals that are indeed similar to those of spoken languages [2]. But that is not the whole story.

I will demonstrate certain fundamental, predictable, and nontrivial differences between spoken and signed languages, most notably, a direct correspondence between bodily articulations and linguistic functions [3,4] . The striking difference between self-organization of speech articulators in the vocal tract, and self-organization of visible parts of the body -- each associated cognitively with a linguistic function (Figure 1) -- will be presented. I will argue that it is precisely these differences that shed light on the nature of the whole human language in our species.

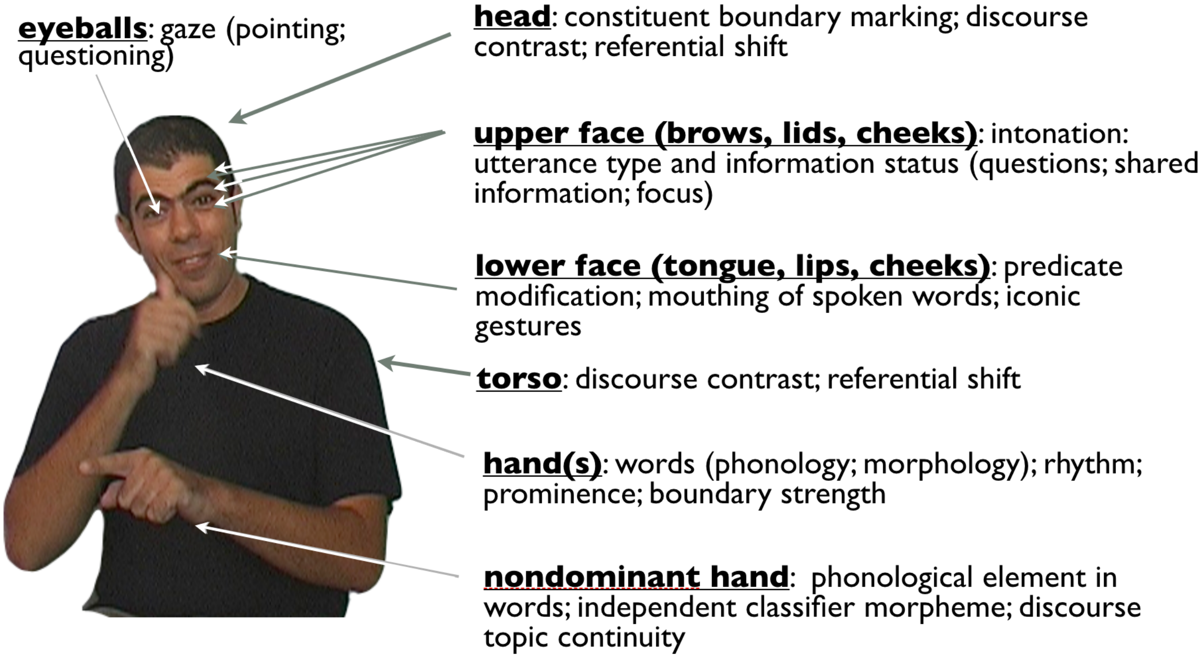

eyeballs: gaze (pointing; questioning)
head: constituent boundary marking; discourse contrast; referential shift
upper face (brows, lids, cheeks): intonation: utterance type and information status (questions; shared information; focus)
lower face (tongue, lips, cheeks): predicate modification; mouthing of spoken words; iconic gestures
torso: discourse contrast; referential shift
hand(s): words (phonology; morphology); rhythm; prominence; boundary strength
nondominant hand: phonological element in words; independent classifier morpheme; discourse topic continuity

Figure 1. Correspondence between bodily articulations and linguistic functions in sign languages

References

[1] Chomsky, N. (2007). Of minds and language. *Biolinguistics. 1*, 009-027.

[2] Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge University Press.

[3] Sandler, W. (2012). Dedicated gestures in the emergence of sign language. *Gesture. 12(3*), 265-307.

[4] Sandler, W. (2018). The body as evidence for the nature of language. *Frontiers in Psychology 9/1782.* 2-21. doi: 10.3389/fpsyg.2018.01782

Speakers, listeners, languages: Coarticulatory variability and contrast in spoken language dynamics

Marianne Pouplier[1]

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

In this talk, I will report on our recent work which seeks to understand to which extent coarticulatory variability in speech production and perception may be shaped by phonological contrast and, at the same time, vary between individuals of the same language community. Language-specific sound systems have been argued to be influential factors governing the degree of coarticulation, and the maintenance of phonological contrast is expected to restrict the temporal extent of articulatory anticipation [1-4]. Yet evidence for the intuitively appealing role of contrast has remained ambiguous [5, 6]; variation between languages does not necessarily pattern as predicted by phonological structure. For nasality, the presence of contrast may indeed constrain coarticulation (e.g., French), or the absence of contrast may enable extensive coarticulation (e.g., English), yet for other languages, nasal coarticulation has been found to be either unexpectedly limited in the absence of a contrast (e.g., Spanish [4]), or unexpectedly extensive despite the presence of a contrast (e.g., Lakota [7]). For labial coarticulation, the picture is similarly complex [8, 9], albeit there being far fewer studies.

Another, related point of discussion in coarticulation research is the role of individual variation. While the individual variation that is observable among speakers of the same language (e.g., [10]) has often been set aside as experimental 'noise', recent work particularly on nasality has put this individual variation at the heart of explanations for sound change. Individuals within a language community may differ systematically in how they produce and perceive coarticulation, rather than contrast acting as a constraining force on coarticulation in a monolithic fashion [11]. The implications of this shift in perspective on variability within a language for our understanding of between-language variation in coarticulation are yet to be explored.

Due to small sample sizes in many studies, our knowledge about the nature of language specific effects on coarticulation does not always stand on firm ground. Indeed, Noiray et al. [8] contested the notion that coarticulation is language specific at all: their study on anticipatory lip rounding in English and Canadian French, yielded no evidence of systematic patterns by language, despite French, but not English contrasting rounding on vowels. They therefore concluded that the implementation of anticipatory lip rounding is purely speaker-dependent, contra older work on labial coarticulation by Lubker & Gay [9] .

To broaden our understanding of how the scope of anticipatory coarticulation may interact with a language's phonology, how it is planned, and perceived, we have undertaken a large-scale comparative study on contextual vowel nasalization as well as lip rounding in three languages. We recorded nasalance and 'blue lip' video data [12] for French, German, and American English for 25-30 speakers per language [13, 14]. For English, neither nasality nor rounding is contrastive for vowels, for French, both are contrastive whereas German contrasts lip rounding only. This allows us to study, for the same speakers, the coarticulatory behavior for two independent articulators, in the presence or absence of a phonological contrast (across languages). Production data will be complemented by perception experiments probing the production-perception link.

For nasality, our results so far confirm the expectation of language specific effects: French is most limited in its coarticulatory scope, whereas English has the greatest temporal extent of anticipatory nasalization. German falls between the two other languages, differing from neither one [15]. For all languages we find a range of individual variation: some speakers show an extensive temporal range of anticipatory coarticulation while others coarticulate comparatively little. First analyses suggest that speakers who coarticulate most extensively for nasality do not necessarily do so for lip rounding, meaning there is no general speaker-specific coarticulatory setting.

For lip rounding, coarticulation is, perhaps surprisingly, similarly extensive in both English and German with a high degree of individual variation. Our results align with Noiray et al.'s previous observation for French vs. English that individual variation dominates any group-level effects. Our data analysis for lip rounding in French is still in progress.

Overall, our results thus resonate with other studies which cast doubt on phonological contrast being a good predictor of the temporal extent of coarticulation (among others, [7, 16]). Within each language, coarticulation is both speaker and articulator specific. Language-specific effects, to the extent present, do not necessarily align with phonological contrast.

Work supported by DFG grant PO 1269/5-1

References

[1] Öhman, S.E., Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 1966. **39**(1): p. 151-168.

[2] Manuel, S., The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, 1990. **88**(3): p. 1286-1298.

[3] Delvaux, V., D. Demolin, B. Harmegnies and A. Soquet, The aerodynamics of nasalization in French. *Journal of Phonetics*, 2008. **36**(4): p. 578-606.

[4] Solé, M.-J., Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech*, 1995. **38**: p. 1-23.

[5] Mok, P., Language-specific realizations of syllable structure and vowel-to-vowel coarticulation. *Journal of the Acoustical Society of America*, 2010. **128**(3): p. 1346-1356.

[6] Mok, P., Does vowel inventory density affect vowel-to-vowel coarticulation? *Language and Speech*, 2012. **56**(2): p. 191-209.

[7] Scarborough, R., G. Zellou, A. Mirzayan and D.S. Rood, Phonetic and phonological patterns of nasality in Lakota vowels. *Journal of the International Phonetic Association*, 2015. **45**(3): p. 289-309.

[8] Noiray, A., M.-A. Cathiard, L. Ménard and C. Abry, Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *The Journal of the Acoustical Society of America*, 2011. **129**(1): p. 340-349.

[9] Lubker, J. and T. Gay, Anticipatory labial coarticulation: Experimental, biological, and linguistic variables. *The Journal of the Acoustical Society of America*, 1982. **71**(2): p. 437-448.

[10] Grosvald, M., Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, 2009. **37**(2): p. 173-188.

[11] Beddor, P.S., A.W. Coetzee, W. Styler, K.G. McGowan and J.E. Boland, The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, 2018. **94**(4): p. 931-968.

[12] Lallouache, M.T., *Un poste "Visage-parole" couleur. Acquisition et traitement automatique des contours des lèvres [A "face-speech" interface. Automatic acquisition and processing of labial contours]*. 1991, PhD thesis, Institut National Polytechnique de Grenoble.

[13] Lo, J.H., C. Carignan, M. Pouplier, F. Rodriquez, R. Alderton, B.G. Evans, and E. Reinisch, Language specificity vs speaker variability of anticipatory labial coarticulation in German and English. *Proceedings of the 20th International Congress of Phonetic Sciences, Prague*, 2023.

[14] Rodriquez, F., M. Pouplier, R. Alderton, J.H. Lo, B.G. Evans, E. Reinisch, and C. Carignan, What French speakers' nasal vowels tell us about anticipatory nasal coarticulation. *Proceedings of the 20th International Conference of the Phonetic Sciences, Prague*, 2023.

[15] Pouplier, M., F. Rodriquez, R. Alderton, J.H. Lo, B.G. Evans, E. Reinisch, and C. Carignan, The window of opportunity: Anticipatory nasal coarticulation in three languages. *Proceedings of the 20th International Conference of the Phonetic Sciences, Prague*, 2023.

[16] Beddor, P.S., J.D. Harnsberger and S. Lindemann, Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, 2002. **30**: p. 591-627.

Syllable position in secondary dorsal contrasts: an ultrasound study of Irish

Ryan Bennett[1], Jaye Padgett[1], Máire Ní Chiosáin[2], Grant McGuire[1], Jennifer Bellik[1]

[1]University of California, Santa Cruz (USA), [2]University College Dublin (Ireland)
padgett@ucsc.edu, rbennett@ucsc.edu, maire.nichiosain@ucd.ie, gmcguir1@ucsc.edu, jbellik@ucsc.edu

This study examines the articulatory robustness of secondary dorsal /Cʲ Cˠ/ contrasts in Irish, across different word/syllable positions, using ultrasound imaging. We find that /Cʲ Cˠ/ contrasts are more articulatorily distinct in onset position than in coda position, and speculate that syllable-based differences in the articulation of /Cʲ Cˠ/ may help explain why /Cʲ Cˠ/ contrasts are preferentially realized in onset position across languages.

Every consonant in Irish is either contrastively palatalized /Cʲ/ or contrastively velarized /Cˠ/. These /Cʲ Cˠ/ contrasts occur both word-initially and word-finally (1). Word-final /Cʲ Cˠ/ contrasts can also mark morphosyntactic distinctions, such as plural vs. singular inflection (2).

(1)	/bʲɔːnˠ/ ‘peak’	/bˠɔːnˠ/ ‘white’
	/pʲɔːnˠ/ ‘pen’	/pˠɔːnˠ/ ‘pawnshop’
	/bˠrˠɔːdʲ/ ‘neck’	/bˠrˠɔːdˠ/ ‘drizzle’
	/sˠkˠɔːlʲ/ ‘shadow’	/sˠkˠɔːlˠ/ ‘supernatural being’
(2)	/kˠatʲ/ ‘cats’	/kˠatˠ/ ‘cat’
	/bˠɔːdʲ/ ‘boats’	/bˠɔːdˠ/ ‘boat’

Work on the typology of /Cʲ C⁽ˠ⁾/ contrasts has shown that such contrasts are more susceptible to loss in word/syllable-final position [1, 2], particularly for labials. Word-final /Cʲ C⁽ˠ⁾/ contrasts seem to be less perceptible than word-initial /Cʲ C⁽ˠ⁾/ contrasts in both Russian and Irish [3-6], possibly due to differences in the availability and robustness of acoustic cues in each of these contexts [7]. However, there is relatively little work examining possible articulatory bases for these perceptual asymmetries, and none on Irish. Kochetov [2, 8] found that the palatalization gesture of [pʲ] is reduced and differently timed in onset position compared to coda position. These articulatory differences may contribute to the perceptual and typological asymmetries noted above regarding /Cʲ C⁽ˠ⁾/ contrasts across word/syllable contexts.

Our study considers comparable contrasts in Irish. We test the hypothesis that sound changes affecting /Cʲ C⁽ˠ⁾/ contrasts, and the resulting typology, stem from patterns of articulatory reduction and coordination, likely working in tandem with perceptual asymmetries across syllabic contexts. Specifically, we expect that dorsal positions reflecting /Cʲ Cˠ/ contrasts will show less articulatory separation in coda (word-final) compared to onset (initial) contexts.

We've collected ultrasound data from 7 Irish speakers, representing all major dialects (Ulster, Connacht, Munster). We present data from 4 speakers here, and will analyze data from the remaining 3 speakers prior to the conference. Speakers uttered 5 repetitions of a list of C-initial and C-final Irish words. Target consonants were all stops (labial, coronal, velar), paired for secondary articulation (/Cʲ/ vs. /Cˠ/), syllabic position (onset vs. coda), and vowel context (adjacent to [iː], [uː], or [ɔː]). All target consonants were in word-initial stressed syllables; target onsets were always word-initial, and target codas always word-final. In each pass through the list, words were presented in random order, embedded in the carrier phrase [ˈdˠuːrtʲ ˈiːfˠə ___ əˈnˠuˠrə] ‘Aoife said ___ last year’. Ultrasound data was collected using a Terason T3000 ultrasound system with a model 8MC3 probe, mounted in an Articulate Instruments Ultrasound Stabilization Headset [9], at 60 frames/second. The tongue surface in these images was traced with EdgeTrak [10] (Fig. 1).

We assess dorsal position in target consonants in three ways: using loess-smoothed curves; comparing the position of the highest point of the tongue dorsum; and computing Root Mean Sum of Squared Distances between two curves [11] (Fig. 2).

All three measures find that /C^{j} C$^{(ɣ)}$/ contrasts are more widely separated in word-initial (onset) position than in word-final (coda position). This is especially true for labials and dorsals. Consonants show more coarticulation with neighboring vowels when in coda position; this is again especially true for labials. These observations hold whether we compare onsets vs. codas at C release, or instead compare onsets at CV transition with codas at VC transition. We conclude that typological asymmetries in the distribution of /C^{j} C^{ɣ}/ contrasts are reflected in articulatory asymmetries in the production of these contrasts in the synchronic phonetics of Irish.

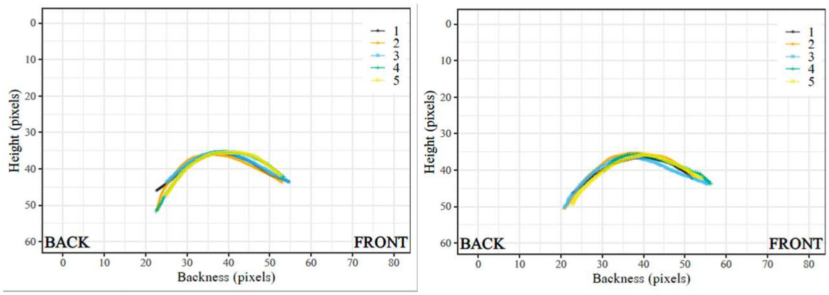


Fig. 1: tracings for [b^{j}] in word-initial (left) vs. final (right) position adjacent to [uː], C offset.

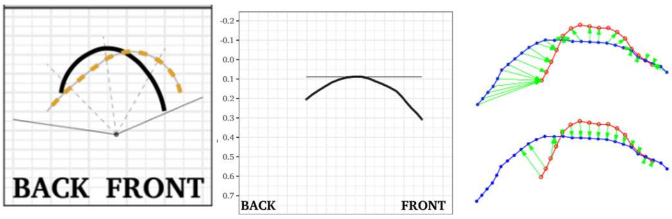


Fig. 2: loess-smoothed comparisons; peak dorsal position; RMSSD measures (Csapó et al. 2017)

References

[1] Takatori, Y., 1997. *A study of constraint interaction in Slavic phonology*. Yale University PhD.

[2] Kochetov, A., 2002. *Production, perception, and emergent phonotactic patterns*. Routledge.

[3] Kochetov, A., 2004. Perception of place and secondary articulation contrasts in different syllable positions: language-particular and language-independent asymmetries. *Language and speech* 47.4.

[4] Kochetov, A., 2006a. Testing licensing by cue. *Phonetica* 63.

[5] Ní Chiosáin, M. & J. Padgett, 2012. An acoustic and perceptual study of Connemara Irish palatalization. *Journal of the International Phonetic Association* 42.2.

[6] Padgett, J. & M. Ní Chiosáin, 2018. The perception of a secondary palatalization contrast: a preliminary comparison of Russian and Irish. In R. Bennett et al. (eds.), *Hana-bana: a festschrift for Junko Ito and Armin Mester*.

[7] R. Wright, 2004. A review of perceptual cues and cue robustness. In B. Hayes et al. (eds.), *Phonetically based phonology*. Cambridge University Press.

[8] Kochetov, A., 2006b. Syllable position effects and gestural organization: Evidence from Russian. In: L. Goldstein et al. (eds.), *Papers in Laboratory Phonology VIII*. Mouton de Gruyter.

[9] Wrench, A., 2008. *Articulate Assistant user guide, version 1.17.*

[10] Li, M. et al., 2005. Automatic contour tracking in ultrasound images. *Clinical linguistics and phonetics* 19.6-7.

[11] Csapó, T. et al., 2017. Comparison of distance measures in tongue contour traces of ultrasound images. Poster presented at Ultrafest 2017.

An acoustic and articulatory study on variation of high vowel devoicing across prosodic contexts and speakers in Korean

Jungyun Seo[1], Sahyang Kim[2,4] & Taehong Cho[3,4]
[1]University of Michigan (USA), [2]Hongik University (Korea), [3]Hanyang University (Korea), [4]HIPCS
sjungyun@umich.edu, sahyang@gmail.com, tcho@gmail.com

High vowels, particularly when surrounded by voiceless consonants, often undergo devoicing. This phenomenon can be attributed to physiological influences stemming from the tongue position during the production of high vowels [1]. Robust patterns of high vowel devoicing are observed in Japanese (e.g., [2]). In Japanese, high vowel devoicing is generally viewed as a phonological process whereby the [+voice] feature becomes disassociated and assimilated to the [-voice] feature of neighboring consonants (e.g., [3]), as evident in completely devoiced tokens that lack any trace of the vowel (e.g., [4]). In contrast, Korean devoicing is often described as a phonetic process, which can be considered to be due to the overlap between the glottis abduction gesture for voiceless consonants and the voicing gesture for the vowel ([5, 6]). Greater devoicing occurs when the overlap between these gestures is more extensive. Furthermore, the position of a vowel within a phrase has been found to exert influence on devoicing. For instance, completely devoiced tokens are more commonly found in phrase-initial positions as opposed to phrase-medial positions. This distinction can be accounted for by the larger overlap caused by the expanded glottis opening gesture at the domain-initial position, driven by the heightened [+spread glottis] feature due to domain-initial strengthening [7].

The present study aims to delve deeper into the nature of high vowel devoicing in Korean by examining the correlation between high vowel devoicing and tongue height. This investigation sheds new light on how this phonetic process relates to the physiological constraints imposed by tongue height. Furthermore, the study incorporates speaker variation to determine whether high vowel devoicing is an automatic process or one controlled by individual speakers. In doing so, the study explores variations in high vowel devoicing across different positions and in contexts where focus-induced prominence occurs. This exploration is important not only because devoicing patterns can be influenced by prosodic structural factors such as position and prominence but also because it enriches the contexts in which variation in high vowel devoicing can be observed.

An acoustic and articulatory study was conducted with 13 Seoul Korean speakers to investigate high vowel devoicing. The target words (p^hip^ha, p^hip*a) consisted of a high vowel /i/ surrounded by voiceless consonants, and they were produced in various prosodic structural contexts. Note that the same test word occurred in both IP-initial and IP-medial positions, as well as in focused and unfocused conditions. The focused condition involved contrasting the two target words, as presented in Table 1. Devoicing proportion was determined by calculating the ratio of acoustic duration between the voiceless portion and the entire syllable in the first syllable of CV.CV target words. Furthermore, tongue height (maxima during the vocalic movement) was measured using Electromagnetic Articulography (EMA) for the same set of target syllables. A total of 1483 tokens were examined, comprising 2 test words, 15 repetitions, 2 positions, 2 focus conditions, and 13 speakers.

The results revealed a gradient distribution of devoicing proportion (Fig. 1a), highlighting the phonetic nature of the devoicing process. Notably, completely devoiced tokens (20% of the data) were more prevalent in prosodically weak positions, indicating their susceptibility to coarticulatory effects from neighboring voiceless consonants. However, there was significant variability in devoicing patterns across speakers (Fig. 1b), suggesting individual differences in coarticulatory influences. Importantly, the correlation between devoicing proportion and tongue height was not clearly established ($\rho = 0.089$, $p < 0.001$), indicating that tongue height does not directly impact devoicing. This lack of correlation can be attributed to speaker variation. For instance, Fig. 2 illustrates speaker-specific effects of focus on tongue height and devoicing. Some speakers (Fig. 2c) exhibited increased tongue height with no effect on devoicing under focus, while others (Fig. 2b) showed unexpected patterns where heightened tongue position coincided with decreased (rather than increased) devoicing. These findings suggest that higher tongue position does not consistently induce more devoicing due to biomechanical factors. It appears that contrastive focus may enhance the [high] feature of the high vowel, potentially reinforcing the voicing feature. Yet, some speakers utilized both features, while others selectively suppressed devoicing under focus (Fig. 2a), and some did not demonstrate consistent modulation of tongue height and devoicing in relation to featural enhancements. (Position effects also exhibited speaker variation not relying on tongue height, although it is not discussed in detail here due to space limitations.)

In conclusion, the findings suggest that high vowel devoicing in Korean exhibits a gradient phonetic process. However, the presence of speaker variation indicates that individual speakers adjust devoicing based on linguistic factors such as prosodic structure and phonological constraints, rather than relying solely on tongue height. Nevertheless, devoicing is more prevalent in prosodically weak positions, and certain speakers demonstrate a stronger inclination for devoicing. A broader implication is that if the number of speakers demonstrating robust devoicing increases, it has the potential to initiate sound changes, resulting in a more categorical devoicing process similar to what is observed in Japanese.

Table 1. Carrier sentences. Target words are underlined and a contrastive focus falls either on the target word or on the word that immediately follows it. Corrective contrast information was presented with the bold characters in the sentence.

Word	Boundary	Focus	Question sentence	Target sentence
피파 pʰipʰa	Phrase-initial	Focused	이번 단어는 삐빠뒤에 놔? [ipʌn tanʌnɯn p*ip*atwinoa?] Should I put the word behind p*ip*a this time?	아니야, **피파**뒤에 놔. [aniya.] [**pʰipʰa**twienoa.] No, put it behind **pʰipʰa**
		Unfocused	이번 단어는 피파 앞에 놔? [ipʌn tanʌnɯn pʰipʰaapenoa?] Should I put the word in front of pʰipʰa this time?	아니야, 피파**뒤**에 놔. [aniya.] [pʰipʰa**twi**enoa.] No, put it **behind** pʰipʰa
	Phrase-medial	Focused	이번 단어는 누나 삐빠 뒤에 놔? [ipʌn tanʌnɯn nunap*ip*atwienoa?] Should I put the word behind sister's p*ip*a this time?	아니야, 누나**피파**뒤에. [aniya.] [nuna**pʰipʰa**twie.] No, put it behind sister's **pʰipʰa**.
		Unfocused	이번 단어는 누나 피파 앞에 놔? [ipʌn tanʌnɯn nunapʰipʰaapenoa?] Should I put the word in front of sister's pʰipʰa this time?	아니야, 누나피파**뒤**에. [aniya.] [nunapʰipʰa**twi**e.] No, put it **behind** sister's pʰipʰa.

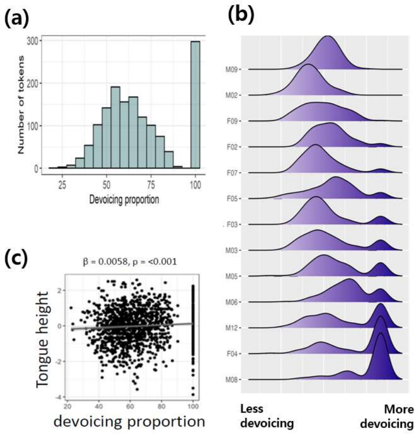


Fig.1 (a) Distribution of devoicing with 100 on the x axis indicating complete devoicing; (b) variation in distribution of devoicing across speakers ; (c) corelation between tongue height and devoicing.

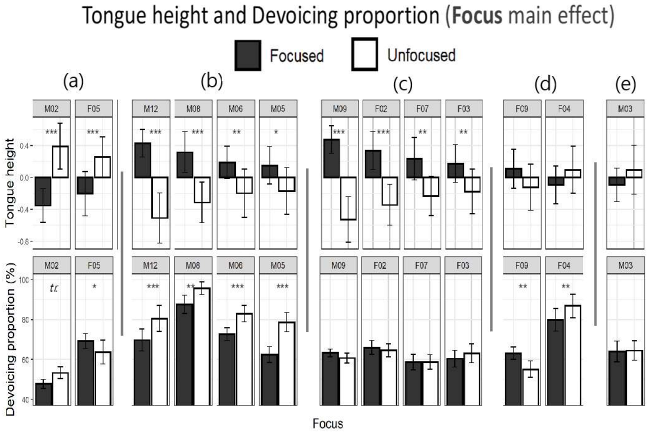


Fig.2 Prosodic modulation driven by Focus type in tongue height (upper panels) and devoicing proportion (lower panels) per speaker (***, **, * and tr. refer to $p < 0.001$, $p < 0.01$, $p < 0.05$, and $0.05 < p < 0.06$ in statistical analyses, respectively). Note that group (c) showed a higher tongue position in the focused condition compared to the unfocused condition without changing devoicing proportion; group (b) showed the focus effect only in the devoicing proportion ; and group (d) showed both a higher tongue position and less devoicing (more voicing) under focus.

References

[1] Ladefoged, P. (1973). The features of the larynx. *Journal of phonetics*, *1*(1), 73-83.

[2] Vance, T. J. (2008). The sounds of Japanese with audio CD. Cambridge University Press.

[3] Tsuchida, A. (2001). Japanese vowel devoicing: Cases of consecutive devoicing environments. *Journal of East Asian Linguistics*, 10(3), 225–245

[4] Beckman, M. E., & Shoji, A. (1984). Spectral and perceptual evidence for CV coarticulation in devoiced/si/and/syu/in Japanese. *Phonetica*, 41(2), 61–71.

[5] Jun, S. A., & Beckman, M. E. (1993). A gestural-overlap analysis of vowel devoicing in Japanese and Korean. Paper presented at the 67th annual meeting of the Linguistic Society of America, Los Angeles.

[6] Jun, S. A., & Beckman, M. E. (1994). Distribution of devoiced high vowels in Korean. In Third International Conference on Spoken Language Processing.

[7] 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.

Contrast enhancement and the distribution of vowel duration in Japanese

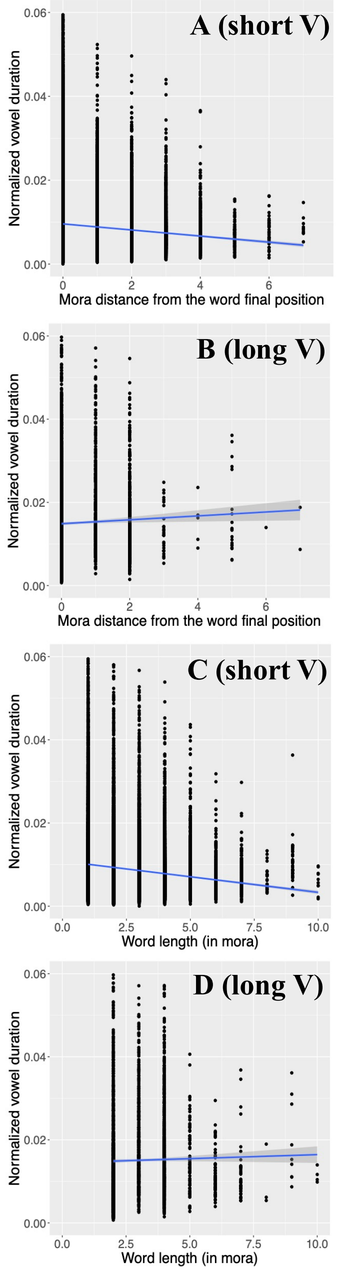
Shinichiro Sano[1], Céleste Guillemot[2]

[1, 2]Keio University (Japan)

shinichirosano@gmail.com, celesteguillemot@gmail.com

Background: Previous research has shown that patterns in phonetic implementation of segments and subsegmental features are controlled by information-related factors [1-5]. Phonetic cues that contribute less information are more prone to undergo reduction or neutralization [1,2,5]. This is illustrated by the crosslinguistic tendency for phonological processes involving neutralization to exhibit a preference for word-ends over beginnings [5,6]. On the other hand, words with low predictability tend to be longer [7], and their segments need to convey more disambiguating information [8].

Vowel duration provides useful test cases to consider the role of information: [9,10] demonstrated that, in English, more predictable or less informative vowels are shorter. In Japanese, in which vowels contrast in length (short vs. long, [11,12]), [13] show that preceding consonant and information-related measures (Surprisal and Entropy) play a role in variations of vowel duration at the sub-phonemic level. Building upon previous research, this paper focuses on the vowel length contrast to examine (i) how the positional bias and difference in intonation phrase (IP)/word length is reflected in the distribution of vowel duration in Japanese, (ii) how the distribution differs depending on the type of linguistic unit, and (iii) how these are related to the vowel length contrast.



Method: Data were retrieved from the CSJ-RDB (Corpus of Spontaneous Japanese – Relational Database, National Institute for Japanese Language and Linguistics 2012), among which the present study targeted 12 speech samples. An exhaustive search of the data in the CSJ- RDB resulted in 44,219 tokens, of which 40,703 (92%) were short vowels and 3,516 (8%) were long vowels, where tokens with filled pauses, word fragments, and other non-linguistic events (e.g., laughter) were excluded. The duration of each token was analyzed in terms of position in IP and word, and length of IP and word. All distributional skews discussed below were tested by the linear mixed-effects model using *lmer* of the lmerTest package in R [14]. We fit separate models for short and long vowels. In the models, response variable was duration of short/long vowels normalized by speech rate (duration of IP divided by the number of moras); we included factors of interests (position in IP/word and length of IP/word) and other control variables (e.g., kinds of vowels, accented or not); random intercepts for speaker and item (lemma) and by-speaker and by-item random slopes were also included in the model.

Results and discussion: **[Position]** We measured positions of vowels in IP by word distance from the IP-final position and in word by mora distance from the word-final position. At the IP level, short and long vowels showed the same pattern: duration is longer at more back positions than more front positions (short V: t = -11.309, p < 0.01, long V: t = -3.7, p < 0.01). This can be attributed to the effect of final lengthening, which occurs in utterance-final and phrase-final position, but almost never in word-final position in Japanese [15,16]. At the word level, however, short and long vowels showed different patterns. At more front positions (rightward in A and B), short vowels become shorter (t = -11.348, p < 0.01), while long vowels become longer (t = 2.954, p < 0.01), resulting in a larger durational gap between them that provides enhanced cues for short vs. long contrast. This suggests contrastive hyperarticulation at informationally salient positions. At more back positions (leftward), short vowels are longer, while long vowels are shorter, making the durational difference between them closer. As a result, the durational distinction is more likely to be neutralized, which is consistent with the fact that back positions are informationally non-salient [5]. **[Length]** We measured IP length by word count and word length by mora count. At the IP level, short and long vowels showed the same pattern. The duration is longer when the IP is shorter (short V: t = -3.352, p < 0.01, long V: t = -2.566, p <

0.01). This may be due to physiological reasons: a limitation of breath entails a limited IP duration; hence when an IP is longer, each segment becomes shorter (cf. Respiratory Code for f0; [17,18]). On the contrary, at the word level, distinct patterns were again observed in short and long vowels, as C and D illustrate. In longer words, short vowels become shorter (t = -4.900, p < 0.01), while long vowels become longer (t = -3.011, p < 0.01), making the durational distance between short and long vowels greater, that is, enhanced cues for short vs. long contrast in longer words. This may be due to the lexical distribution of shorter and longer words. In token frequency, shorter words are more frequent than longer words (59.8% and 59.2% of all words are less than two (for short V) and three moras (for long V)). Since shorter words are more frequent and predictable, phonetic signal in these words tend to be phonetically reduced (probabilistic reduction, [4,9]), while longer words are less frequent and less predictable, and therefore phonetic signal in these words should be enhanced ([8]). In type frequency, however, longer words are more frequent than shorter words (75.8% and 61.4% of all words are more than three moras (for short V) and four moras (for long V)). With more lexical competitors, in longer words the predictability with which a target segment is identified becomes lower, and thus requires the phonetic signal to be more informative or salient to differentiate the target from other competitors.

The results suggest that, at the word level, duration is effectively controlled (enhanced cues for salient positions and words with less predictability or more competitors, and reduced cues for non-salient positions and words with more predictability or less competitors) to give appropriate degree of speech signal to balance the successful transmission of lexical information and the cost for phonetic implementation. However, this is not the case with IP, which contributes to sentence-level information (e.g., intonation). In addition, hyperarticulation (reduction or enhancement) not only targets a particular linguistic unit independently, making it shorter or longer, but also a contrast in such a way as to increase the durational distance between contrasting segments.

References

[1] Hall K. C., Hume E., Jeager F. & Wedel A. (2016). *The message shapes phonology*. ms.

[2] Hall K. C., Hume E., Jeager F. & Wedel A. (2018). The role of predictability in shaping phonological patterns. Linguistics Vanguard, 4(S2), 20170027.

[3] Cohen-Priva U. (2015). Informativity affects consonant duration and deletion rates. *LabPhon*, 6(2), 243-278.

[4] Turnbull R. (2018). Effects of lexical predictability on patterns of phoneme deletion/reduction in conversational speech in English and Japanese. Linguistics Vanguard, 4(S2), 20170033.

[5] Wedel A., Ussishkin A. & King A. (2019). Crosslinguistic evidence for a strong statistical universal: Phonological neutralization targets words-ends over beginnings. Language, 95(4), e428-e446.

[6] Houlihan K. (1975). *The role of word boundary in phonological processes*. Doctoral dissertation. The University of Texas at Austin.

[7] Zipf G. K. (1935). The psycho-biology of language. Boston: Houghton Mifflin.

[8] King A., & Wedel, A. (2020). Greater Early Disambiguating Information for Less-Probable Words: The Lexicon Is Shaped by Incremental Processing. *Open Mind: Discoveries in Cognitive Science*, 1-12.

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

[10] Aylett M. & Truk A. (2006). Language redundancy predicts syllable duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America,* 119(5), 3048-3059.

[11] Vance T. (1987). *An introduction to Japanese phonology*. New York: SUNY Press.

[12] Vance T. (2008). *The sounds of Japanese.* Cambridge: Cambridge University Press.

[13] Shaw J. & Kawahara S. (2017). Effects of surprisal and entropy on vowel duration in Japanese. *Language and Speech*, 62, 80-114.

[14] R Core team (2019). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

[15] Hoequist C. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, 40(3), 203-237.

[16] Takeda K., Sagisaka Y. & Kuwabara H. (1989) On sentence-level factors governing segmental duration in Japanese. *The Journal of the Acoustical Society of America,* 86, 2081.

[17] Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. *Proceedings of Speech Prosody* 2002, 47-57.

[18] Gussenhoven C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science*, 8(2), 425-434.

Lexicons and phonologies co-evolve under pressure from incremental word processing

Andy Wedel
University of Arizona

Over the last century, much evidence has accumulated suggesting that language structures evolve under pressure to maintain a balance between effort and communication accuracy. However, both 'effort' and 'accuracy' are constrained by the particular processes involved in encoding and decoding the speech signal. The work described here concerns adaptations of the lexicon and phonology of a language to the fact that listeners process the phonetic signal incrementally. When listeners hear the beginning of a word, they begin narrowing down hypotheses about what word they are hearing from the very first segment, rather than waiting until the end of the word. As a consequence of this incremental processing, early segments contribute more disambiguating information than later segments. Independent evidence suggests that speakers hyperarticulate phonetic cues that distinguish them from near lexical neighbors [1]. Putting these two facts together, we expect lexicons should evolve so that informative segments are concentrated early in words, because that is where they can do the most work in disambiguation. Conversely, less effortful combinations of sounds - which are often less informative - should be more common near ends of words because there, expenditure of greater effort does little to increase information transmission accuracy. Recent evidence shows that this is the case [2].

Turning to phonology, we expect that the ability of early phonemes in a word to provide more disambiguating information should inhibit the development phonological rules that reduce information word-initially. As a way to investigate this general hypothesis, we assembled a set of geneologically and areally balanced set of 50 languages and asked whether contrast-neutralizing rules are significantly more common at word ends than beginnings. For example, a common type of neutralizing rule is 'word-final obstruent devoicing', in which the voicing contrast in obstruents is neutralized at the ends of words. We find that indeed, neutralizing rules are significantly more frequent word-finally than word-initially (Figure 1). This pattern was found in every language family and in every language area in the dataset, as we would expect if the asymmetry is due to a basic fact about human language processing [3].

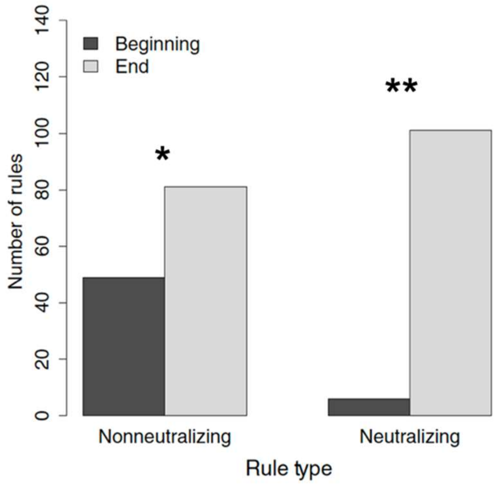
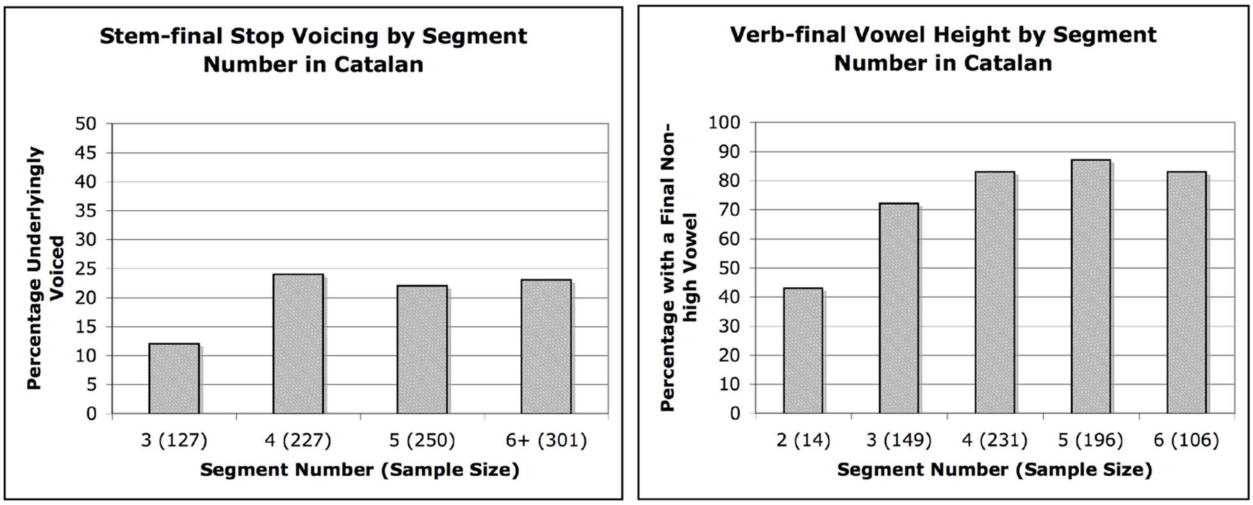


Figure 1. The left pair of bars shows the number of non-neutralizing rules in the dataset that apply to word beginnings versus ends. Non-neutralizing rules are ~2 times more likely to apply to the ends of words than the beginnings. The right pair of bars shows the number of neutralizing rules applying to word beginnings versus ends. Neutralizing rules are ~20 times more likely to apply to word ends. This word-edge difference is significantly greater than the difference for non-neutralizing rules.

Recall that we think that neutralization rules are more common word-finally because segment contrast late in the word is less informative. But what about very short words? In short words the end of the word is very close to the beginning, so final segments are more informative. Given that a language has a final-neutralization rule in its phonology, how might the lexicon evolve to respond? A way that a word can avoid the effects of a final-neutralization rule in the phonology is to end with a segment that does not alternate. In new data expanding on [4] we focus on neutralizing phoneme pairs in which one neutralizes to the other, as in final obstruent devoicing. In this dataset we find that short words are less likely to end with the alternating member of a final neutralization, exemplified below with Catalan, which shows both final devoicing and vowel-height neutralization in unstressed syllables.

Figure 2. The first panel shows the percentage of stems in Catalan with underlyingly voiced word-final obstruents by word length. The second panel shows the percentage of stems that have in a non-high vowel in the final syllable, by word length. In Catalan, non-high vowels neutralize to a higher vowel in unstressed syllables; for example /e, o/ neutralize to [i, u]. In both cases, shorter stems are significantly less likely to end with a phoneme whose surface form neutralizes to that of another phoneme.



These observations suggest that phonologies and lexicons co-evolve under pressure from incremental processing, and further under pressure from each other. Here, we show data consistent with a model in which incremental processing in concert with relative hyperarticulation of high-information cues inhibits the development of word-initial neutralization rules. In turn the lexicon evolves such that most short words end in word-final segments that do not alternate, which has the effect of localizing word-final neutralization in longer words where word-final segments are least informative.

References

1. Wedel, Andrew, Noah Nelson & Rebecca Sharp (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language* 100: 61-88.
2. King, Adam & Wedel, Andrew. (2020) Greater Early Disambiguating Information for Less-Probable Words: The Lexicon Is Shaped by Incremental Processing. *Open Mind* 4(1): 1-12
3. Wedel, Andrew, Ussishkin, Adam, King, Adam (2019). Crosslinguistic evidence for a strong statistical universal: Phonological neutralization targets word-ends over beginnings. *Language* 95(4): e428-e446.
4. Ussishkin, A., & Wedel, A. (2009). Lexical access, effective contrast and patterns in the lexicon. *Perception in phonology*, 267-92.

Acoustic and linguistic influences on f0 imitation

Kuniko Nielsen[1], Rebecca Scarborough[2]

[1]*Oakland University (USA),* [2]*University of Colorado Boulder (USA)*
nielsen@oakland.edu, rebecca.scarborough@colorado.edu

Studies in phonetic imitation have shown that speakers imitate some phonetic patterns to which they are exposed [1, 2]. However, it is still unclear what aspects of the speech signal speakers are responding to when they change their speech behavior: a specific acoustic value or a linguistically-interpreted target. To address this issue, we conducted two online pitch imitation experiments: one in which participants were exposed to a linguistically-unmarked overall pitch difference (in this case, low f0) and one in which participants were exposed to a linguistically-salient manipulated pitch accent realization (in this case, extra high H in L+H*). If imitation targets the specific acoustic value of a model talker's naturally low pitch, we expect participants to converge acoustically toward the talker's low f0 in both experiments. On the other hand, if imitation targets linguistic patterns, participants should imitate the linguistically meaningful pattern, i.e., the high contrastive focus pitch accent, even if it results in acoustic divergence from the model talker's generally low f0.

Both experiments included 4 blocks: 1) baseline, in which American English speaking participants produced sentences based on information presented on-screen; 2) exposure, in which participants heard stimulus sentences presented auditorily; 3) shadowing, in which the participants repeated sentences presented auditorily; and 4) post-test, which was like the baseline task. The model talker whose speech was presented in the exposure and shadowing blocks was a male with a naturally low f0 (mean=101Hz in carrier phrases). In *Experiment 1*, eighteen participants (9M, 9F, data collection is on-going) produced 80 utterances of the form "The word is [X]," where the target word could be a color, animal, or shape pictured on the screen. In *Experiment 2,* fourteen participants (7M, 7F, also on-going) were shown a 3x3 grid composed of different shapes in different colors and were asked to describe the location of an animal on the grid. The animal moved from trial to trial to elicit 60 contrastive sentences, e.g., "Now the mouse is on the *red* square." In the exposure and shadowing tasks, listeners heard versions of the model talker's speech acoustically manipulated so that L+H* contrastive peaks were 1.2 times their natural peak height (unedited mean=179Hz; edited mean=215Hz). The degree of manipulation was chosen to ensure that the peak was saliently enhanced but would still be lower than female participants' f0 peaks. In Expt. 1, participants' mean f0 was measured as the average f0 across each utterance. In Expt. 2, f0 was measured at the hand-labeled f0 peak in each target word; a relative peak height was calculated as the height of the f0 peak divided by the utterance average f0. Participants were recruited and paid through the online recruitment platform Prolific, and the online experiment was set up and administered using Gorilla [3].

Results showed that female participants in Expt. 1 lowered their utterance average f0 in shadowing (168Hz) and post-test (170Hz), relative to baseline (174Hz), converging toward the lower f0 of the model talker; male participants showed no change (baseline:101Hz, shadow:101Hz, post:101Hz). In Expt. 2, on the other hand, male participants (but not female participants) increased their peak f0 (realized in a contrastive pitch accent) in both shadowing (132Hz) and post-test (132Hz), relative to baseline (126Hz).

Separate mixed effects linear regressions on f0 for each experiment (utterance mean f0 for Expt. 1 and relative f0 peak for Expt. 2) included block and gender and their interaction as fixed factors, and random intercepts by-participant. The Expt. 1 model confirmed that both shadowing [t=-3.47, p<.001] and post-test [t=-2.92, p<.01] f0 were significantly lower than baseline. As expected, men were lower in pitch than women [t=-9.95, p<.0001]. There were significant gender by block interactions as well: female participants lowered their f0 more than male participants in both shadowing [t=2.470, p>.05] and post-test [t=2.053, p<.05].

The Expt. 2 model of relative peak f0 showed no main effect of block or gender, but there were significant interactions between block and gender in both tasks [shadowing: t=2.683, p<.001; post-test: t=3.134, p<.001], showing that only male participants increased the relative peak f0 in these tasks. A second model looking at raw peak f0 (not relative to utterance f0) showed that the shadowing peaks were lower than the peaks at baseline [t= -2.27, p<.05] (reflecting a lowering of f0 overall, as in Expt. 1), but post-test peaks were not [t=-0.46, p>0.1]. There was also a significant interaction between shadowing and gender [t=-2.526, p<.05], due to the lowering of shadowing peaks for women, while men's peaks were raised. This pattern seems to reflect the overall f0 lowering of females in shadowing also seen in Expt. 1.

Across the experiments, then, participants did imitate f0, but their behavior indicates two different patterns of imitation. In contrastively unmarked utterances (Expt. 1), female participants converged toward the low f0 of the model talker in both shadowing and post-test blocks, replicating previous studies on f0 imitation in which pitch (with no phonologically relevant distinctions) was imitated [4, 5]. Male participants in our study, whose f0 was already similar to the model talker, showed no lowering. When producing contrastively focused target items (Expt. 2), female participants again shifted downward, producing lower pitch accent peak f0 (but with no change in relative peak) during shadowing. In other words, they shifted toward the model talker's actual f0, but opposite his pattern of raised pitch accents. But in Expt. 2, this lowering was not carried over to the post-test block as in Expt. 1. Male participants did exhibit imitation (i.e., increased relative peaks), but we cannot determine from the current data whether they were imitating a specific acoustic value or a linguistic target (i.e., a high contrastive pitch accent), since both the actual f0 target and the pattern of raised pitch accent were higher than the male participants' baseline f0 peaks.

We argue nonetheless that speakers take into account linguistic factors when imitating. Absent linguistically meaningful structures, speakers may imitate acoustic targets directly, as seen in overall f0 lowering for female talkers in Expt. 1. When encountering and interpreting a linguistically meaningful structure (as in the pitch rise of a contrastive focus pitch accent in Expt. 2), the patterns of imitation are altered, especially if the acoustic and linguistic target are incongruent. Our results suggest that speakers' speech behavior reflects sensitivity to both a specific acoustic value and a linguistically-interpreted target. Additional data from female participants is being collected in order to further investigate the extent to which the presence (or salience) of incongruent linguistic and acoustic targets affects imitation.

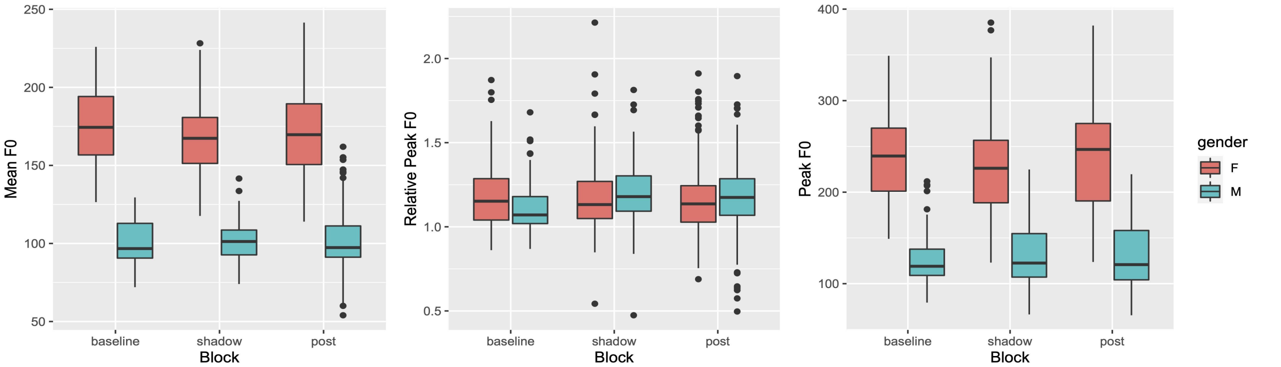


Fig 1. *Mean f0 (left) from Expt. 1 and relative peak f0 (middle) and peak f0 (right) from Expt. 2, by block. (Note that the model talker had a mean f0 of 101Hz, a mean relative peak of 1.99, and a mean peak f0 of 215Hz.)*

References

[1] Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013) Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *JML*, 69(3), 183-195.

[2] Nielsen, K. (2011) Specificity and abstractness of VOT imitation. *J Phon*, 39(2), 132-142.

[3] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020) Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, 52(1), 388-407.

[4] Dilley, L. C. (2010) Pitch Range Variation in English Tonal Contrasts: Continuous or Categorical? *Phonetica*, 67, 63–81.

[5] Babel, M., & Bulatov, D. (2012) The role of fundamental frequency in phonetic accommodation. *Lang & Sp*, 55, 231-248.

What are you sinking about? Effects of phonetic learning on online lexical processing of accented speech

Yevgeniy Melguy and Keith Johnson
University of California, Berkeley (USA)
ymelguy@berkeley.edu, keithjohnson@berkeley.edu

Speech produced with an unfamiliar accent may pose a challenge for listeners, resulting in delayed processing or lower intelligibility [1]. Such costs may be due to a mismatch between listener expectations about how a given sound category should be phonetically realized, and how it is implemented by non-native speakers. Phonetic mismatches can increase processing time [2], but listeners could avoid them by adjusting their expectations for a given speaker or speech variety. There is evidence that listeners use just such a strategy to perceptually adapt to an unfamiliar accent via *phonetic recalibration* of perceptual category boundaries. For instance, following exposure to an artificial accent involving a realization of /s/ that is phonetically intermediate between [s] and [f] (e.g., the word *moss* realized as *mo*[s/f]), listeners are more likely to categorize ambiguous tokens along a phonetic continuum between /s/ and /f/ as the trained phoneme /s/ [3].

Despite such adaptation being well-attested in the literature (see [4] for a review), the mechanisms involved in such category re-tuning are still underexplored. Namely, it is unclear whether listeners use a targeted mechanism specific to the phonetic patterns they encounter, or whether they use a more general mechanism of "criteria relaxation" that is insensitive to phonetic detail [5]. Recent literature has suggested that recalibration of category boundaries is achieved by a relatively general mechanism by which listeners expand phonetic categories in perceptual space, generalizing beyond the specific phonetic pattern they are exposed to [6]. In this study, the authors found that following exposure to an atypical accent where the dental fricative /θ/ was produced as [θ/s] (e.g., *throat* as [θ/s]*roat*), listeners shifted their category boundary toward /s/ on a /θ/-/s/ continuum. They also generalized learning to a novel contrast involving the same target phoneme /θ/-/ʃ/, classifying more ambiguous tokens as /θ/. However, no shift was observed for /θ/-/f/. The authors explain this finding based on the high degree of phonetic similarity between [s] and [ʃ], as measured by perceptual confusability data. This suggests that phonetic learning involves some sensitivity to phonetic detail, but that it is general enough to allow for transfer to a distinct pronunciation. Given that non-native speakers may be especially variable [7], maintaining this kind of relatively tolerant strategy may be beneficial for achieving speaker-independent accent adaptation.

However, as recent literature has pointed out [8], it is unclear whether adjustments to category boundaries in fact underlie improvements to comprehension and/or processing of accented speech. The current study tests the question of whether the same mechanism found in [6] leads to improvements in lexical processing following accent exposure. Across two experiments, 137 adult listeners recruited on the Prolific web platform completed a cross-modal priming lexical decision task, following exposure to an unfamiliar accent where a target phoneme was manipulated to be ambiguous (/θ/ = [θ/s]). This task involved presentation of an auditory prime followed by a written word, and listeners were asked to decide whether the latter was a real word or not. Critical trials involved the presentation of written /θ/ words (e.g., <therapy>) paired with either (1) ambiguous 'identity' primes either equivalent or highly similar to the exposure accent (e.g., [θ/s]*erapy* or [θ/ʃ]*erapy)* + <therapy>), (2) unambiguous but phonetically similar related primes (e.g., *serapy* or *sherapy* + <therapy>), or (3) unrelated primes (e.g., *banana* + <therapy>). In Exp. 1, results of linear mixed-effects modeling found a significant interaction of group and experimental condition ($\chi2(2) = 6.08$, $p < 0.05$), indicating that prior accent exposure affected word processing. Both controls and listeners with prior accent exposure saw similarly large 'identity' priming effects with /θ/ = [θ/s] primes (Fig.1), although there was a trend toward faster RTs for the accent exposure group ($b = -0.127$, $SE = 0.11$, $p = 0.25$). This suggests that these words were sufficiently similar to natural /θ/ that they did not pose significant processing problems. However, listeners in the accent exposure group showed a significantly larger difference between related trials (*serapy* + <therapy>) and identity trials (b = 0.26, SE = 0.11, p < 0.05) compared to controls. In Exp. 2, listeners with prior accent exposure saw significant related priming (e.g., *sherapy* + <therapy>) in the first half of trials, (b = -0.18, SE = 0.09, p < 0.05), whereas controls saw none. However, controls saw significant learning over the course of the task, with an increase in the size of the priming effect for both identity (b = -0.28, SE = 0.08, p < 0.001) and related prime trials (b = -0.23, SE = 0.08, p < 0.01), whereas listeners with prior experience saw no significant learning effect.

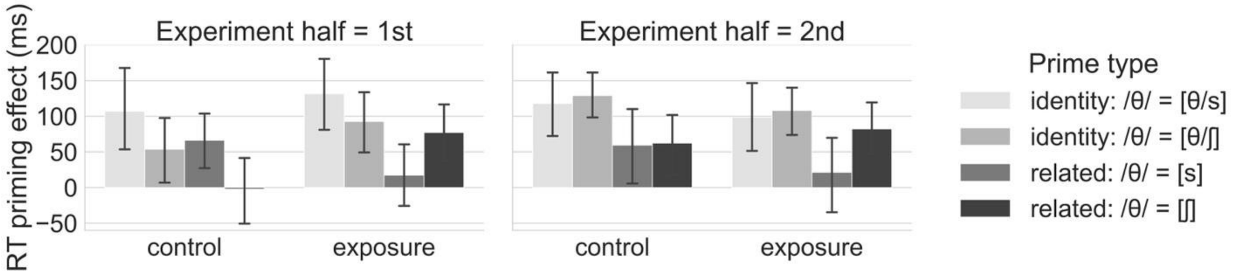


Fig. 1. Cross-modal priming results from Experiments 1 and 2 for listeners with and without prior exposure to a phonetically ambiguous /θ/ = [θ/s] pronunciation. Priming effect calculated by subtracting RTs from trials with 'identity' or related primes from trials with unrelated primes. Error bars indicate boot-strapped 95% confidence intervals.

Together this set of findings shows that phonetic detail plays a complex role in perceptual learning for speech. Although trained listeners showed a trend for stronger 'identity' priming with the ambiguous /θ/ primes vs. controls, accent experience did not yield a significant processing advantage. However, trained listeners did show changes to lexical processing elsewhere, as illustrated in weaker /s/-word priming (*serapy* + <therapy>) but stronger /ʃ/-word priming (*sherapy* + <therapy>) compared to untrained controls, suggesting that training resulted in listeners becoming more tolerant of atypical productions of the target phoneme in some cases (Exp.2) but less tolerant in others (Exp.1). This suggests that the learning mechanism is sensitive to phonetic detail and similarity to previously encountered speech, but that listeners can abstract over differences, facilitating lexical processing in certain novel contexts.

References

[1] Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *JASA*, *116*(6), 3647–3658.

[2] Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Percept. & Psychophys. 35*(1), 49–64.

[3] Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

[4] Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attent., Perception, & Psychophysics*, *71*(6), 1207–1218.

[5] Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. Developmental Science, 15(6), 732-738.

[6] Melguy, Y. V., & Johnson, K. (2022). Perceptual adaptation to a novel accent: Phonetic category expansion or category shift? *JASA*, *152*(4), 2090–2104.

[7] Wade, T., Jongman, A., & Sereno, J. (2007). Effects of Acoustic Variability in the Perceptual Learning of Non-Native-Accented Speech Sounds. *Phonetica*, *64*(2–3), 122–144.

[8] Zheng, Y., & Samuel, A. G. (2020). The relationship between phonemic category boundary changes and perceptual adjustments to natural accents. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(7), 1270–1292.

Lexical Sources of Phonological Alternation: a role for Voting Bases

Canaan Breiss[1] & Donca Steriade[1]

[1]*Massachusetts Institute of Technology (USA)*
canaan@mit.edu, steriade@mit.edu

Summary: We explore the lexicon's influence on the shape of new derivatives, using data from Romanian derived verbs. The key notions in what follows are: (i) *Derivatives* (Ds), forms created by affixation to a word or root; (ii) *local Bases* (B_L), exponents of an immediate syntactic constituent of D; and (iii) *remote Bases* (B_R), forms that are lexically related to D, but distinct from B_L. For example, the English *sỳntactíc-ian* has *syntáct-ic* as its B_L; and *sýntàx* as a B_R.

We start from the observation that in some morphological systems, the phonology of some Ds is determined not by the shape of their B_Ls, as expected ([1]), but by that of a B_R. The stress of trochee-initial *sỳntactícian*, for instance, is a version of the trochaic B_R *sýntàx*, and different from that of its iamb-initial B_L *syntáctic*. An extensive pattern of B_R-influenced Ds has been reported for Romanian in [2]. The present contribution is an effort to extensively verify one aspect of that study's findings, to explore a different analysis, and to clarify through a *wug*-test if the pattern we found in the Romanian lexicon mirrors native speaker preferences.

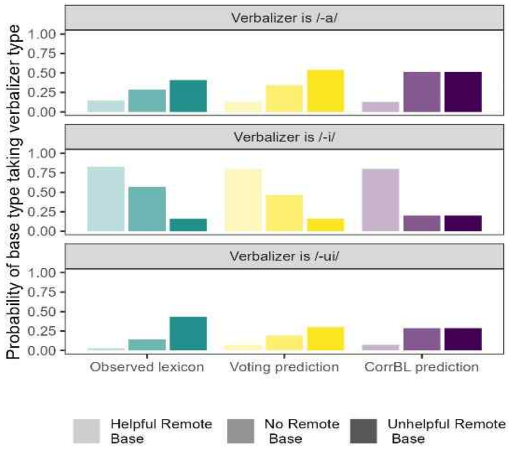
Romanian verbs and palatalization: Romanian derived verbs are formed by suffixes -a, as in [ɨm-pʌjenʒen-á] 'to cover in spiderwebs', B_L [pʌjánʒen] 'spider'; -í, e.g. [ɨm-pʌdur-í] 'to cover in forests,' from [pʌdúre] 'forest'; and -uí, e.g. [ɨn-vʌl-uí] 'to veil', B_L [vʌl] 'veil'. While free in general, the suffix choice is restricted after velars, a fact due to the process of velar palatalization, which turns velars [k, g] into palato-alveolars [tʃ, dʒ] before front vocoids: [e, i, j]. A first analysis of these restrictions [2] builds on two facts. First, like most front suffixes in Romanian, the verbalizer -í, triggers velar palatalization: e.g. [ɨn-furtʃ-í] 'to bifurcate (intrans.)' from B_L [fúrk-ʌ] 'fork'. Second, the verbalizer -í attaches freely to velar-final bases like [fúrk-ʌ], causing velar palatalization *iff these bases already possess a palatalized allomorph in their inflectional paradigm*. Thus, the plural of [fúrk-ʌ] is [fúrtʃ-i]. If the base lacks a palatalized plural – because it lacks *any* plural, or because its plural suffix does not trigger palatalization – the verbalizer -í is avoided, and one of the alternatives, -á or -uí, is used instead. Thus, [tsark] 'fenced space', plural [tsárk-uri], no palatalization, gives rise to [ɨn-tsʌrk-uí] 'to fence in'. Nouns like [tsark], with invariant velars in their inflectional paradigm, rarely give rise to /-í/-suffixed verbs: forms like *[ɨn-tsʌrtʃ-í], with palatalized root allomorph found just in the derived verb, not in its base N, are rare and marginal, and no verbs like *[ɨn-tsʌrk-í], with surface [k-i] across the suffixal boundary, exist at all. Thus both markedness, i.e. avoidance of [ki], and faithfulness play a role in selecting the verbal suffix and in velar palatalization. Since verbs like [ɨn-furtʃ-í] 'bifurcate' don't refer to a plurality of participants, the B_L of the derived verb is not its plural. Thus plural [furtʃ-i] 'forks,' is a B_R since its presence in forming the verb is not syntactically justified. From these assumptions, it follows that the selection of the verbalizer /-í/ and the application of palatalization in forms derived with this /-í/ are licensed by a B_R, the palatalized plural stem of the base N.

Analyses: We explored two analyses. One ([2, 3]) uses four ingredients: (i) a markedness constraint *KE, banning velars before front vocoids; (ii) a faith constraint ID B(ASE)- D(ERIVATIVE) requiring D's stem to correspond in its consonantism to that of *some* Base, B_L or B_R; (iii) a preference for D's stem to be in correspondence to its B_L (Corr B_L); and (iv) a preference to use the -í suffix in change-of-state intransitives (USE /-í/). The ranking ID-BD >> *KE >> USE /-í/ >> CORRB_L models the ban on palatalized items like *[ɨn-tsʌrtʃ-í] (given invariant [tsark]); the avoidance of /-í/ after bases with such invariant velars; and the default preference for the verbalizer /-í/ in derived intransitives, as in [ɨn-furtʃ-í]. A similar model invoking violable CorrB_L proves useful elsewhere ([3, 4]). Tableau A (bottom p. 2) presents a schematic analysis with select candidates in a standard categorical OT.

An alternative model, used in [6], starts from the idea that each form in a lexical paradigm exerts some attraction on the stem of the D, independently of improvements in D's markedness. There are multiple active faithfulness constraints (ID$_L$, ID$_R$), each expressing a preference for correspondence between D’s stem and a specific B, either the B_L or a B_R. Markedness (*KE), ganging up with these constraints, can result in the selection of B_R-based stems if a B_R improves the stem’s markedness. Importantly, however, the two faithfulness constraints can gang up: when the B_L and B_R do not improve markedness, they should jointly discourage palatalization even more than in nouns lacking any B_R at all: therefore, they should select non/-í/ even more often.

Although the two theories make identical predictions about forms with palatalized B_Rs (e.g. pl. [fúrtʃ-i], called *helpful* B_Rs here) and about forms lacking any B_R (e.g. [vlʌg-uí] ‘exhaust’ based on singular-only [vlágʌ] ‘force’), they make different predictions about the effect of B_Rs that resemble the B_L, e.g. [tsark], pl. [tsark-uri]. We call plurals like [tsark-uri] *unhelpful* B_Rs: if *two* bases ‘voting’ to preserve [k], thus *against* palatalization, have more effect than just one, such nouns could inhibit palatalization in the derived verb more than plural-less [vlágʌ].

Evidence for Voting Bases in the Romanian lexicon: Drawing on dictionary data (dexonline.ro), we find that the distribution of verbalizer allomorphs is sensitive not only to the presence of helpful B_Rs (with -í more likely in cases where the paradigm already has a palatal-final stem allomorph), but also that B_Ls with an unhelpful B_R (that is, a non-palatalized plural) are even less likely to take the non-í verbalizers, -á or -uí, which do not trigger palatalization, than nouns lacking any B_R. These data are plotted in the figure below, with predictions from the two theories derived from a Maximum Entropy grammar ([6]) implementing each analysis fully and faithfully (not shown). A χ-squared test with one degree of freedom finds that the CORRB$_L$ model fits (right) fits the lexical data (left) significantly less well than the Voting Bases model (center); $p < 0.01$. An experiment with Romanian native speakers is in progress to test the generality of this pattern in words that lack a lexicalized corresponding verb. If speakers’ allomorph selection is sensitive to the presence of both helpful and unhelpful Remote Bases, we will take this to support the lexical evidence we have advanced here to favor the Voting Bases model.



References

[1] Chomsky, N., & Halle, M. (1968). The sound pattern of English.

[2] Steriade, D. (2008). A pseudo-cyclic effect in Romanian morphophonology. *Inflectional identity*, *18*, 313-360.

[3] Steriade, D., & Stanton, J. (2020). Productive pseudo-cyclicity and its significance. *Talk at LabPhon*, *17*.

[4] Steriade, D., & Yanovich, I. (2015). Accentual allomorphs in East Slavic: inflection dependence. *Understanding Allomorphy*, 254-314.

[5] Breiss, C. (2021). *Lexical Conservatism in phonology: theory, experiments, and computational modeling*. UCLA

[6] Goldwater, S., Johnson, M., Spenader, J., Eriksson, A., & Dahl, Ö. (2003, April). Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory* (Vol. 111, p. 120).

Tableau A: CORRB$_L$ analysis (select candidates); H = Harmony, candidate subscripts indicate Base-Derivative correspondence.

B_L [furk]$_L$-ʌ B_R [furtʃ]$_R$-i	ID-BD	*KE	USE /-í/	CORRB$_L$		B_L [tsark]$_L$ B_R [tsark]$_R$-uri	ID-BD	*KE	USE /-í/	CORRB$_L$
D_1: [furtʃ]$_L$ -í	*					D_1: [tsʌrtʃ]$_L$-í	*			
D_2: [furk]$_L$ -í		*				D_2: [tsʌrk]$_L$-í		*		
D_3: [furtʃ]$_R$ -í				*		D_3: [tsʌrtʃ]$_R$-í	*			*
D_4: [furtk]$_R$ -í	*	*		*		D_4: [tsʌrk]$_R$-í		*		*
D_5: [furk]$_L$ -uí			*			D_5: [tsʌrk]$_L$-ui			*	

Tableau B: Voting Bases analysis (select candidates) for illustration of violations and ganging only; H = Harmony

B_L [furk]$_L$-ʌ B_R [furtʃ]$_R$-i	ID-B$_L$	ID-B$_R$	*KE	USE /-í/	H		B_L [tsark]$_L$ B_R [tsark]$_R$-uri	ID-B$_L$	ID-B$_R$	*KE	USE /-í/	H
	1	1	1	1				1	1	1	1	
D_1: [furtʃ] -í	*				1		D_1: [tsʌrtʃ]-í	*	*			2
D_2: [furk] -í		*	*		2		D_2: [tsʌrk]-í			*		1
D_3: [furk] -uí		*		*	2		D_3: [tsʌrk]-ui				*	1

Perceptual bases for the compensatory lengthening typology

Suyeon Yun

Chungnam National University (Korea)
suyeon.yun@cnu.ac.kr

This study provides perceptual bases for the typological patterns of compensatory lengthening (CL) from experimental results. CL refers to the lengthening of a vowel triggered by the deletion of a neighboring consonant, e.g., CVC → CVː. Moraic theory [1] assumes that (i) the deletion of postvocalic codas is likely to trigger CL, whereas the deletion of prevocalic onsets never does, and (ii) the deletion of moraic codas adjacent to the preceding vowel is likely to trigger CL, whereas the deletion of non-moraic codas not adjacent to the preceding vowel never does.

More recently, Yun [2, 3] presented the results of a cross-linguistic survey of 141 languages showing CL triggered by consonant loss either synchronically or diachronically, arguing that the typological patterns of CL may be characterized as implicational universals in terms of the position and adjacency of the trigger consonant relative to the target vowel. Specifically, it is shown that (i) if the loss of a prevocalic consonant triggers CL, so does the loss of a postvocalic consonant, and (ii) if the loss of a consonant not adjacent to the vowel triggers CL, so does the loss of a consonant adjacent to a vowel. In addition, when an intervocalic consonant is deleted, it is in most cases the preceding vowel that is lengthened, i.e., VCV → VːV.

This typology further indicates that the domains where CL applies are string-based, not syllable-based. The domains of duration preservation in CL triggered by the loss of a postvocalic consonant and by the loss of a prevocalic consonant, respectively, are [VC] and [CV], and the domains of CL triggered by the loss of a consonant adjacent to a vowel and that triggered by the loss of a consonant not adjacent to a vowel are [VC] and [VCC], respectively.

This study hypothesizes that the typological asymmetries in CL may result from differences in the perceptual salience of duration change in these sequential domains. This is based on the P-map hypothesis [4], which posits that perceptually smaller modifications are preferred, and it is encoded in grammar that the rankings in which faithfulness constraints forbidding more drastic change always dominate those forbidding less drastic change. The current perceptual hypothesis for CL typology is that the loss of a consonant duration results in a perceptually more drastic change to the [VC] domain than to the [CV] and [VCC] domains; therefore, CL is more likely to occur when a postvocalic and vowel-adjacent consonant in [V$\underline{\mathbf{C}}$] is deleted than when a prevocalic consonant in [$\underline{\mathbf{C}}$V] or non-vowel-adjacent consonant in [VC$\underline{\mathbf{C}}$] is deleted.

To test this hypothesis, ABX discrimination and ABX similarity judgment experiments were conducted with 13 English and 53 Korean speaker participants. The AXB discrimination task used nonce word stimuli that differed in the duration of a consonant, [m] or [s], located in the (i) postvocalic position (V_(C)), (ii) prevocalic position ((C)_V), (iii) postconsonantal and non-vowel-adjacent position (VC_), and (iv) preconsonantal and non-vowel-adjacent position (_CV). The baseline duration of the target consonant was 150 ms, modified in four steps: 50 ms, 100 ms, 200 ms, and 250 ms. For example, the participants heard an ABX triplet that consisted of a$\underline{\mathbf{m}_{\mathbf{[100\ ms]}}}$ vs. a$\underline{\mathbf{m}_{\mathbf{[150\ ms]}}}$ vs. a$\underline{\mathbf{m}_{\mathbf{[150\ ms]}}}$ and chose whether the third sounded the same as the first or second. The results show that listeners were more sensitive to the duration modification of a postvocalic consonant (*d′* = 1.94 for English, *d′* = 2.18 for Korean) than to that of a prevocalic consonant (*d′* = 1.6 for English, *d′* = 2.09 for Korean), postconsonantal consonant (*d′* = 1.38 for English, *d′* = 1.8 for Korean), or preconsonantal consonant (*d′* = 1.59 for English, *d′* = 1.75 for Korean).

The ABX similarity judgment task was designed to directly test whether the duration of the [VC] domain was more likely to be preserved than the duration of the [CV] domain in VCV sequences. For example, participants were asked whether $\mathbf{a}_{[150ms]}\mathbf{m}_{[150ms]}\mathbf{a}_{[150ms]}$ sounded more similar to $\mathbf{a}_{\mathbf{[200ms]}}\mathbf{m}_{[100ms]}\mathbf{a}_{[150ms]}$ (lengthening the preceding vowel) or $\mathbf{a}_{[150ms]}\mathbf{m}_{[100ms]}\mathbf{a}_{\mathbf{[200ms]}}$ (lengthening the following vowel) when the intervocalic consonant was reduced. Listeners were more likely to preserve the lost duration of the intervocalic consonant in the preceding vowel (53.8% for English and 56.6% for Korean) than in the following vowel (47.2% for English and 43.4% for Korean).

These results suggest that when a postvocalic and vowel-adjacent consonant in [V**C**] is reduced, listeners can identify the loss of its duration more accurately than when a prevocalic consonant in [**C**V] or a postconsonantal consonant in [VC**C**] is reduced. In addition, when an intervocalic consonant, which is simultaneously postvocalic and prevocalic, is shortened in a [VCV] sequence, listeners prefer to preserve the lost duration in the preceding vowel in [**V**C] than in the following vowel in [C**V**]. Therefore, the impetus to compensate for duration loss would be stronger for the consonant in [V**C**] than for the consonant in [**C**V] and in [VC**C**] by the P-map hypothesis, leading to frequent occurrences of CL through postvocalic and vowel-adjacent consonant loss.

References

[1] Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry, 20*(2): 253-306.

[2] Yun, S. (2010). The typology of compensatory lengthening: A phonetically-based Optimality Theoretic approach. MA thesis, Seoul National University.

[3] Yun, S. (2023). Implicational universals of compensatory lengthening. Manuscript submitted for publication.

[4] Steriade, D. (2008). The phonology of perceptibility effects: the P-Map and its consequences for constraint organization. In K. Hanson & S. Inkelas, (eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky*. (pp. 151-179). Cambridge: MIT Press.

Resyllabification revisited: arguments for V-to-V intervals as units of rhythm

Donca Steriade

MIT

Introduction: When words are combined in connected speech, changes occur in the timing of articulatory gestures and in the feature composition of segments near junctures [1, 2], as when consonant-final words precede vowel-initial ones. Thus *an aim* becomes similar to *a name*. Such changes are standardly attributed to resyllabification [3, 4], the reassignment of consonants from the coda of one syllable to the onset of the next, $V_1C.\#V_2 \rightarrow V_1.C\#V_2$. The phonetic study of resyllabification in several languages (e.g. [5], [6], [7], [8]) shows that differences remain between 'resyllabified' $V_1.C\#V_2$ and basic $V_1.\#CV_2$. This fact is consistent with an alternative hypothesis, entertained in [5] and elsewhere: segmental changes of $V_1C\#V_2$ in connected speech, relative to $V_1C\#$ in isolation, are unrelated to the *prosodic structure* of $V_1C\#$ finals.

The present study takes up this question by reconsidering cross-juncture sequences from the different angle of metrical weight. Resyllabification ($V_1C.\#V_2 \rightarrow V_1.C\#V_2$) should turn $V_1C\#$, considered a heavy syllable, into light V_1. This change is detectable in quantitative meters, based on periodic alternations of heavy and light units. With this in mind, I consider the quantitative meters of A.Greek and Latin, having re-examined metrically parsed samples of the Iliad (first 3 songs) and the Aeneid (first 2 songs). Two results emerged that bear on resyllabification. First, the weight of $V_1C\#V_2$ is identical to that of $V_1\#CV_2$, as if resyllabification has indeed applied. But, second, the metrical texts reveal striking gaps in the distribution of V(C)# sequences in lines of verse, gaps that turn out to be governed by general laws: [9], [10], [11], [12]. A unified interpretation of all this data is possible, if the units of rhythm counted in stress and meter are not syllables, but *Vowel-to-Vowel (V-to-V) intervals* ([13], [14], [15], [16], [17] [18]).

Intervals: A V-to-V interval (abbrev. *interval*) is a unit of rhythm containing, like a syllable, one nucleus. Unlike a syllable, it begins at the left edge of its nucleus (or at the P-Center, [14]) and ends right before the next nucleus, or at the end of the domain. All segments in the interval contribute to its weight. Interval boundaries are invariably defined by nucleus boundaries, so a VCV string is always parsed |VC|V|, where '|' = interval boundary. An interval changes its contents if one or more Cs are added to its right, as in V#CV parsed as |V#C|V|. This reparse in connected speech differs from resyllabification: it's automatic, not the outcome of a specific constraint ranking, and it yields different weight results from resyllabification, as seen below.

Intervals and Weight: Weight distinctions are defined for intervals in (1). The illustrative forms in (1) are from Latin, where heavy penults attract stress. Words are divided into syllables in (a), into intervals in (b). Consistent with the definition of intervals, initial Cs don't belong to any interval at all. Note that VC intervals must be light, unlike C_0VC syllables. This is supported by the common occurrence of stress systems with light final VC. For *extra-light* intervals, see [17].

<table>
<tr><th>(1)</th><th>Extra light</th><th>Light</th><th>Heavy</th></tr>
<tr><td>(a) syllables</td><td></td><td>$C_0\breve{V}$ (<b><i>ci</i></b> in <i>fá.<b>ci</b>.lis</i>)</td><td>$C_0\breve{V}CC_0$, $C_0\bar{V}C_0$ (<b><i>cul</i></b> in <i>fa.<b>cúl</b>.tās</i>)</td></tr>
<tr><td>(b) intervals</td><td>$\breve{V}$ (<b><i>i</i></b> in <i>m|úl|<b>i</b>|er|</i>)</td><td>$\breve{V}C$ (<b><i>il</i></b> in <i>f|ác|<b>il</b>|is|</i>)</td><td>$\breve{V}CCC_0$, $\bar{V}C_0$ (<b><i>ult</i></b> in <i>f|ac|<b>últ</b>|ās|</i>)</td></tr>
</table>

Weight-changes in phrasal contexts: The present study and earlier work in [9–12] documents restrictions on word-final $\breve{V}(C)\#$ sequences in A.Greek and Latin quantitative meters. These restrictions are outlined in (2). The top row in (2) identifies the weight of the first position in a $\breve{V}CC_0V$ string containing a word juncture; the middle row identifies several types of $\breve{V}CC_0V$ sequences containing a word boundary; the bottom row indicates which among these are under-attested or banned in verse. Shaded cells highlight restricted, i.e. underattested or banned sequences. Clear cells correspond to favored or unrestricted sequences: e.g. $\breve{V}C\#V$ is favored over $\breve{V}\#CV$, [11], and $\breve{V}CC\#V$ is unrestricted, compared to restricted $\breve{V}C\#CV$, [10]. Unlike $\breve{V}C\#$ and $\breve{V}\#$ finals, finals like $\bar{V}C_0$ # and $\breve{V}CC(C)\#$ are never restricted, in any context [10].

(2)	First position is metrically light		First position is metrically heavy		
	(a) V̌C#V	(b) V̌#CV	(c) V̌#CCV, CC=sT	(d) V̌C#CV	(e) V̌CC#V
		restricted	prohibited/restricted	restricted	

Interval-based vs. syllable-based analyses: A unified, interval-based analysis of these restrictions is outlined below. The sequences in (2) are parsed into intervals in (3), and into syllables in (4). Shaded cells correspond, as in (2), to underattested or banned sequences:

(3)	(a) |V̌C|#V	(b) |V̌#C|V	(c) |V̌#CC|V	(d) |V̌C#C|V	(e) |V̌CC|#V
(4)	(a) V̌.C#V	(b) V̌.#CV	(c) V̌#C.CV	(d) V̌C.#CV	(e) V̌C.C#V

The interval parses in (3) reveal the reason for the restrictions observed: in all the restricted cells, and only there, the *weight of the word-final interval differs from its weight in isolation*. That's because the line medial context adds one or more Cs to it. By contrast, the word-final intervals of the unrestricted cases (a) and (e) maintain line-internally the same weight they have in isolation.

On an interval analysis, all patterns of underattestation outlined in (2-3) emerge as driven by a preference for *weight invariance between intervals in the word in isolation and their line-medial correspondents*. All changes of weight between isolation and the line-medial contexts are disfavored, and cause poets to avoid the relevant sequence. By contrast, syllable-based parses of the same sequences, in (4), are unable distinguish restricted sequences from unrestricted ones.

References

[1] Stetson, R. H. (1951) *Motor phonetics*. 2nd ed. Amsterdam: North-Holland

[2] Tuller, B. and Kelso, J.A.S (1991) The production and perception of syllable structure*. Journal of Speech and Hearing Research*, 34, 501–508.

[3] Kahn, Daniel (1980) *Syllable-based generalizations in English phonology.* New York: Garland

[4] Kiparsky, Paul (1979) Metrical structure assignment is cyclic. *Linguistic Inquiry*, 10.3: 421-441

[5] de Jong, Ken (2001) Rate-Induced Resyllabification Revisited. *Language and Speech,* 434 (23), 197–216

[6] Fougeron, C. (2007) Word boundaries and contrast neutralization in French enchaînement. Cole, J. and Hualde J.I. *Papers in Laboratory Phonology IX: Change in Phonology*, Berlin: Mouton de Gruyter, pp. 609-642.

[7] Scobbie, J. and Pouplier, M. (2010) The role of the syllable in external sandhi, *Journal of Phonetics*, 38, 240-259

[8] Son, M. (2011) Projection of Syllable Structure: Korean /VkV/. *Korean Journal of Linguistics,* 36, 395-414.

[9] Hartel, W. (1873) *Homerische Studien,* Berlin, Vahlen.

[10] Hilberg, I. (1879) *Princip der Silbenwaegung* Wien, Hölder

[11] Soubiran, Jean (1966) Ponctuation bucolique et liaison syllabique en grec et en latin*, Pallas*, No. 13, pp. 21-52

[12] Hoenigswald, H. (1949) A Note on Latin Prosody: Initial S Impure after Short Vowel, *Transactions and Proceedings of the American Philological Association* (TAPA), Vol. 80, 271-280.

[13] Farnetani, E. and Kori, S. (1986) Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication 5,17-34*

[14] Barbosa, P. and Bailly, G. (1994) Characterisation of rhythmic patterns for text-to-speech synthesis. *Speech Communication*, 15 (1-2), 127-137.

[15] Hirsch, A. (2014) What is the domain for weight computation? *Proceedings of AMP*. 2013

[16] García-Duarte, G. (2017) Weight gradience and stress in Portuguese. *Phonology 34.* 41–79

[17] Steriade, D. (2019) CiV-Lengthening and the weight of CV, in T. Bradley et al. (eds) *Schuh-Schrift*, UCLA.

[18] Sturtevant, E. (1922) Syllabification and syllabic quantity in Greek and Latin, *TAPA*, Vol. 53, 35-51

Poster Presentations

Day 2

(Saturday, May 27, 2023)

Place Assimilation of the Moraic Nasal to /r/ in Japanese

Maho Morimoto[1, 2], Ai Mizoguchi[3, 4] & Takayuki Arai[1]

[1]Sophia University (Japan), [2]JSPS (Japan), [3]Maebashi Institute of Technology (Japan), [4]NINJAL (Japan)
maho.morimoto.jp@gmail.com, aimizoguchi@maebashi-it.ac.jp, arai@sophia.ac.jp

Background: The Japanese moraic nasal /N/ is known to regressively assimilate to the following plosives in terms of place of articulation (e.g., [1, 2]), as is demonstrated by several articulatory studies (e.g., [3, 4, 5]). /N/ is generally thought to assimilate not only to plosives but also to the liquid consonant /r/ (e.g., [1, 6]). However, the paucity of experimental data observing the actual articulatory behavior limits our understanding of the gradient nature of the process.

We address two aspects of liquids for which the assimilation may not work as in the other coronal consonants. First, the constriction location of Japanese /r/ is known to vary considerably [7], including those that are more posterior than in other coronals (e.g., [7, 8]). Second, in the framework of Articulatory Phonology (AP; [9]), the possibility that liquids are cross-linguistically gesturally complex segments involving the coordination of coronal and dorsal gestures has been explored [10]. Different views exist as to whether this applies to Japanese /r/ as well [8, 11]. In this study, we report the results from an ultrasound study to examine the place assimilation of the Japanese moraic nasal /N/ to /r/ and the extent to which it differs from the assimilation to /d/ and /n/ as well as the utterance-final /N#/.

Methods: The current study reports data from two native speakers of Tokyo area Japanese (one female, AJF01, and one male, AJM01, both in their 20s). We report the articulatory results for the items shown in Table 1. Each word was presented 10 times in a random order and the participants read them aloud in isolation. Audio and ultrasound recordings were obtained simultaneously. For the analysis, the midsagittal tongue contours at the acoustic midpoint of each target consonant were manually traced using GetContours [12].

Results & Discussion: Fig. 1 shows the tongue contours for the target segments in each speaker, predicted from the repetitions with 95% confidence intervals using the generalized additive model (GAM; [13]). Overall, the results confirm the traditional description in which coda nasals assimilate in place to plosives and liquids. An additional finding is that the constrictions for /d/, /n/, and /r/ are more apical, as inferred from the concave shape at the tongue blade, while the constrictions for /N/ in /Nd/, /Nn/, and /Nr/ are more laminal, as indicated by the more convex shape. This may be a result of a gestural planning specific to assimilated nasals through which speakers achieve a prolonged and more stable tongue tip (TT) constriction. Alternatively, this may be explained as gestural blending in the framework of AP. Assuming that there is an underlying tongue body (TB) gesture for /N/, as suggested in [4], the need to raise the TT and TB simultaneously may result in the slight raising of the tongue blade. This is schematized in Figs. 2a and 2b, following the box notation in [9].

The tongue contours for /r/ and /Nr/ suggest that the constriction location of the TT is higher and more posterior than that of the obstruents, especially for AJF01. Furthermore, we observe a slight bulge around the TB area for /r/ and /Nr/ for both speakers. While this may simply be viewed as a by-product of the retracted TT constriction, possibly involving a concentration of the mass due to the contraction of the tongue surface muscles, it is also consistent with the view that the Japanese /r/ is gesturally complex. As shown in Fig. 2c, the bulge in the TB area may be accounted for as the cumulative effect of the intrinsic dorsal gestures of /r/ and /N/.

Conclusion: The current study reported the differences among the tongue contours of /N/ followed by /r/ and the other consonants, /d/ and /n/. Further investigation is needed to test the analyses discussed above and obtain a full picture of the assimilation process. Our results also highlight the need to combine point-tracking techniques such as EMA and whole-tongue imaging techniques such as ultrasound or rtMRI to better understand the nature of the lingual gesture involved in the production of /r/ (i.e., retraction as a horizontal displacement of the tongue body vs. sliding back of the tongue surface).

Table 1 Experimental stimuli (target segments are indicated in red).

/NC/				/C/			
/Nd/	/aNda/	あんだ	‘a hit’	/d/	/hadaka/	はだか	‘naked’
/Nn/	/aNna/	あんな	‘like that’	/n/	/anata/	あなた	‘you’
/Nr/	/haNra/	はんら	‘half-naked’	/r/	/ara/	あら	‘coarseness’
/N#/	/kaNaN/	かんあん	‘consideration’	–	–	–	–

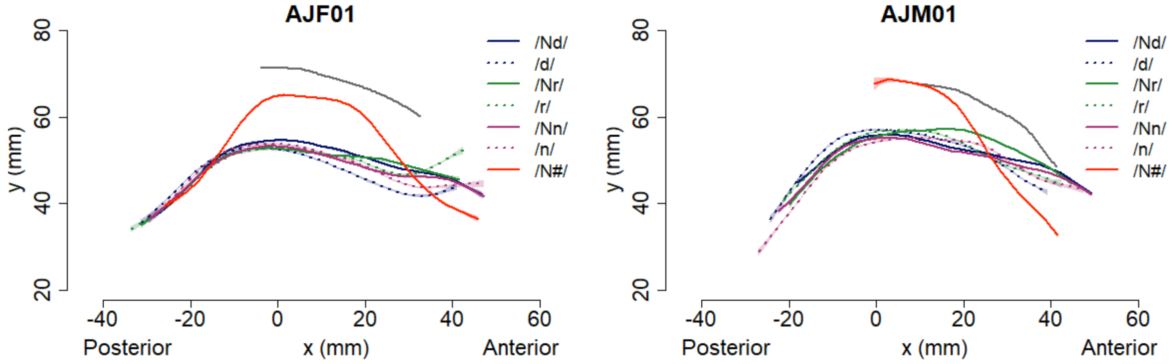


Figure 1 GAM contours with 95% confidence intervals for the seven target segments in each speaker. The gray line above the tongue contours shows the palate contour.

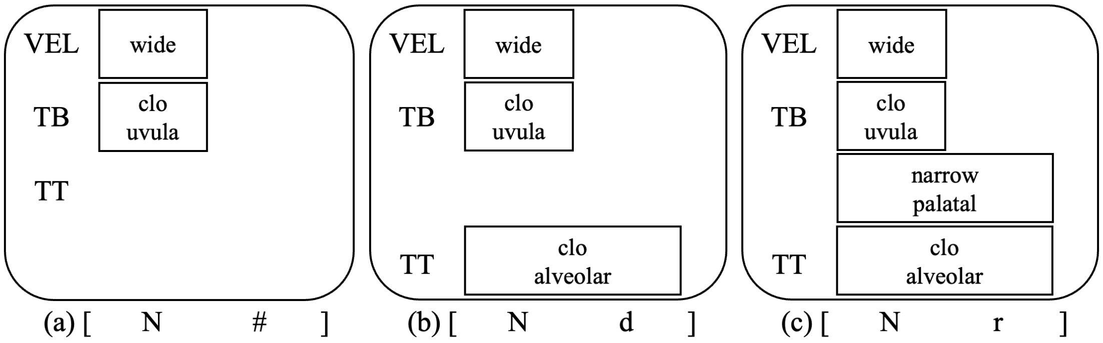


Figure 2 Partial gestural scores showing the activation intervals of consonantal gestures.

References

[1] Vance, T. J. (2008). *The Sounds of Japanese*. Cambridge University Press.

[2] Labrune, L. (2012). *The Phonology of Japanese.* Oxford University Press.

[3] Stephenson, L., & Harrington, J. (2002). Assimilation of place of articulation: Evidence from English and Japanese. *Proceedings of the 9th Australian International Conference on Speech Science and Technology*, 592–597.

[4] Mizoguchi, A. (2019). *Articulation of the Japanese Moraic Nasal: Place of Articulation, Assimilation, and L2 Transfer.* Ph.D. dissertation.

[5] Maekawa, K. (2021). Production of the utterance-final moraic nasal in Japanese: A real-time MRI study. *Journal of the International Phonetic Association*, 1–24.

[6] Akamatsu, T. (1997). *Japanese Phonetics: Theory and Practice*. (Vol. 3). Lincom Europa.

[7] Kawahara, S., & Matsui, F. M. (2017). Some aspects of Japanese consonant articulation: A preliminary EPG study. *ICU Working Papers in Linguistics*, *2*, 9–20.

[8] Morimoto, M. (2020). *Geminated Liquids in Japanese: A Production Study*. Ph.D. dissertation.

[9] Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*(2), 201–251.

[10] Proctor, M. (2011). Towards a gestural characterization of liquids: Evidence from Spanish and Russian. *Papers in Laboratory Phonology*, *2*(2), 451–485.

[11] Yamane, N., Howson, P., & Wei, P.-C. G. (2015). An ultrasound examination of taps in Japanese. *Proceedings of ICPhS XVIII*, 1–4.

[12] Tiede, M. K. (2022). *GetContours*. GitHub repository, https://github.com/mktiede/GetContours.

[13] Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.

Prominence in Mundari Disyllables and Inflected Polysyllabic Nouns

Luke Horo, Pamir Gogoi & Gregory D.S. Anderson

Living Tongues Institute for Endangered Languages (USA)
luke.horo@livingtongues.org, pamirgogoi11@gmail.com, gdsa@livingtongues.org

In this paper, we describe our preliminary findings from an ongoing study of intonation in Mundari, an Austroasiatic language spoken by some two million people in at least four dialects. Here we present a comparative analysis of the system of prominence attested in two such dialects, viz. Hasadaʔ and Naguri. We use as a basis for this preliminary study disyllabic forms of any function and polysyllabic nouns that are inflected for a variety of case, possession, etc. categories. Future studies will cover the significantly more complicated system of intransitive and transitive verb forms.

When examining previous studies on Mundari, one encounters a wide array of perspectives– there are almost as many different analyses as there have been analysts. Thus, according to [1, 2] and [3], Mundari is a stress language, while [4] considers it to be a pitch accent language. Moreover, [5] claims that in disyllabic words the accent is on the first syllable, with (lexical) exceptions. [3] also claims that Mundari stresses the second syllable in disyllabic words if it is of the shape of $C^1V^1C^2V^2$ or $C^1V^1C^2V^2C^3$ but in words of the shape $C^1V^1C^2C^3V^2$, stress falls on the initial syllable, suggesting a QS iambic system. If the word is trisyllabic, stress falls on the 2nd syllable regardless of the shape according [3]. Further, [1] finds that only if the final syllable is closed, it is accented, otherwise it is the initial syllable in disyllabic words, thus QS trochaic. Most recently, [6] states that if a word is trisyllabic, stress can only be on the second or the third syllable: on the third syllable if that is not a suffix, otherwise it falls on the second syllable in Mundari trisyllabic words, but never on the first syllable, regardless of syllable weight. Also, [4] states that "in Mundari a phonological word maximally consists of three syllables". However, these previous studies are impressionistic and not verified by instrumental analysis, nor supported with statistical data.

Overall, it has been assumed that all Munda languages show a trochaic pattern of prominence [7, 8, 9], but recent instrumental analyses of Sora [10, 11, 12] and Assam Santali [13], supported by statistical data, suggest that these two sister languages to Mundari rather consistently show second syllable prominence. The prominence is cued by intensity, duration and/or fundamental frequency on the second syllable.

In this report we offer a new instrumental analyses of Mundari focusing for this study on disyllables and inflected polysyllabic nouns. We compare these findings with the claims made in the literature about the language, as well as with the findings from the more recent studies on related languages. This includes the role of quantity sensitivity (if any) in determining patterns of prominence, what the acoustic cues of prominence in Mundari are and how they conspire to encode the prominent syllable, and whether the maximal phonological word is three syllables or not. We also compare these results with an exercise in writing words by native speakers that speaks to the fact that psychological "reality" of word shapes and boundaries may not coincide with phono-prosodic data on the nature of the word in Mundari. All data are taken from field notes.

References

[1] Cook, Walter A. (1965). *A descriptive analysis of Mundari: A study of the structure of the Mundari language according to the methods of linguistic science.* Washington, DC: Georgetown University dissertation.

[2] Langendoen, D. Terence (1963). *Mundari Phonology*. Unpublished paper.

[3] Sinha, N.K. (1975). *Mundari Grammar*. Mysore: Central Institute of Indian Languages.

[4] Osada, Toshiki (1992). *A reference grammar of Mundari*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

[5] Hoffmann, Johann. (2001). [1903] *Mundari Grammar*. Calcutta: Bengal Secretariat Press.

[6] Osada, Toshiki. (2008). Mundari. In Gregory D.S. Anderson (ed.), *The Munda languages*. Routledge Language Family Series. London: Routledge (Taylor and Francis), pp. 99-164.

[7] Donegan, Patricia J. & David Stampe. (1983). Rhythm and the holistic organization of language structure. In Richardson, John F., Mitchell Marks, & Amy Chukerman (eds.), *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax*, 337–353. Chicago: Chicago Linguistic Society.

[8] Donegan, Patricia J. & David Stampe. (2004). Rhythm and the synthetic drift of Munda. In Singh, Rajendra (ed.), *Yearbook of South Asian languages and linguistics*, 3–36. Berlin: Mouton De Gruyter.

[9] Donegan, Patricia J. (1993). Rhythm and vocalic drift in Munda and Mon-Khmer. *Linguistics of the Tibeto-Burman Area* 16(1). 1–43.

[10] Horo, Luke and Priyankoo Sarmah. (2015). Acoustic analysis of vowels in Assam Sora. In L. Konnerth et al. (eds.) *Northeast Indian Linguistics* 7. 69-86. Canberra: ANU.

[11] Horo, Luke. (2017). *A phonetic description of Assam Sora*. Guwahati, India: Indian Institute of Technology dissertation.

[12] Horo, Luke, Priyankoo Sarmah and Gregory D. S. Anderson (2020). Acoustic phonetic study of the Sora vowel system. *Journal of the Acoustical Society of America* 147(4). 3000-3011. https://doi.org/10.1121/10.0001011.

[13] Horo, Luke and G. D. S. Anderson. (2021). Towards a prosodic typology of the Kherwarian Munda languages: Santali of Assam. In M. Alves and P. Sidwell (eds.) *Proceedings of the 30th Annual SEALS meeting*. Honolulu; UH Press, pp. 298-317.

Korean speakers' perception of cues to liquid contrasts: an EEG study

JIWON HWANG[1], HYUNAH BAEK[2] & ELLEN BROSELOW[3]

[1]Stony Brook University (USA), [2]Ajou University (Korea), [3]Stony Brook University (USA)
jiwon.hwang@stonybrook.edu, hyunahbaek@ajou.ac.kr, ellen.broselow@stonybrook.edu

In Korean, the lateral liquid and tap [ɾ] are in complementary distribution, with a single liquid pronounced as [l] in syllable coda and as tap [ɾ] in syllable onset (e.g., [mul]/[muɾ-i] 'water/water+NOM'). However, for words borrowed into Korean, English intervocalic [r] is adapted as [ɾ] ([tɕʰeɾi]) 'cherry'), as expected, but intervocalic single [l] is generally adapted as geminate [ll] ([kʰolla] 'cola'), even though Korean [ll] is considerably longer than English [l] [1]. This suggests that Korean listeners are more sensitive to the lateral/non-lateral contrast than to the singleton/geminate contrast for liquids. We present evidence from an event-related potential (ERP) study testing the hypothesis that Korean listeners are more likely to process intervocalic single [l] (illegal in Korean) as geminate [ll] than as [ɾ], to explain why Korean listeners realize the English [r-l] contrast in terms of the Korean [ɾ-ll] contrast ([daɾi] 'bridge', [dalli] 'differently').

We conducted an ERP experiment in which Korean participants were presented with auditory stimuli consisting of Korean sentences and asked to decide whether the sentence was semantically well-formed. Three different types of auditory stimuli were created: (i) semantically well-formed sentences containing either an [ll]-word or a [ɾ]-word, each in a semantically congruent context; (ii) semantically ill-formed sentences in which an [ll]-word was replaced by an [ɾ]-word (or vice versa), where the substituted word was incongruent with the preceding context; and (iii) ambiguous sentences in which a context-appropriate word containing either [ɾ] or [ll] was edited to contain a single [l], inducing a change from the target word either in duration (ll→l) or in laterality (ɾ→l). We expected semantically anomalous sentences to elicit an N400, a negative-going ERP component found about 400 milliseconds following semantic stimuli that are incongruous given the preceding context. Given the loanword adaptation pattern, in which English single [l] is mapped to Korean [ll], we predicted that Korean listeners would also display a larger N400 to stimuli that were anomalous by virtue of a change in laterality than a change in duration. That is, the stimuli containing a word with single [l] should elicit a larger N400 in contexts where an [ɾ]-word was appropriate than in contexts where an [ll]-word was appropriate.

The behavioral results from 27 Korean participants show that, as expected, Korean listeners were much more likely to accept ambiguous sentences with a single [l]-word as well-formed when the [l]-word appeared in a context where an [ll]-word was expected (81.4%) than a context where an [ɾ]-word was expected (39.4%) (Figure 1). Similarly, a larger N400 amplitude was observed in response to an [l]-word presented after an [ɾ]-word context than in a context where an [ll]-word was expected (see the red line in the rectangle box in figure 2). However, although both the behavioral results and the N400 patterns were consistent with the adaptation of intervocalic single [l] as geminate [ll], N400 to /l/ forms was unexpectedly large in both contexts. That is, while almost 40% of the ambiguous sentences containing an [l]-word were accepted as well-formed, the N400 response elicited by the same sentences was as large as the response elicited by clearly ill-formed sentences. We propose that the N400 pattern reflects a response to the illegality of the single [l] stimuli at early levels of processing, while the behavioral results reflect the higher level mapping of these stimuli to legal native structures.

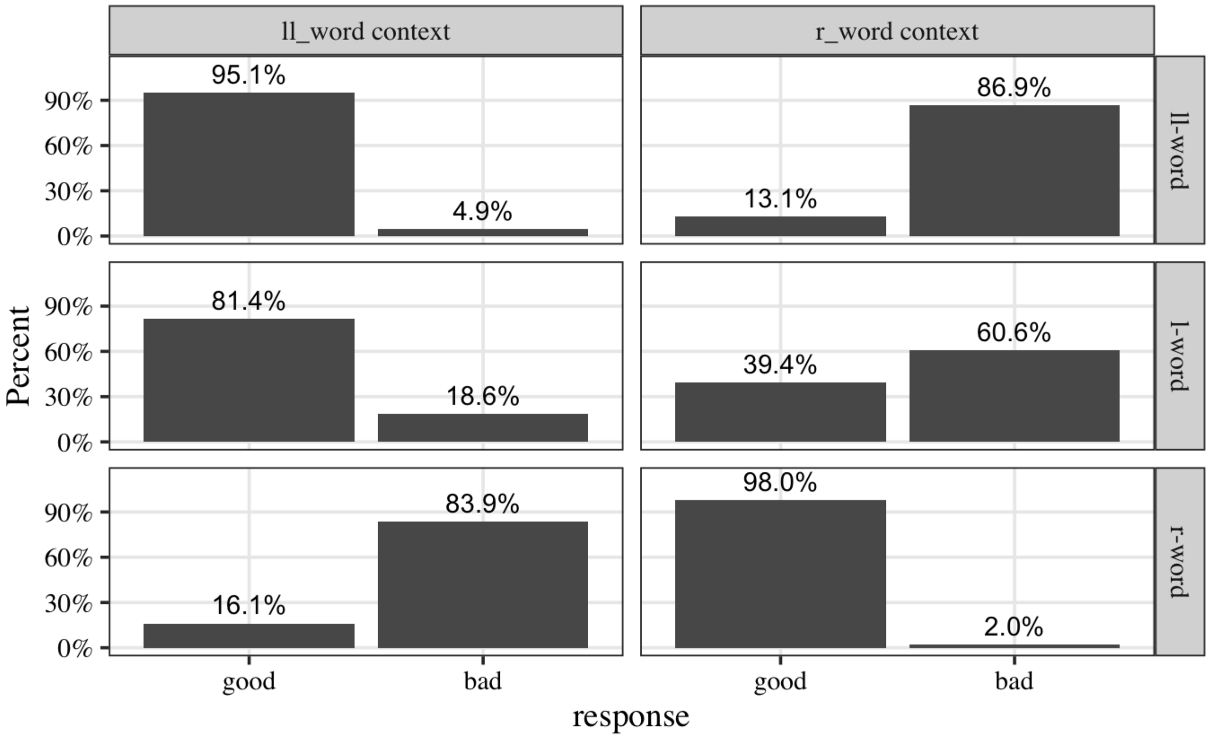


Fig 1. SENTENCE WELL-FORMEDNESS JUDGEMENT TASK

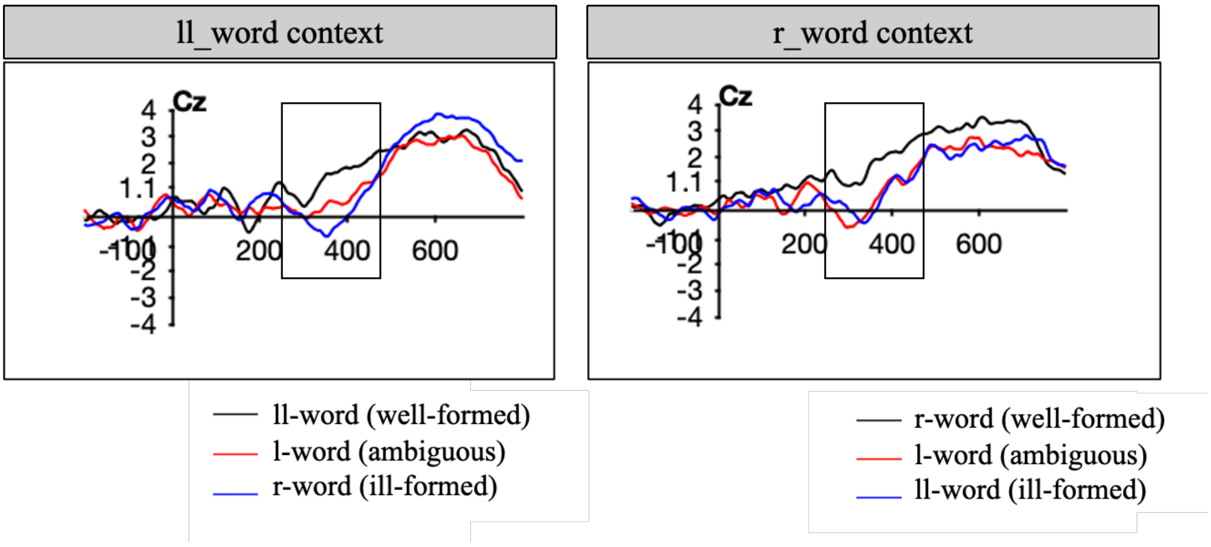


Fig 2. ERP WAVEFORMS AT ELECTRODE SITE CZ: WAVEFORMS REFLECT THE AVERAGE ACTIVITY AND THE ONSET OF DEVIANCE (BEGINNING OF THE VOWEL BEFORE [ll], [l], [ɾ] IN EACH TARGET WORD) IS INDICATED BY THE VERTICAL BAR

References

[1] Oh, M. (2005). Phonetic and spelling information in loan adaptation. *Korean Journal of Linguistics 20*, 347-368.

Phonetic cue-weighting in the production of Mandarin rising [T2] and low [T3] tones by Japanese learners

Yufei Niu[1] & Ricky Chan[2]

[12]*The University of Hong Kong (HK)*
yufeiniu@connect.hku.hk, rickykwc@hku.hk

L2 tone acquisition can be challenging and previous research has shown that T2/T3 confusion is common for L2 Mandarin learners [1]. For native Mandarin speakers, in addition to the primary F0 cue, voice quality (T3 allophonic creakiness) and duration can serve as important secondary cues in distinguishing T2 and T3 [2, 3]. However, the relative importance of these cues for L2 learners remains unclear. Similar to Mandarin, Japanese contrasts some words with F0 cues and it is interesting to determine if Japanese learners of Mandarin will use similar acoustic cues in lexical tone production. We also explore whether there is a shift towards a more native-like cue weighting pattern with increasing Mandarin proficiency.

We recruited two groups of Japanese learners of Mandarin (JAPH: high Mandarin proficiency and JAPL: low Mandarin proficiency) as the target groups and one group of native Mandarin speakers (MAN) as the control group. This contribution reports preliminary results from five speakers (2 from JAPH, 2 from JAPL, and 1 from MAN). Participants read aloud monosyllabic words in a Mandarin carrier phrase “wo du X zhe ge zi” (‘I read the word X’) in a quiet room, with 1000 tokens in total. 10 equidistant F0 measurements were obtained and F0 tone contours were modelled with quadratic polynomials ($y = a + bx + cx^2$), in which ‘a’ refers to intercept, ‘b’ to slope, and ‘c’ to curvature. For the secondary cues, we measured the duration and extracted the minimum values of six voice quality parameters (H1*-H2*, H2*-H4*, H1*- A1*, H1*-A2*, H1*-A3*, and CPP) from VoiceSauce. We then conducted the Linear Discriminant Analysis on T2 and T3 for each participant to examine the weighting of each acoustic parameter for distinguishing the two tones (Table 1). In addition, to explore the extent of creakiness in T3 production, we performed mixed-effects linear regression across language groups on each voice quality cue and the result of Tukey’s post-hoc test is shown in Table 2.

Figure 1 shows that the T3 contour of the JAPL group substantially rises in the later part (even reaching the height of T1 at the endpoint), probably because at the beginning stage, T3 contour is the focus of teaching and L2 learners were taught to produce a very complete T3. Then, with the improvement of Mandarin proficiency, L2 learners gradually got a better grasp of the T2/T3 distinction and their T3 contours converged towards that of native speakers (only a small rise in the second half). For the cue-weighting pattern, results in Table 1 suggest that F0 cues are the most robust, accounting for the top three weights across language groups (except for JAPL2). Duration plays an essential role for low-proficiency learners and relatively speaking, it has more weight in the JAPL group compared to JAPH and MAN group, i.e., when F0 cannot clearly distinguish T2 and T3 for beginners, secondary cues such as duration comes into play. However, as Mandarin proficiency improves, L2 learners tend to weigh more on voice quality than duration, which is consistent with native speakers’ cue-weighting pattern. For the specific voice quality cues, there are between-group differences in most of them except for H2*-H4* and H1*-A1*. All the other parameters were significantly lower for the MAN group than for the L2 learners, and the JAPH group had significantly lower values for H1*- H2* and CPP than the JAPL group (Table 2). Since lower values suggest a higher degree of creakiness, we may infer that high-proficiency learners produce creakier T3 than their low counterparts, indicating a native-like tendency. In sum, this preliminary study suggests that the cue-weighting pattern in Japanese learners’ T2/T3 distinction skew towards native speakers as their Mandarin proficiency increases.

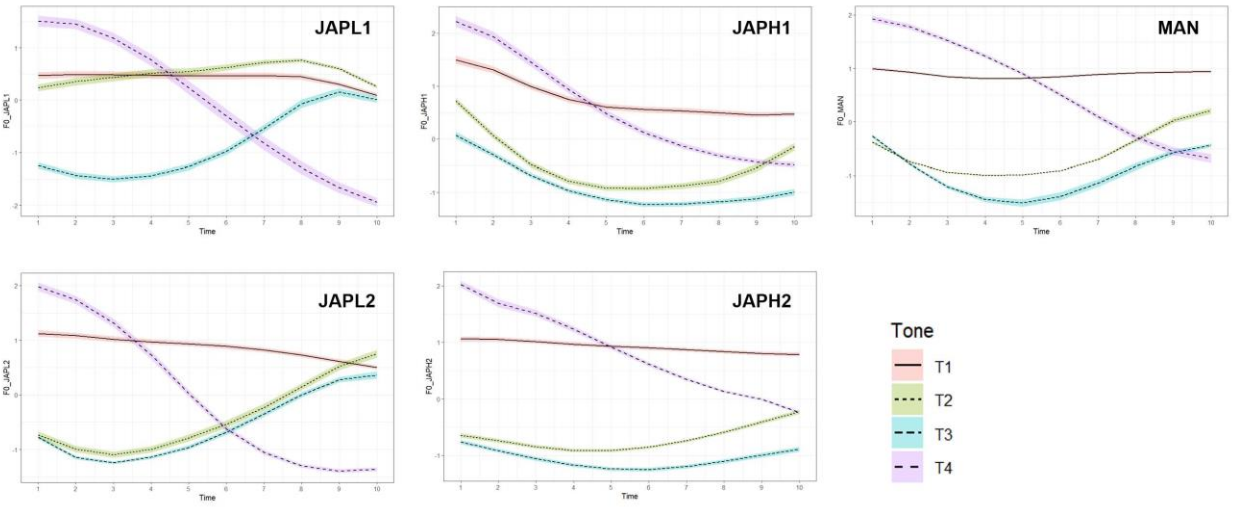


Figure 1. Individual tone contour patterns across language groups

Table 1. Individual LDA results of Mandarin T2/T3 distinction (The five highest weighted cues are bolded)

	JAPH1	JAPH2	JAPL1	JAPL2	MAN
a	**-1.4804**	**-1.6223**	**-1.5180**	**-0.6855**	**-0.6327**
b	**-3.0432**	**-3.7727**	**-0.6433**	**-2.1609**	**-5.2126**
c	**-3.1628**	**-2.6683**	**0.9427**	**-1.8471**	**-4.0688**
duration	0.1973	**0.2145**	**0.5627**	**1.1038**	**0.2647**
H1*-H2*	-0.1852	**-0.2747**	-0.0396	**0.4441**	**-0.6252**
H2*-H4*	-0.2758	-0.0106	0.1759	0.0027	-0.2041
H1*-A1*	**0.3064**	-0.1218	**-0.4124**	-0.2205	0.1051
H1*-A2*	**-0.3992**	-0.0010	-0.1829	-0.1391	0.1549
H1*-A3*	0.2061	-0.0210	0.3415	-0.0298	-0.2032
CPP	-0.1629	-0.1472	0.1290	-0.2125	0.0715

Table 2. Results of Tukey's post-hoc comparisons of voice quality parameters in T3 production

	Language	estimate	SE	*p*.value
H1*-H2*	JAPH - JAPL	-3.11	0.53	<.0001
	JAPH - MAN	2.88	0.83	0.0050
	JAPL - MAN	5.99	0.83	<.0001
H1*-A2*	JAPH - JAPL	0.80	1.37	0.8283
	JAPH - MAN	7.08	1.85	0.0005
	JAPL - MAN	6.28	1.88	0.0029
H1*-A3*	JAPH - JAPL	-1.87	1.33	0.3395
	JAPH - MAN	6.32	2.05	0.0127
	JAPL - MAN	8.18	2.06	0.0011
CPP	JAPH - JAPL	-1.29	0.16	<.0001
	JAPH - MAN	1.94	0.20	<.0001
	JAPL - MAN	3.23	0.21	<.0001

References

[1] Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonallanguage speakers. *Journal of Phonetics, 40*(2), 269–279.

[2] Van Way J (2014) Effect of creaky voice simulation on third-tone perception in Mandarin Chinese. *University of Hawai'i at Mānoa: Working Papers in Linguistics* 45(3): 1-13.

[3] Yang, J., Zhang, Y., Li, A., & Li, X. (2017). On the duration of Mandarin tones. *Proceedings of theAnnual Conference of the International Speech Communication Association*, 1407–1411.

Intergestural CV timing of homophonous words with different morphological structures: A preliminary report on liquid /l/ in Korean

Jiyoung Lee[1], Sahyang Kim[2], and Taehong Cho[1]
[1]*Hanyang Institute for Phonetics & Cognitive Sciences of Language, Hanyang University (Korea),*
[2]*Hongik University (Korea)*
Ljy1004@hanyang.ac.kr , sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

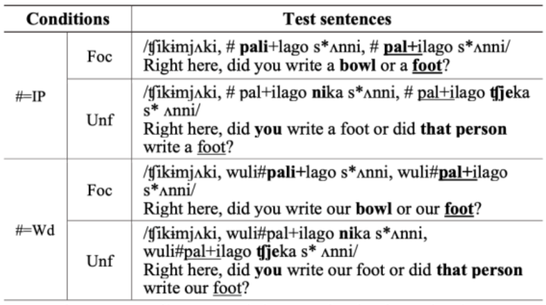
The Korean liquid phoneme /l/ has several allophones, which are highly dependent on its position in syllable (e.g., a lateral in coda and a flap in onset or intervocalic position) [1]. This EMA study investigates whether and how the same allophone of /l/ may be distinguished in the articulatory dimension when it has different underlying syllable compositions. To examine this question, homophonous sequences with different underlying structures were compared. Two pairs of $C_1V_1\underline{\mathbf{C_2}}$-$C_1V_1.\underline{\mathbf{C_2}}V_2$ words (e.g., /pal/ 'foot' - /pali/ 'bowl' where /l/ is underlyingly in the coda or in the intervocalic onset position) were used, and they were followed by the grammatical particle /(i)lago/. In this way, C_2, /l/, in the words with different syllable structures would both be produced as a flap at the surface phonetic level (e.g., a derived flap condition as in /pal+ilago/ and an underlying flap condition as in /pali+lago/).

A couple of competing predictions can be made on the production of the two types of flaps. One possibility is that there may be no articulatory differences between the two homophonous sequences. This is based on previous studies suggesting that the spelled-out segments go through the same process and create the syllabified form, resulting in little chance to reflect its internal syllable structure in production [2]. Alternatively, some differences may be observed between the derived versus the underlying flap. In Articulatory Phonology (AP) [3, 4], gestural coordination is specified in the lexicon, and understood by phase relations within a syllable. AP assumes that the CV sequence shows an in-phase relationship, and the two gestures start almost synchronously. Based on the assumption, the underlying flap in /pali/ should show an in-phase relationship while the resyllabified flap in /pal+i/ may not show the same pattern as its relationship, which is not specified in the lexicon. Some differences in gestural coordination depending on underlying compositions have indeed been found in previous studies [5, 6]. This study further tests whether and how prosodic prominence modulates the potential underlying structural differences. It is well-known that the presence of prominence contributes to maximizing lexical contrasts and enhancing the gestural bonding strength [7, 8, 9, 10]. It is therefore hypothesized that, if any, the structural differences would be maximized under prominence.

Articulatory data were collected using EMA (AG501, Carstens Electronics) from twelve native Seoul Korean speakers but only a subset of the data was analyzed here (4F, 3M). As shown in Table 1, two prosodic factors were manipulated: Boundary (IP-initial or Wd-initial) and Focus (Focused or Unfocused). Each speaker produced 240 test sentences (4-target * 2-boundary * 2-focus * 15-repetition). By excluding 45 tokens with unintended prosodic renditions, 1635 tokens were collected for analysis. Kinematic data from tongue tip (TT) for the consonantal /l/ gesture and tongue body (TB) for the vocalic gestures were analyzed by using MVIEW [11]: C_2 duration (from the onset to the target of C_2 /l/), C_2V_2 duration (from the onset of C_2(or V_2[i]) to the target of V_2), Intergestural timing (between the onsets of C_2 and V_2[ii]). A series of linear mixed-effects models were fitted separately for each measurement.

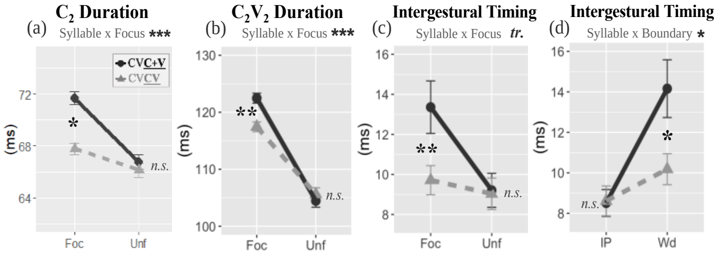
Results showed that there was a temporal difference in the articulatory duration of TT movement for /l/ as a function of underlying structure, but only in the focused condition. While /l/ was realized as a flap on the surface in both conditions, the resyllabified flap ($C_1V_1\underline{C_2}$+V_2), a lateral underlyingly, was longer than the canonical flap ($C_1V_1.\underline{C_2}V_2$) under focus (Fig. 1a, p=.028). This is in line with the lexical contrast maximization under prominence reported in previous studies [8, 10]. The results, therefore, seem to indicate that when there is a need to deliver an informational locus, speakers put a deliberate effort into making a distinction between the two homophonic sequences by referring to the underlying syllable structures. Another important finding was that C_2V_2 duration was shorter under focus in the monomorphemic ($C_1V_1.\underline{C_2V_2}$) than in the heteromorphemic condition ($C_1V_1\underline{C_2+V_2}$) (Fig. 1b, *p*=.0029), showing underlying syllable differences. In addition, the absolute distance between the C_2 and V_2 gestures (i.e., intergestural timing between C_2 and V_2 gestures) was again shorter for $C_1V_1.\underline{C_2V_2}$ than for $C_1V_1\underline{C_2+V_2}$ in the focused condition (Fig. 1c, *p*=.009). According to Articulatory Phonology [3, 4], the gestural coordination and corresponding phase relationship is specified in the lexicon, and C and V gestures are assumed to start synchronously as they are in an in-phase relationship. Thus, the smaller temporal interval between C and V gestures can be understood as a tighter in-phase relationship with stronger gestural cohesiveness. The differences were further augmented under prosodic prominence, suggesting that speakers make efforts to differentiate the underlying structural difference. What is interesting is the difference in intergesutral timing between the two homophonous sequences was found in the Wd condition. A stronger prosodic boundary is generally known to induce a stronger bonding relationship between C and V [7, 9], resulting in shorter temporal interval in the IP-initial condition (Fig 1d). On the contrary, in the phrase-medial position, loosened gestural bonding strength may display the underlying structural differences, resulting in a longer temporal interval between C and V gestures especially for $C_1V_1\underline{C_2+V_2}$ than in $C_1V_1.\underline{C_2V_2}$ (Fig 1d, *p*=.02). In conclusion, the present study suggests that speakers fine-tune the articulatory realization of gestures and their coordination to encode and maintain the underlying structural difference, further modulating them by referring to higher-order prosodic structure.

Table 1. Examples of test sentences. Targets are underlined, and contrasted words are marked in bold. '#' and '+' refer to phrase boundary and morphological boundary, respectively.



Conditions		Test sentences
#=IP	Foc	/ʧikɨmjʌki, # **pali**+lago s*ʌnni, # **pal+i**lago s*ʌnni/ Right here, did you write a **bowl** or a **foot**?
	Unf	/ʧikɨmjʌki, # pal+ilago **ni**ka s*ʌnni, # pal+ilago **ʧje**ka s* ʌnni/ Right here, did **you** write a foot or did **that person** write a foot?
#=Wd	Foc	/ʧikɨmjʌki, wuli#**pali**+lago s*ʌnni, wuli#**pal+i**lago s*ʌnni/ Right here, did you write our **bowl** or our **foot**?
	Unf	/ʧikɨmjʌki, wuli#pal+ilago **ni**ka s*ʌnni, wuli#pal+ilago **ʧje**ka s* ʌnni/ Right here, did **you** write our foot or did **that person** write our foot?

Figure 1. Syllable x Focus interaction on C_2 duration (a), C_2V_2 duration (b), and intergestural timing (c). Syllable x Boundary interaction on intergestural timing (d). Error bars represent standard errors ('*' refers to *p*<.05, '**' to *p*<.01, and '***' to *p*<.001).



References

[1] Lee, S. & Kang, S. (2003). Acoustic and phonological properties of Korean liquids. Journal of Language Sciences, 10(2), 77-92.

[2] Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and brain sciences, 22(1), 1-38.

[3] Browman, C.P., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. Phonetica, 45, 140–155.

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

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

[6] Lee, J., Kim, S., & Cho, T. (2019). Effects of morphological structure on intergestural timing in different prosodic-structural contexts in Korean. In Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia.

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

[8] Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. Journal of Phonetics 39.

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

[10] de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. The journal of the acoustical society of America, 97(1), 491-504.

[11] Tiede, M. (2010). MVIEW: Multi-channel visualization application for displaying dynamic sensor movements.

[i] This is the case when the V_2 gesture starts earlier than the C_2 gesture.

[ii] In order to measure the synchronicity of the two gestures, the actual degree of proximity between the onsets of C_2 and V_2 was measured as an absolute value.

Estimating tongue stiffness during phonation using ultrasound passive shear wave elastography

Chenhao Chiu[1], Wei-Cheng Hsiao[2], Huang-Yu Shih[2], Bao-Yu Hsieh[2,3], Yining Weng[1]

[1]National Taiwan University (Taiwan), [2]Chang Gung University (Taiwan), [3]Chang Gung Memorial Hospital at Linkou (Taiwan)

chenhaochiu@ntu.edu.tw, v25293626@gmail.com, lamboipe888@gmail.com, byhsieh@mail.cgu.edu.tw, wynjdws@gmail.com

Recent biomechanical modeling studies have proposed that sound variants or alternations may be associated with physiological preferences, such as muscle-induced stress and strain of the tongue surface [1,2]. But the account of physiological preference was largely based on simulation results, while *in vivo* measurements on the articulatory apparatus (e.g., the tongue) remain sparse. A recently developed technique of ultrafast imaging captures up to thousands of ultrasound images per second, making it possible to analyze the mechanical waves that travel through the scanned tissues [3,4]. By calculating the amplitude and velocity of this mechanical wave, the stiffness of the scanned tissues can then be estimated (a method known as ultrasound elastography). Miura et al. [5] employed ultrasound elastography with ultrafast imaging to estimate the stiffness (hardness, in their term) and pressure of the tongue when external forces were applied. However, it is yet to be determined whether the stiffness of the tongue may be different during phonation, or vary during productions of different sounds.

We used ultrasound passive elastography with ultrafast imaging to capture the state of the tongue during vowel articulation. One native speaker of Mandarin participated in data collection. The speaker produced multiple repetitions of Mandarin [a], [i], and [u] vowels, all in high level tone. Each vowel vocalization lasted for 4 ~ 5 seconds. During each vocalization, a roughly 100 ms long ultrasound video was captured at a frame rate of 3k to 9k fps. Ultrasound ultrafast imaging allowed us to track the propagation of the mechanical waves produced by the intrinsic vibrations of the vocal folds through the tongue tissues. We then measured the wave amplitude, propagation velocity, and the associated stiffness of the tongue, as well as spectral properties of the mechanical waves such as vibration frequency. Additionally, acoustic recordings were carried out in parallel with ultrasound data collection. The acoustic fundamental frequencies (F0) of vowel vocalizations were calculated for cross comparison.

Our results show that the frequencies of the mechanical waves measured at the tongue match with those measured from acoustic recordings (Fig. 1), affirming that the vibrations of the vocal folds propagate mechanical waves all the way through the tongue. This suggests that the vocal folds can serve as an intrinsic source of vibration, suitable for shear wave estimation and quantification of tongue stiffness using ultrafast imaging. The propagating velocities of the mechanical waves also differ for each vowel (Fig. 2, bottom), indicative of vowel-dependent local tongue stiffness. Taken together, the findings of the current study demonstrate the viability of ultrasound passive shear wave elastography for examining speech production. This technique could be a promising tool for probing into the detailed physiological operations of the speech apparatus during articulation, and potentially further account for patterns of sound combinations or changes.

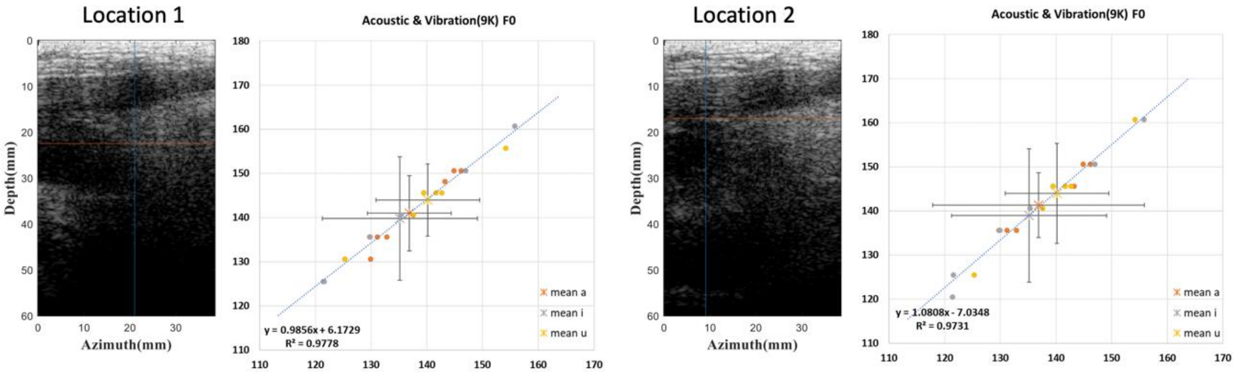


Fig.1 Correlations between estimated ultrasound vibration frequency (y axis) and recorded acoustic F0 frequency (x axis, both in Hz) at two different locations of the midsagittal tongue. The B-mode images indicate where the ultrasound vibration frequency was obtained.

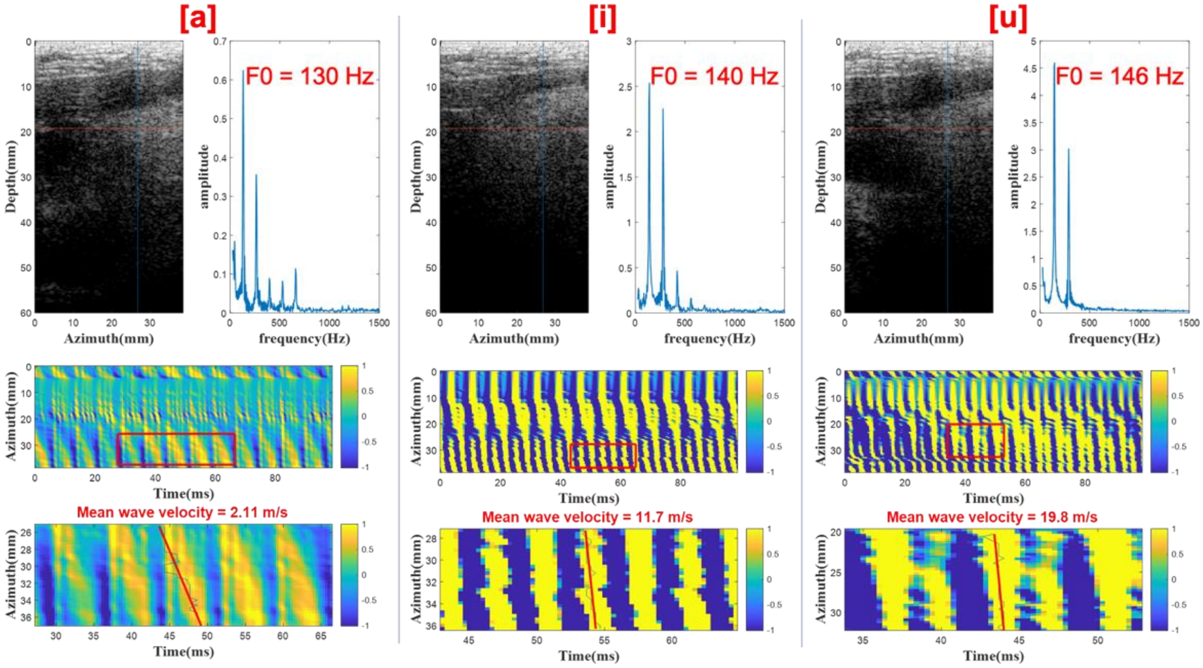


Fig.2 Mechanical vibrations of tongue tissues at the intersection of the blue and orange lines in the B-mode images (top left) were submitted to spectral analysis (top right). Shear wave velocities were calculated for each vowel and visualized in M-mode images (bottom).

References

[1] Stavness, I., Gick, B., Derrick, D., & Fels, S. (2012). Biomechanical modeling of English/r/variants. *The Journal of the Acoustical Society of America*, *131*(5), EL355-EL360.

[2] Chiu, C., & Lu, Y. A. (2021). Articulatory evidence for the syllable-final nasal merging in Taiwan Mandarin. *Language and Speech*, *64*(4), 771-789.

[3] Jing, B., Tang, S., Wu, L., Wang, S., & Wan, M. (2016). Visualizing the vibration of laryngeal tissue during phonation using ultrafast plane wave ultrasonography. *Ultrasound in Medicine & Biology*, *42*(12), 2812-2825.

[4] Jing, B., Ge, Z., Wu, L., Wang, S., & Wan, M. (2018). Visualizing the mechanical wave of vocal fold tissue during phonation using electroglottogram-triggered ultrasonography. *The Journal of the Acoustical Society of America*, *143*(5), EL425-EL429.

[5] Miura, K., Ohkubo, M., Sugiyama, T., Tsuiki, M., & Ishida, R. (2021). Determination of the Relationships Between intra-and Extraoral Tongue Hardness, Thickness, and Pressure Using Ultrasonic Elastography. *Dysphagia*, *36*(4), 623-634.

The effects of ultrasound image feedback on Korean L2 learners' production of English /ɹ/ in production training

Joo-Kyeong Lee

University of Seoul
jookyeong@uos.ac.kr

This study investigates the effects of production training with using ultrasound image feedback on Korean L2 learners' production accuracy of English retroflex. It has been widely known that Korean learners of English have a problem with production and perception of English /ɹ/ and /l/. It may arise from a one-to-two match relation between the Korean L1 phoneme /l/ and the English L2 phonemes /ɹ/ and /l/. Moreover, both Korean /l/ and English /ɹ/ and /l/ have two allophonic variants respectively depending on syllable position. When the variants are further considered, the learning mechanism of English /ɹ/ and /l/ may be more complicated. Numerous studies have examined L2 learners' production and perception of the sounds and extended to L2 perception training (Aoyama et al., 2004; Bird & Gick, 2018; Bradlow et al., 1999 among others), but there have been very few about production training, particularly using a high technology methodology, for instance, 'ultrasound images.' This sheds light on the current work of articulatory training of the English L2 retroflex with ultrasound imaging feedback and its effect on learning improvement.

In the experiment, nine Korean learners of English, who were rated as intermediate in English proficiency from a Foreign Accentedness (FA) task, participated in the production training sessions of English /ɹ/ with ultrasound imaging feedback. They went through 6 sessions of production training and took three production tests: a pretests, a posttest and a generalization test. In each test, they recorded 72 English words containing /ɹ/ (24 words * 3 repetitions) in onset and coda positions respectively and a half of them were minimal pairs with /l/. In the posttest and the generalization test, 12 novel words each were used. The posttest was carried out right after they finished the last training session, and the generalization test was given a month away from the last training. Participants were provided a 10-minute instruction about the speech organs and the tongue tip movement during /ɹ/ articulation only in the first training. Each training lasted one hour where a learner read a list of sentences 'Please say _____ for me' while they looked at the ultrasound videos of their production. When the tongue tip was not satisfactorily raised and curled back, the instructor showed a native speaker's ultrasound image of the same word for corrective feedback. The instructor did not present any oral explanation about how the trainees were incorrect. Ultrasound images were aligned with spectrograms and splined at the midpoint of the acoustic duration of /ɹ/. Spline coordinates (over 42 points) were extracted and submitted to SSANOVA in R for the tongue contour analysis.

Results showed that Korean learners of English with intermediate proficiency mostly produced the Korean tap [ɾ] in onset position before training as shown in Figure 1. However, they successfully raised up the tongue tip to make a retroflex in coda position though the tongue tip curling-back gesture was not sufficiently distinctive. In the posttest, the retroflexation articulation of /ɹ/ was successfully produced in both onset and coda similarly to native speakers' retroflex as in Figure 2. In the generalization test, individual differences were found; one of them failed to maintain the retroflexation gesture in onset position, and two of them showed that its degree decreased. For the production of coda /ɹ/, most of the participants maintained their learning of tongue tip curling gesture while tongue tip rising was unwaveringly observed. One of the participants showed a hyperarticulation; the tongue tip rising was toward further back of the hard palate.

The corrective feedback of ultrasound image seemed to be significantly effective on Korean learners' accuracy of English onset /ɹ/ production as most participants showed significant improvement after training. The ultrasound imaging of a retroflexation gesture seemed to be satisfactorily applied to a participant' motor control over articulating the English retroflex.

Korean learners of English with intermediate proficiency were likely to be accurate at producing the coda retroflex before training though the location and the degree of tongue tip raising was various. Clear-[l], which is allophonically produced in coda position of Korean, is phonetically and in some languages phonologically very distant from [ɹ]. According to the L2 Speech Learning Model (Flege, 1995) that an L2 sound, which is more distant from its corresponding L1 sound, is perceptually more accessible and ultimately easier to learn to produce, the coda [ɹ] has been learned to a great extent. On the other hand, both tap and retroflex involve a concave tongue shape and a lower F3 in CV structure (Cathcart, 2012). This serves a substantial account for the inaccurate production of the onset /ɹ/ before training. Korean /l/ is allophonically realized as [ɾ] in onset position, and the English [ɹ] is phonetically closer to the Korean tap [ɾ], which may result in more perceptual confusion. It will therefore take more time to learn to perceive and produce, and even when it is learned, it will be hard to maintain the learning.

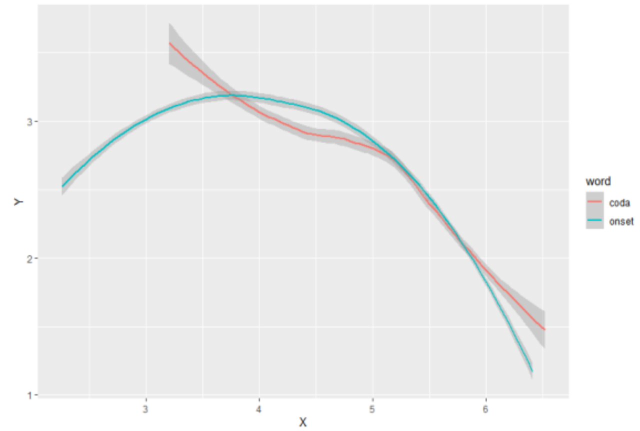


Figure 1. Smooth splines of [ɹ] in pretest

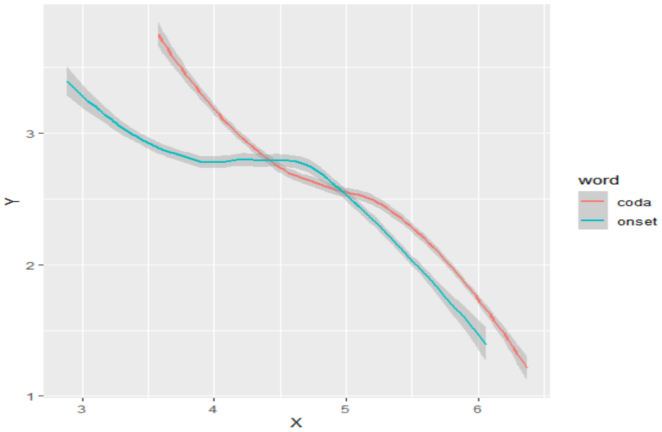


Figure 2. Smooth splines of [ɹ] in posttest

References

[1] Aoyama, K., Flege, J. E., Guion, S. G., Akahane - Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* , 32 (2), 233 - 250.

[2] Bird, S. and Gick, B. (2018) Ultrasound biofeedback in pronunciation teaching and learning. *Proceedings of the 2nd International Symposium on Applied Phonetics*. (ISAPh 2018), 5 - 11.

[3] Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. Perception & psychophysics , 61 , 977-985.

[4] Cathcart, C. (2012) Articulatory variation of the alveolar tap and implications for sound change. UC Berkeley Phonology Lab Annual Report, 76-110.

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

Focus Prosody in Fijian: a Pilot Study

Tsz Ching Mut[1], Candide Simard[2], Apolonia Tamata[2] & Albert Lee[1]

[1]*The Education University of Hong Kong (Hong Kong),* [2]*University of the South Pacific (Fiji)*
stcmut@gmail.com, candide.simard@usp.ac.fj, apolonia.tamata@usp.ac.fj, albertlee@eduhk.hk

Studying focus prosody of verb-initial languages can be difficult because non-prosodic focus markers such as fronting (i.e. moving the focussed item to the front) are often involved (see review in [1]). For example, in Samoan fronting can be used to mark (contrastive) focus [2], alongside prosodic markers. This means that comparisons of prosodic cues among focus conditions are not based on otherwise identical utterances – a potential source of confounds. Possibly in part due to this challenge, Fijian focus prosody has yet to be empirically investigated with any systematic production experiment, making it an understudied topic in phonetics to this day.

Fijian is an Austronesian language spoken by about 400,000 as a first language [3] in Fiji. Its basic word order is often considered verb-object-subject (but note alternative accounts such as [3]), and focus is often marked by word order [4]. Currently, there is no published production study of Fijian focus prosody (except one pilot study [5]).

While not much is known about Fijian focus prosody, researchers have investigated a related language verb-initial Samoan [2]. It was found that individual speakers varied in focus-marking strategies. The initial phonological phrase was always the most prominent. In verb-agent-object sentences, the verb and agent were in the initial phrase. Speakers raised the accent on the object in object focus, and lowered it in agent focus; although they did not do this consistently. No prosodic marking of focus on the agent was found.

To elicit prosodic focus markers in situ with fronting suppressed, one possible strategy is by avoiding natural sentences. Alternatively, one could use strings such as phone numbers or, in the present study, items. With such a paradigm, one could answer research questions such as: (i) Are narrow focus different from neutral focus? (ii) Is narrow focus marked differently across different locations? (iii) What acoustic cues (e.g. f_0, intensity, duration) are used to mark focus? We designed a production task using item strings to answer these questions.

Ten native speakers of Fijian from the University of South Pacific were recruited. They have no (history of) hearing or language impairment. Participants completed a sequence naming task. The sentences are composed of three adjacent noun phrases (NPs), i.e. *uvi, uto, dalo* 'yam, fruit, taro'. Each sentence has four focus conditions based on the NP positions, namely initial focus, medial focus, final focus (i.e. narrow focus), and neutral focus (i.e. board focus) (Table 1). The focuses were elicited by the presentation of pictures of yam, fruit and taro, followed by a precursor question asked by the interviewer. Altogether, we recorded 120 utterances (1 sentence * 4 focus conditions * 3 repetitions * 10 speakers).

Figure 1 displays the f_0 of all focus conditions with SS ANOVA [6]. The neutral focus condition has significantly lower f_0 than the narrow focus conditions starting in the final word. We also fitted linear mixed effects models to the f_0 data using *lmerTest()* [7]. Model construction followed a bottom-up approach. Post-hoc comparisons were done using *emmeans()* [8]. The best fitting model contained the fixed factor of the focus condition (initial, medial, final, neutral), and by-subject random intercept. Intensity and duration data were analysed using the same approach. The main effect of focus on f_0 was significant, $X^2(3) = 508.57$, $p <.001$. Post-hoc test shows that initial, medial and final focus had significantly higher f_0 than neutral focus ($p <.0001$). It means that, regardless of focus locations, a general elevation of f_0 is observed for all narrow focus conditions.

Figure 2 shows that the intensity of the focus conditions of medial, final (narrow focus) and neutral focus. The main effect of focus on intensity was significant, $X^2(3) = 69.829$, $p <.001$ too. Post-hoc test indicates that intensity of final and medial focus is significantly greater than neutral focus ($p < .0001$), but the intensity of initial and neutral focus is not significantly different ($p = 0.9440$).

It is likely that in situ prosodic focus in Fijian is mainly marked by elevation of f_0 and intensity in narrow focuses, the difference in duration is not significant, $X^2(3) = 5.6281$, $p = 0.1312$. Generally speaking, mean syllable duration is longer for narrow focus conditions, but that for initial focus is the only exception, meaning that syllable duration is shorter for than neutral.

Our findings suggest that in situ prosodic focus in Fijian is mainly marked by a general elevation of f_0 in narrow focus conditions and increase in intensity (in medial and final focus conditions), and not by syllable duration. Although we have found that focus locations significantly affected prosodic focus markers in scripted sentences, there are different ways to mark focus in natural speech, in addition to, cross-speaker variability. To gain a better understanding of Fijian focus prosody, further systematic production studies of focus marking strategies are needed.

Precursor question	Target sentence	Focus condition
uto, *uto, dalo?*	***uvi***, *uto, dalo*	narrow (initial)
uvi, ***uvi***, *dalo?*	*uvi,* ***uto***, *dalo*	narrow (medial)
uvi, uto, ***uto***?	*uvi, uto,* **dalo**	narrow (final)
uvi, uto, dalo?	*uvi, uto, dalo*	broad (neutral)

Table 1: Summary of stimuli used

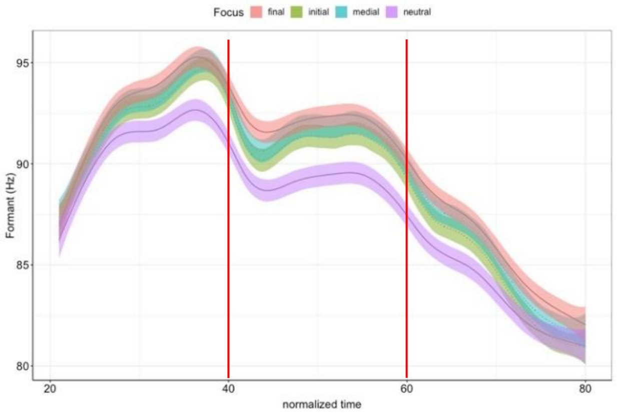


Figure 1: SS ANOVA comparing f_0 (Hz) of different focuses (word boundaries in red)

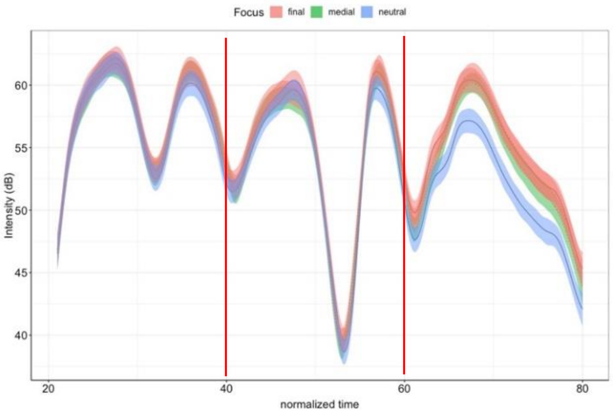


Figure 2: SS ANOVA comparing intensity (dB) of final, medial and neutral focus

References

[1] Kügler, F., Calhoun, S. 2020. Prosodic encoding of information structure: A typological perspective. In C. H. M. Gussenhoven & A. Chen (eds.), *The Oxford Handbook of Language Prosody*. Oxford University Press. 454–467

[2] Calhoun, S. 2015. The interaction of prosody and syntax in Samoan focus marking. *Lingua*, 165, 205–229.

[3] Geraghty, P. A. 2006. Fijian. In K. Brown (ed.), *Encyclopedia of Language and Linguistics* (2nd ed.). Elsevier, 465.

[4] Milner, G. B. 1989. On prosodic relations between Fijian bases and verbal suffixes. In J.H.C.S. Davidson (ed.) *South-East Asian Linguistics*. University of London, 59-88.

[5] Albert Lee, Candide Simard, Apolonia Tamata, Jiaying Sun, Tsz Ching Mut. Submitted. Focus prosody in Fijian: a pilot study. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*. Prague.

[6] Davidson, L. S. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *J. Acoust. Soc. Am.* vol. 120, no. 1, 407–415.

[7] Kuznetsova, A. Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest package: Tests in linear mixed effects models. *J. Stat. Softw.*, vol. 82, no. 13, 1–26.

[8] Lenth, R. V. 2020. Emmeans: Estimated marginal means, aka least-squares means. [Online]. Available: https://cran.r-project.org/package=emmeans.

Social evaluations of speech vary as a function of perceived speaker nativeness

Wai Ling Law[1], Olga Dmitrieva[2], Lan Li[3] & Jette Hansen Edwards[4]

[1]The Hong Kong University of Science and Technology (Hong Kong), [2]Purdue University (USA), [3]The Chinese University of Hong Kong, Shenzhen (China), [4]The Chinese University of Hong Kong (Hong Kong)
lcclaw@ust.hk, odmitrie@purdue.edu, lanli@cuhk.edu.cn, jhansen@cuhk.edu.hk

Foreign accented speakers and speakers of regional varieties are consistently rated less favorably on competence traits than native speakers and speakers of standard varieties, although they can be rated more favorably on solidarity traits (see [1] for a meta-analysis). However, the relationship between raters' own linguistic group membership and their social evaluations of talkers with different linguistic backgrounds remains unclear. It is also unclear whether and how prestige associated with different languages within a speech community affects social evaluation of speech. Therefore, the present study investigates the effects of listeners' native-speaker status and perceived nativeness of the talkers on the social evaluations of talkers' speech.

Following a matched-guise approach, we recorded four speakers from Hong Kong who are trilingual in Cantonese, English and Mandarin (in this proficiency order) reading a semantically neutral text in each language. All four spoke Cantonese as their first language (L1) and varied in the degrees of Cantonese accent in their English and Mandarin pronunciation. We recruited 576 trilingual participants from Mainland China to act as raters in the study. They were asked to evaluate the recordings on competence and solidarity traits using a seven-point semantic differential scale. The raters were also asked to judge how native they thought each talker sounded in each of the three languages. Among the raters, 378 spoke Cantonese as L1 and Mandarin as L2, while 198 spoke Mandarin as L1 and Cantonese as L2. Raters were not made aware that the same four speakers provided speaking samples in the three languages.

The evaluation of perceived talker nativeness was not affected by the raters' native language. All four talkers were perceived as native Cantonese speakers. Talker 1 was judged to speak both Mandarin and English with a strong Cantonese accent. Talker 3 was rated as near native-like in English and heavily Cantonese-accented in Mandarin. Talkers 2 and 4 were rated as near-native in English and moderately (Talker 2) to mildly (Talker 4) Cantonese-accented in Mandarin.

Separate repeated measures ANOVAs with competence and solidarity ratings as dependent variables and guise (Cantonese, English, or Mandarin) as well as talker identity as independent factors revealed significant main effects for guise and for talker identity, and significant interactions between the two in each model. The rating patterns were not visibly affected by raters' native language. That is, both native speakers of Cantonese and native speakers of Mandarin were affected by guise and talker identity/perceived accentedness in similar ways in their evaluations. Specifically, Mandarin guise elicited the lowest ratings on both solidarity and competence, while English guise received the highest ratings on both attributes, with Cantonese guise occupying the middle position (see Fig. 1 and Fig. 2). The pattern may reflect a combination of the effect of perceived accentedness of the talkers as a group and the perceived prestige of the language spoken. The majority of the speakers were perceived as strongly to moderately accented in Mandarin, leading to low social evaluations. While all were perceived as native in Cantonese, most were also perceived as near-native in English, which, combined with the prestige of English, could have resulted in higher ratings.

As to the effect of the talker identity, on average, Talker 4 was ranked the most favorably, in agreement with this talker's lowest perceived degree of accentedness in both Mandarin and English. In contrast, Talker 1 was rated the lowest on average, in accordance with the high degree of accentedness perceived in this talker's Mandarin and English speech. The remaining two talkers patterned in the middle.

Nevertheless, the interaction between guise and talker identity indicated that these two factors were not independent of each other. The interaction pattern that emerged suggested that talkers' levels of accentedness in each specific guise is what guided listeners' evaluations. For example, in the English guise, talker 1 (perceived as strongly accented in English) was consistently ranked the lowest on both solidarity and competence, while the three remaining talkers (all judged to be near-native in English) received higher ratings. In the Mandarin guise, talkers 1 and 3 (perceived as strongly accented in Mandarin) were judged less favorably on solidarity and competence than talkers 2 and 4. In each language, talkers who are perceived to be more native-like are considered more competent and more attractive.

Interestingly, despite the equally native Cantonese speech, the four talkers received different competence and solidarity ratings in their Cantonese guises, as well as in their Mandarin and English guises. This result suggests that evaluations were not affected by the perceived accentedness alone and that other qualities of the talkers' voices contributed to their ratings on solidarity and competence.

Overall, these findings suggest that listeners use exemplars associated with social and linguistic groups as a reference in social evaluations of speech [2]. Listeners' evaluation of the talkers' personal attributes is strongly affected by the perceived nativeness of the talkers' speech, independently of the listener's own nativeness status. That is, both native and non-native speakers of Mandarin penalize perceived accentedness in Mandarin speech. In addition, the prestige associated with different languages in a given speech community plays an important role. Speakers demonstrating native-like competence in prestigious languages such as English may be evaluated more favorably than native speakers of less prestigious languages such as Cantonese even by the L1 listeners of these less prestigious languages.

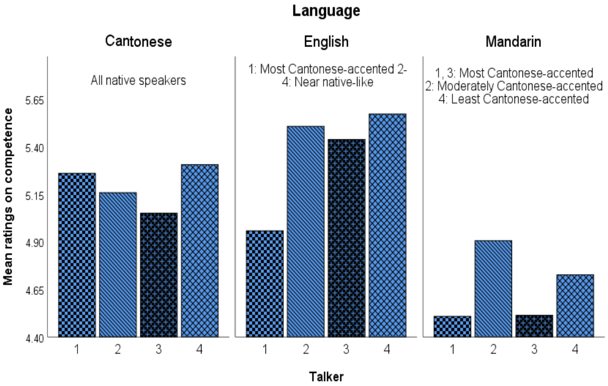


Fig. 1 The mean ratings of each talker in each language on competence (collapsing across L1 Cantonese and L1 Mandarin listeners).

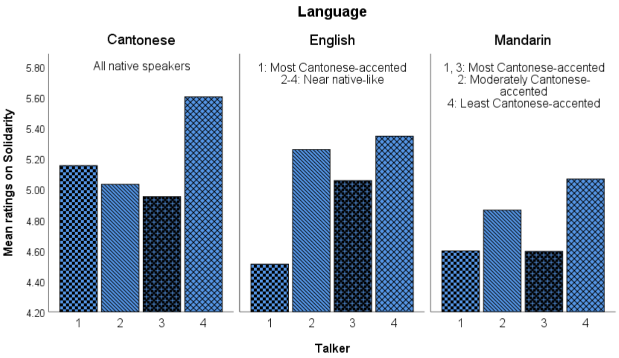


Fig. 2 The mean ratings of each talker in each language on solidarity (collapsing across L1 Cantonese and L1 Mandarin listeners).

References

[1] Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology, 42*, 120–133.

[2] Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 23*(3), 351–379.

Dialect levelling across generations: A socio-phonetic study of the medial [i] and vowel shift in the Jin dialect spoken in Baotou, China

Xinyue LIU & Peggy MOK

The Chinese University of Hong Kong (Hong Kong, China)

xinyueliu@cuhk.edu.hk, peggymok@cuhk.edu.hk

In this study, the medial [i], a vernacular and unique feature of the Jin dialect which is not found in Mandarin Chinese, was investigated to explore the dialect levelling and the vowel shift phenomena in *Baotou* from a socio-phonetic perspective.

The Jin dialect is a variety of Chinese that is mainly spoken in the northern parts of China [1]. It was not regarded as a separate dialect type of Chinese until the 1980s, when Li [2] proposed that the Jin dialect should be separated because of its entering tones which are usually marked with the glottal stop [ʔ]. The mutual intelligibility rate between the Jin dialect and Mandarin Chinese is around 60% regarding the lexicon [3].

Baotou is an immigrant city of northern China which has undergone two large-scale waves of immigration and most of the locally-born residents are still using the Jin dialect with high proficiency. Influenced by the popularization of Mandarin Chinese as a supralocal language with higher prestige and the frequent contact with Mandarin-speaking immigrants, the Jin dialect spoken in *Baotou* has undergone some changes. Due to language contact [4], the medial [i] as a traditional variant in the Jin dialect may be levelled down under the influence of Mandarin Chinese which does not have the medial [i] in the same phonetic contexts as the Jin dialect. The medial [i] is possible to disappear under the force of dialect levelling.

To investigate the change of the medial [i], four sociolinguistic factors were considered in the study: (1) AGE of the speaker (younger vs. older); (2) GENDER of the speaker (female vs. male); (3) language STYLEs of the speech (informal and casual interview; careful picture description; formal wordlist reading); and (4) language ATTITUDES of the recorded participants (scores of attitudinal questionnaires as a continuous variable).

Two types of Chinese characters with the medial [i] in the Jin dialect were used in the study. The ORI (original) type of characters with the medial [i] was found and confirmed by many scholars [5,6,7] as an original and stable feature of the Jin dialect, while the EME (emerging) type was found to be a developing and unstable feature in the Jin dialect [8,9,10]. Comparing the pronunciation of these two types of characters with the medial [i], the phonetic contexts with regard to the initial consonants and the nucleus vowels show different features: (1) ORI characters have bilabial initial consonants and two nucleus vowels ([ə], [a]); (2) EME characters have velar initial consonants and another two nucleus vowels ([ɛ], [u]).

The nucleus vowels were investigated because past studies [8,10] suggested that the raising and fronting of the nucleus vowels may facilitate the emergence of the medial [i] because the articulation place of a higher and fronter nucleus vowel is closer to that of the high front vowel [i]. With increasing language contact with Mandarin Chinese, the nucleus vowels after the medial [i] were predicted to be lowered or retracted which could accelerate the disappearance of the medial [i]. In the experiment, two types of materials with the medial [i], the ORI type and the EME type, in four nucleus vowel contexts ([ə], [a], [ɛ], [u]) were elicited and recorded in three different speech styles proposed by Labov [11] (informal interview, careful picture description, and formal wordlist reading with speakers' increasing awareness of the language used in the tasks). The formant frequencies of the nucleus vowels after the medial [i] were examined to explore the potential vowel

shift phenomenon in relation to the medial [i]. In addition, a questionnaire with 20 attitudinal questions was designed to investigate the participants' language attitudes towards the Jin dialect spoken in *Baotou* on a 6-point scale. The higher the score, the more positive the attitudes they had.

The results showed that the medial [i] occurred less frequently in the ORI type than in the EME type. As for the speech styles, the usage of the medial [i] decreased with the increasing formality of the style, i.e. decreasing from informal interview, to careful picture description, to formal wordlist reading. Particularly, this change along the language style was more significant regarding the ORI types than the EME types, which suggested that the usage of the ORI type of medial [i] might be under the speaker's control to a large extent. However, the finding of language attitude suggested that it did not act as a good predicator for the production rate of the medial [i], which was against our hypothesis.

The usage of the medial [i] decreased among the younger generations who were also leading in the retraction of the vowel [u] with relatively lower normalized F2 frequencies. The age differences found in the current study are typical evidence of the dialect levelling phenomenon, as the younger generations have much more contact with Mandarin Chinese than the older speakers do. As for the gender differences, although the statistical analysis showed no significant difference regarding the usage of the medial [i], the female speakers showed different trends of vowel shift by leading in the lowering of all the four nucleus vowels with relatively higher normalized F1 frequencies. It is possible that the medial [i] may disappear more quickly in female speakers' speech with all the nucleus vowels showing a lowering trend, which is consistent with the previous finding [12] that females tend to use more prestige or modernized forms (without the medial [i]) than males.

Keywords: dialect levelling; medial [i]; vowel shift; the Jin dialect; language styles; socio-phonetic variation.

References

[1] Xiong, Z. H., Zhang, Z. X. & Huang, H. 2012. *Language Atlas of China (2nd edition)*. Beijing: The Commercial Press.

[2] Li, R. 1985. The Distribution of Mandarin Chinese, *Dialect*, 1.

[3] Tang, C., & Heuven, V. J. van. 2008. Mutual intelligibility of Chinese dialects tested functionally. *Linguistics in the Netherlands 2008*, (25), 145-156

[4] Williams, Ann, & Paul Kerswill. (1999). Dialect leveling: Change and continuity in Milton Keynes, Reading and Hull. In P. Foulkes & G. Docherty (eds.), *Urban voices: Accent studies in the British Isles*. London: Arnold. 141–162.

[5] Hou, J. 1999. *Study of Modern Jin Dialect*. Beijing: Commercial Press.

[6] Hou, J. 2002. *Introduction to Modern Chinese Dialects*. Shanghai Education Press.

[7] Qiao, Q. 2003. The Non-Synchronical Development of Jin dialect and Mandarin (II). *Dialects, (3)*, 233-242.

[8] Bai, J. 2009. Colloquial Readings of Level 1 of MC Xian and Shan Final Groups and Vowel Raising of Lüliang Dialects in Shanxi Province. *Dialects. (1)*, 34-39.

[9] Shi, Y. 2013. A Survey of the Medial [j] in the First Division Characters of Opening-mouth Rhyme in Chinese. *Journal of Central South University (Social Science), 19*(3).

[10] Zheng, Z. 2002. The cause of the abnormal medial in Chinese dialects and the phonetic changes of [e] > [ia], [o] > [ua], *Essays on Linguistics, (26)*, Beijing Commercial Press.

[11] Labov, W. 1963. The social motivation of a sound change. *Word, 19*(3), 273–309.

[12] Queen, R. 2013. Gender, sex, sexuality, and sexual identities. In J.K. Chambers & Natalie Schilling (Eds.), *The handbook of language variation and change* (pp. 368–87). Oxford: John Wiley & Sons.

Exploring the impact of phonological restrictions on phonetic implementation patterns using singing voice: The case of kobushi singing in Japanese

Rina Furusawa[1], Shigeto Kawahara[2]

[1]International Christian University, [2]The Keio Institute of Cultural and Linguistic
furusawar72@gmail.com, kawahara@icl.keio.ac.jp

Summary: How phonological patterns may be shaped by phonetic considerations and how phonological restrictions may affect phonetic implementation patterns have continued to be exciting areas of research throughout the history of phonetic and phonological investigations (e.g., [1]). The current work explored the influence of phonological restrictions on phonetic patterns from a novel perspective through an analysis of singing voice known as *kobushi* in Japanese, which involves an abrupt rise-fall of the F0 contour. A previous study has shown that one type of kobushi usually appears with a neighboring voiceless obstruent, whereas another type of kobushi rarely appears with an adjacent voiceless obstruent. The current study has found that these phonological restrictions affect the phonetic implementation patterns of these two types of kobushi.

Background: Kobushi is a singing technique that is found in several traditional Japanese singing styles. While there are various kinds of kobushi, the kinds that we focused on involve an abrupt F0 rise-fall of about 70 Hz that is implemented within as fast as 30 ms. Minami Kizuki is a professional singer who applies this singing technique to pop song music. [2] has shown that Ms. Kizuki uses two types of kobushi, shown in Figure 1: one type appears near a VC-transition ("right-aligned kobushi") and another type appears near a CV-transition ("left-aligned kobushi").

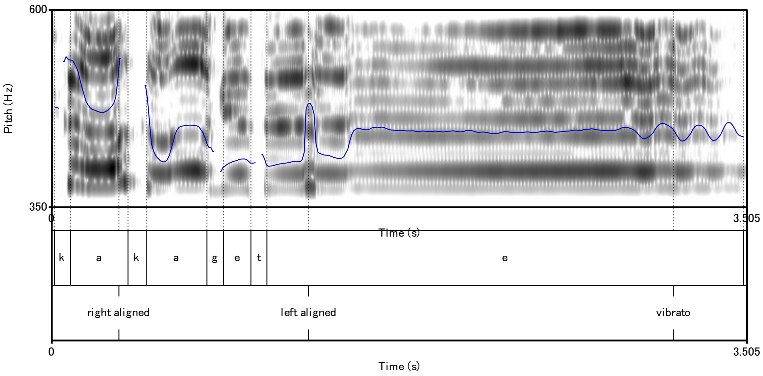


Figure 1: Two types of kobushi; (1) right-aligned kobushi which appears at the end of the first [a] and (2) left-aligned kobushi which appears at the beginning of the long [e].

[2] has also found phonological restrictions on these types of kobushi. The right-aligned kobushi usually appears with a following voiceless obstruent; it looks as if Ms. Kizuki is exaggerating the F0 perturbation effect due to a voiceless obstruent (e.g., [3]) to create this type of kobushi. On the other hand, left-aligned kobushi rarely appears when the preceding consonant is an obstruent, either voiced or voiceless. This phonological restriction may have its roots in the fact that the F0 is perturbed by obstruents, and therefore, the vowel following an obstruent may not be the optimal interval to place left-aligned kobushi; i.e., kobushi is "licensed", in the sense of [4], in a syllable with a sonorant onset. In this presentation, we report on an experiment which tested whether these phonological restrictions impact the phonetic implementation of kobushi.

Methods: Ms. Kizuki sung two songs once with [ta]-syllables and again with [ma]-syllables, as in the research using reiterant speech (e.g., [5]). Based on these sounds, we measured (i) the rise magnitude of kobushi in each condition, (ii) the speed of kobushi implementation (the rise magnitude divided by rise time), and (iii) the distance between the consonantal offset and left-aligned kobushi onset. We expected that left-aligned kobushi may be less easily implemented in the [ta]-rendition than in the [ma]-condition, but the opposite pattern would hold for right-aligned kobushi. For the first two measures, we assessed the results using Bayesian regression models with the [t]-[m] difference and the kobushi-type difference as sum-coded independent variables. We expected that the interaction term between these factors would be meaningful. For the last measure, we analyzed the duration between the consonant offset and the left-aligned kobushi onset and compared the difference between the [t]-condition and the [m]-condition.

Results: *(i) The rise magnitude*. As shown in Figure 2, we found that left-aligned kobushi, which disfavors an obstruent onset, shows smaller rise after [t] than after [m]. The right-aligned kobushi, which takes advantage of a following obstruent, shows larger rises after [t] than after [m]. The Bayesian regression model shows that the central estimate of the crucial interaction term is $\beta_1 = -5.13$, with its 95% credible interval being [-16.82, 2.08]. The probability of this coefficient being negative given the posterior distribution is .87.

(ii) The rise speed. Figure 3 shows the pattern of the speed in which kobushi is implemented in semi-tones per second [6]. As expected, for left-aligned kobushi, which is marked with [t], speed is faster in the [m]-condition than in the [t]-condition, whereas the opposite pattern holds for the right-aligned kobushi. The Bayesian regression shows the estimate of the interaction term is $\beta_1 = -0.37$ with its 95% credible interval being [-0.49, -0.25]. The probability of this coefficient being negative is 1.

(iii) Distance. The distance between the consonantal offset and the left-aligned kobushi's onset showed a result in the expected direction. We found tokens in the [m]-condition in which the kobushi rise starts almost at the same time as [m]'s offset, whereas there were tokens in which left-aligned kobushi was "repelled" after [t]. The overall results appear in Figure 4. The duration is indeed longer after [t] than after [m] ($\beta_1 = 1.89$), but its 95% credible interval was rather large [-9.37, 23.31], and $p(\beta_1>1)=0.58$. While being in the expected direction, the evidence based on this data is a modest one.

Conclusion: The phonological generalizations that [2] has identified are that left-aligned kobushi is phonologically marked with a voiceless obstruent, whereas right-aligned kobushi is unmarked with a voiceless obstruent. The current quantitative study has shown that these phonological restrictions impact the phonetic implementation patterns of singing patterns, as sung by a professional singer, which was most evident in terms of the speed in which kobushi is implemented. Admittedly, our data is limited (singing voice of two songs sung by one singer), and we hope to collect more data to examine how robust the current findings are. Nevertheless, already as it stands, the present study opens up a new area of research in which we can explore the interaction between phonetics and phonology through the analysis of singing voice.

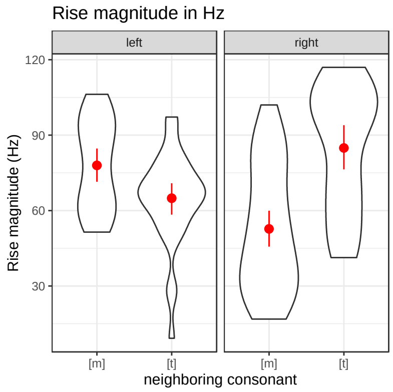


Figure 2: Rise magnitude. Red circles represent the means; the error bars represent bootstrap 95% CIs.

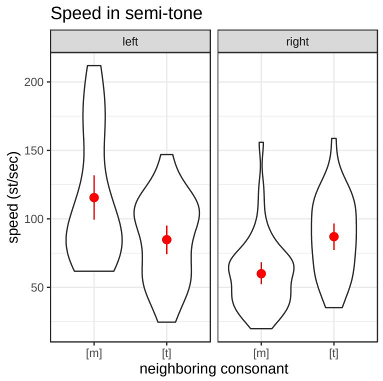


Figure 3: Speed of kobushi.

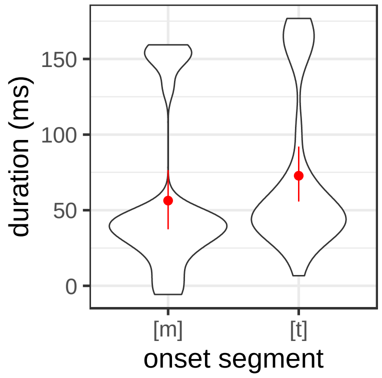


Figure 4: Distance between the consonant and kobushi.

References: [1] Hayes, B. et al. (2004) Phonetically based phonology. CUP. [2] 川原繁人・古澤里菜 (2023) 城南海の「こぶし」の音声学的特徴と音譜上の分布について. 慶應義塾大学言語文化研究所紀要 54. [3] Kingston, J. & R. Diehl (1994) Phonetic knowledge. *Lg*. [4] Steriade, D. (1997) Phonetics in phonology. Ms. [5] Kelso, J.A. et al. (1985) A qualitative dynamic analysis of reiterant speech production. *JASA*. [6] Xu, Y. & Sun, X. (2004) Maximum speed of pitch change and how it may relate to speech. *JASA*.

Nasality and Nasal Excrescence of the Nasal Vowels in Shanghai Chinese

Changhe Chen, Jonathan Havenhill

Department of Linguistics, The University of Hong Kong
riverch8@connect.hku.hk, jhavenhill@hku.hk

The nasal vowels of Shanghai Chinese (Shanghainese) have been described using a range of distinct transcriptions and analyses. These differences reflect both present-day interspeaker variations, as well as sound change throughout the past century. Three general classes of transcription can be found, including (i) phonemic nasal vowels which occur in the absence of nasal coda consonants, transcribed as [Ã][1] or [ã õ̞] [1-3], (ii) nasal vowels followed by a weak [ɲ] or [ŋ], transcribed as [Ã^{ɲ}] or [ã^{ŋ}] [4, 5], which may show only partial nasalization [4], and (iii) contextually nasalized vowels followed by [ŋ] [6, 7]. There is insufficient acoustic and articulatory data to decide between these conflicting accounts, however. This study examines the acoustic and articulatory configuration of the Shanghainese nasal vowels, with the aim of understanding the phonetic features of nasal vowels in general.

The presence of coda [ŋ] following Shanghainese nasal vowels [6, 7], as well as partial vowel nasalization [4], raises the possibility that [ã$^{(ŋ)}$ ɒ̃$^{(ŋ)}$] are in fact contextually nasalized, and/or that the historical nasal coda was never fully dropped during VN → Ṽ sound change. Beddor [8] observes for English that the duration of vowel nasalization is negatively correlated with coda nasal duration. She attributes this finding to variability in the timing of velar lowering, while the velum gesture itself exhibits a constant duration. A similar pattern may be predicted for Shanghainese if vowel nasality is the result of contextual nasalization. On the other hand, incomplete vowel nasalization and coda nasal consonants may also arise as the result of nasal excrescence, in which coda nasal consonants are introduced or restored in a Ṽ → VN sound change. Cross-linguistic data show that excrescent nasals are typically velar, which are favored on perceptual grounds [9, 10]. In Portuguese, nasal vowels with excrescent nasal consonants show varying degrees of nasalization (e.g., [11]), although the proportion of vowel nasalization has not been found to correlate with nasal consonant duration [12]. This study therefore seeks to determine (i) whether Shanghainese nasal vowels show full or incomplete nasalization, (ii) whether they are realized with or without a following nasal consonant, (iii) whether the duration of vowel nasality correlates with that of a following nasal consonant, and (iv) the place of articulation of the following nasal consonant.

Data from 6 native speakers of Shanghainese (3 men, 3 women) were analyzed. All speakers were born and raised in urban Shanghai through age 18 and predominantly speak Shanghainese with their families. Speakers were asked to recite a word list containing 17 (near) minimally contrastive sets containing [a ɐ ã$^{(ŋ)}$ ɒ̃$^{(ŋ)}$] in (C)V and (C)Vʔ syllables, as well as 57 fillers. Each set was composed of four words, e.g., [a˩˧] 鞋 "shoes", [ɐʔ˦] 匣 "box", [ã$^{(ŋ)}$˩˧] 杏 "apricot", [ɒ̃$^{(ŋ)}$˩˧] 項 "item". Words were uniquely pseudo-randomized, and each word was repeated four times in succession. The experimental setup resembled that of Carignan [13]. Simultaneous ultrasound (SonoSpeech Micro in AAA at 84 fps), nasalance (Glottal Enterprises NAS Separator Handle), electroglottographic (Voce Vista), lip video (at 60 fps), and acoustic data were recorded.

Nasalance was measured at 20% intervals throughout the vowel duration and was calculated as the ratio (A_N / (A_O + A_N)) of the RMS amplitude of the oral (A_O) and nasal (A_N) channels; a higher value indicates greater nasalization. Mean nasalance during each speaker's production of oral [a] served as baseline for determining the onset of nasalization. Nasalization was considered to begin at the point when nasalance exceeded the baseline. This point was used to calculate the duration of nasalization throughout the vowel and nasal coda. To examine the place of articulation of the nasal consonant, ultrasound tongue contours were extracted at a single point during the nasal consonant. Reference tongue contours extracted during the stops [t k] (which preceded a low vowel) were analyzed and compared using polar SSANOVA.

[1] In Chinese studies, the central low vowel [ä] is usually transcribed as [A].

Data for the two Shanghainese nasal vowels [ã$^{(ŋ)}$ ɒ̃$^{(ŋ)}$] are presented separately, although results indicate that the two vowels are merged for all speakers. Observed in 96% of tokens, the nasal consonants are velar or post-velar nasal approximants. Representative results for the nasal consonant from a single speaker are presented in Figure 1. Nasalance data presented in Figure 2 reveal that both nasal vowels (which overlap) show little nasalization in the first fifth of their duration and steadily increasing nasality thereafter. This pattern differs from French, in which nasal vowels show a high degree of nasalance as early as the vowel onset [13]. Although this finding indicates that the Shanghainese nasal vowels are only partially nasalized, there is no clear relationship between proportion of vowel nasality and duration or presence of the nasal consonant. A linear mixed effects regression model with fixed effects of nasal consonant duration, vowel, and tone, as well as random effects of speaker and word, was constructed. ANOVA comparison of two models with and without the fixed effect of consonant duration shows no significant difference in model fit ($\chi^2(1) = 0.09, p = 0.77$). Individual speaker data (Figure 4) indicate that while the duration of nasalization varies, duration of the nasal consonant remains relatively stable, suggesting variable duration of the velar lowering gesture. This pattern more closely resembles the nasal vowels of Portuguese [12] than the nasalized vowels of English [8]. While it remains to be determined to what extent the English and Portuguese patterns generalize to other languages, it is possible that the velar/postvelar nasal consonant in Shanghainese, which has weak consonantality [14], was appended as a result of misperception between nasalized vowels and back nasal consonants [9].

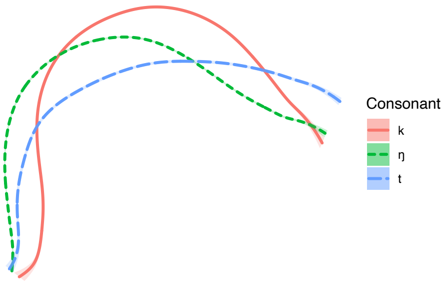


Fig.1 Polar SSANOVA tongue contours of nasal consonant [ŋ] comparing with stops [t k]; the tongue root is to the left.

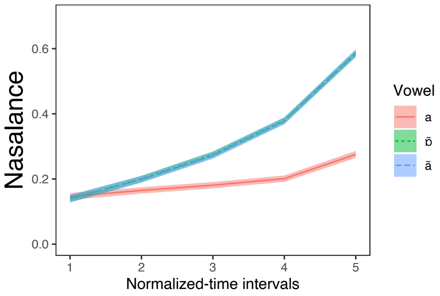


Fig.2 Nasalance of 6 speakers; the two nasal vowels are compared with the oral vowel [a].

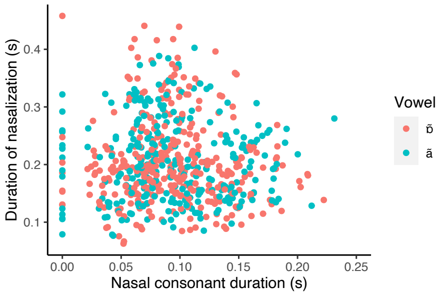


Fig.3 Scatter plot of nasalization duration and nasal consonant duration (zero duration indicates no nasal consonant).

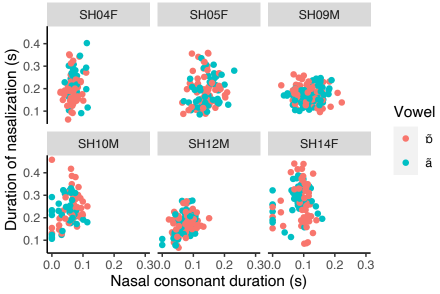


Fig.4 Duration data of individual speakers.

References:

[1] Chao, Y. R. (1928). 現代吳語的研究 [*Studies of the modern Wu dialects*]. Beijing: Tsing Hua College Research Institute. [2] Xu, B. & Tang, Z. (1988). 上海市區方言志 [*A description of urban Shanghainese*]. Shanghai: Shanghai Education Press. [3] Chen, Z. (2019). 開埠以來上海城市方言語音演變 [The historical developments of sound system since the beginning of Concession in the Shanghai urban dialect]. *Bulletin of Linguistic Studies*, *24*, 280-313. [4] Qian, N. (1992). 當代吳語研究 [*Studies of the contemporary Wu dialects*]. Shanghai: Shanghai Education Publishing House. [5] Zee, E. (2016). Shanghai phonology. In G. Thurgood & R. J. LaPolla (eds.), *The Sino-Tibetan Languages* (Second Edition) (pp. 185-192). Routledge. [6] Karlgren, B. (1915–1926). *Etudes sur la phonologie chinoise*. Leyde: E.-J. Brill. [7] Chen, Y., & Gussenhoven, C. (2015). Shanghai Chinese. *Journal of the International Phonetic Association*, *45*(3), 321-337. [8] Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821. [9] Johnson, K., DiCanio, C. T., & MacKenzie, L. (2007). The Acoustic and Visual Phonetic Basis of Place of Articulation in Excrescent Nasals. *UC Berkeley Phonology Lab Annual Reports*, *3*. [10] Chang, K.-Y. (2012). Nasalization of Nasal Finals in Chinese, *Studies in Language and Linguistics*, *32*(2),17-28. [11] Sousa, E. M. (1994). *Para a caracterização fonético-acústica da nasalidade no português do Brasil* [*Towards the phonetic-acoustic characterization of nasality in Brazilian Portuguese*]. (Unpublished doctoral dissertation) Campinas: Universidade Estadual de Campinas. [12] Rothe-Neves, R. (2021). Durational aspects of tautosyllabic vowel nasalization in (Brazilian) Portuguese: An airflow investigation. *Journal of Portuguese Linguistics*, *20*(1), 1-25. [13] Carignan, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *Journal of Phonetics*, *63*, 87-105. [14] Ohala, J. J. & Ohala, M. (1993). The phonetics of nasal phonology: theorems and data, In M.K. Huffman & R.A. Krakow (eds.), *Nasals, nasalization, and the velum* (pp. 225-249). San Diego, CA: Academic Press.

The effect of second-language learning experience on Korean listeners' use of pitch cues in the perception of Cantonese tones

Zhen Qin[1] & Sang-Im Lee-Kim[2]

[1]Hong Kong University of Science and Technology (Hong Kong), [2]Hanyang University (Korea)
hmzqin@ust.hk, sangimleekim@hanyang.ac.kr

Past studies have found the linguistic experience of previously acquired languages, for instance, one's native-language (L1) and second-language (L2) learning experience, modulates the perception of novel sounds from an unfamiliar language (i.e., a third language, L3). It remains unclear whether L1 or L2, or both, is the source of transfer in the very beginning stage of L3 acquisition [1,2]. Lexical tone is a good case for testing the influence of L1 or L2, as listeners with different language backgrounds have a different cue-weighting pattern in tone perception [3]. While tone language (i.e., Chinese) listeners rely more on *pitch contour* (different tone shapes; rising vs. falling tones), non-tone language (e.g., English) listeners often use *pitch height* (difference in height; high vs. low tones). To test the influence of L1 or L2 on the perception of L3 tones, Qin and Jongman [4] examined how English-speaking L2 learners of Mandarin employed pitch contour and pitch height in their perception of Cantonese tones. The results showed that while Mandarin listeners used pitch contour more than pitch height, English listeners who were naïve to lexical tones did not show a difference in their use of pitch cues. Crucially, the L2 learners showed a pattern like Mandarin listeners. The finding suggests an influence of the L2 instead of the L1 in the perception of L3 tones. However, another interpretation of the results would be that the functional use of pitch to lexical contrasts is quite limited in English and the L1 influence was thus not borne out clearly.

The present study is motivated to disambiguate the hypotheses by testing speakers of a language that fully employs pitch cues for lexical contrasts. To that end, we focus on Korean-speaking L2 learners of Mandarin whose L1 (variety) is either Seoul Korean (SK) or Kyungsang Korean (KK). SK is neither tonal nor stressed and does not use pitch to mark lexical prosody [5]. In contrast, KK uses pitch differences to realize lexically contrastive words (e.g. [ka^{L}.tɕi^{H}] 'eggplant' vs. [ka^{H}.tɕi^{H}] 'branch') [6]. If the influence of the L1 is predominant, the two groups of Korean-speaking learners are expected to show different performances, with KK-speaking-L2 learners patterning more like Mandarin listeners by virtue of the contrastive pitch cues in their L1 variety [7]. If the L2 learning experience is more integral in the way L3 prosody is processed, both groups are expected to show greater sensitivity to pitch contour than to pitch height.

The participants completed an AX forced-choice tone discrimination task. 20 intermediate-to-advanced SK-speaking and 15 KK-speaking L2 learners of Mandarin, who were matched in their proficiency in Mandarin and music experience, were recruited as target groups. 15 SK-speaking and 15 KK-speaking (also with limited exposure to SK) participants, who were naïve to any tone languages, were recruited as control groups. As illustrated in Figure 1, four Cantonese tones, that is, one contour tone (Tone 2; T2-rising) and three level tones (T1-high; T3-mid; T6-low), were used for the perception task. Level-Contour (T1-T2; T6-T2) and Level-Level (T1-T6; T3-T6) tonal contrasts were target tone pairs, allowing for testing the *primary* use of pitch contour versus pitch height, respectively [4]. The stimuli were produced by a female native speaker of Cantonese.

Mixed-effects regression models were run on response accuracy (1 for correct and 0 for incorrect). The models were fitted in R using the lme4 package with predictors (cues, groups, and L1 variety) deviation coded (−0.5, 0.5) to test the main effects. The model results, illustrated in Figure 2, showed that naïve Korean listeners, regardless of their L1 varieties, had a greater sensitivity to pitch height than to pitch contour ($\beta = 0.23$, $SE = 0.09$, $z = 2.55$, $p = .01$). In contrast, L2 learners, independent of their L1 varieties, showed greater sensitivity to pitch contour than to pitch height ($\beta = -0.51$, $SE = 0.08$, $z = -6.19$, $p < .001$), consistent with the pattern of Mandarin listeners [4].

Aligned with the cue-weighting theory of speech perception [3, 7], the findings provide evidence for a developmental change in which Korean-speaking L2 learners had *a perceptual cue shift* from pitch height to pitch contour through their L2 experience in Mandarin. Since there is no level tone contrast in Mandarin, subtle differences in pitch height might become within-categorical differences for L2 learners (and Mandarin listeners), resulting in reduced sensitivity to Cantonese level tones [4]. In contrast, the prosodic system of L1 varieties appears to have little influence on L2 learners in their perception of novel tones, which can be potentially explained by the L3 acquisition theory. For instance, the L2 Status Factor Model [1] predicts that L2 plays a privileged role in language transfer due to its non-native cognitive status analogous to L3. The Typological Primacy Model [2], on the other hand, proposes that the source language (L1 or L2) of transfer is determined by the typological similarity between languages. When applied to the current case, L2, rather than L1, is likely to influence the perception of L3 tones either because Mandarin is an L2 or because Mandarin is more typologically similar to Cantonese in that both languages have tone-bearing units as syllables while KK does not [5, 6]. Future studies need to tease apart the two accounts by including other language pairings (e.g., L1 and L3 are both tonal languages).

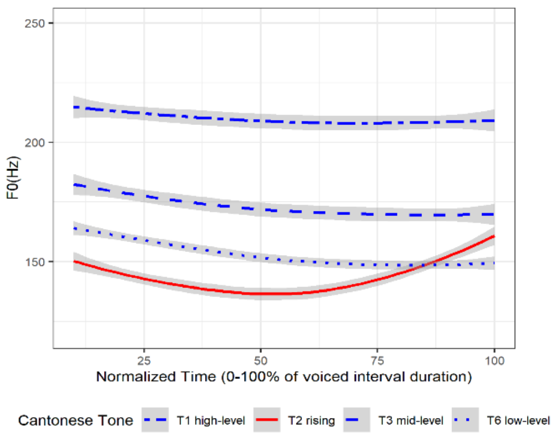


Fig. 1 Time-normalized pitch tracks of a contour tone (T2) in red and three level tones (T1, T3, T6) in blue

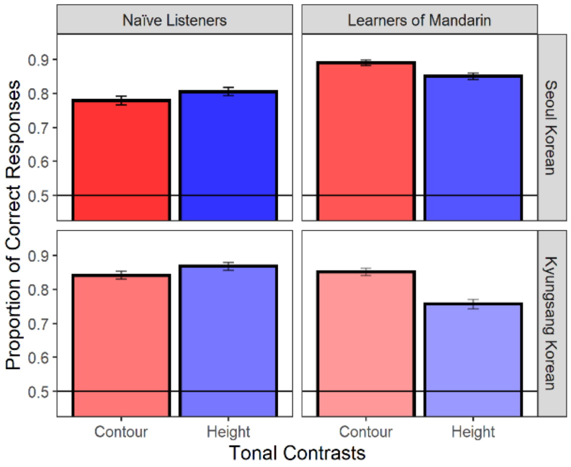


Fig. 2 Discrimination accuracy of Cantonese tones contrasting in pitch contour (red) and pitch height (blue) by SK-speaking (top) and KK-speaking (bottom) naïve listeners (left) and L2 learners of Mandarin (right); the error bars represent 1 SE above/below the mean; the horizontal line represents chance performance (0.5).

References

[1] Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research, 27,* 59-82.

[2] Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research, 27*, 107-27.

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

[4] Qin, Z., & Jongman, A. (2016). Does Second Language Experience Modulate Perception of Tones in a Third Language? *Language and Speech*, *59*, 318-338.

[5] Jun, S. A. (2010). Korean Intonational Phonology and Prosodic Transcription. In Jun, S. A. (eds.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. (pp. 201-229). Oxford, UK: Oxford University Press.

[6] Ramsey, S. R. (1975). *Accent and morphology in Korean dialects: A descriptive and historical study*. Ph.D. Dissertation, Yale University.

[7] Kim, H., & Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics, 87*, 1-15.

Individual Differences in Phonological Proficiency and Correlation with Pitch Sensitivity: Three Types of Perceivers

Renata Kochančikaitė[1] & Mikael Roll[1]

[1] *Lund University (Sweden)*
renata.kochancikaite@ling.lu.se, mikael.roll@ling.lu.se

Native listeners' phonological proficiency has traditionally been analysed as a uniform trait in a healthy population, but there is a growing interest in the estimation of individual differences in phonological processing [1, 2]. Neurophysiological responses recorded during perception of native phonemic contrasts have so far been interpreted to indicate two types of listeners – gradient vs. discrete perceivers, where gradient perceivers rely more strongly on acoustic information felicitously delivered to the cortex, while discrete perceivers incorporate top-down category representation information at that cortical level [3]. However, recent findings in neuroanatomy motivate a hypothesis that the neural basis of phonological proficiency can include *several* processing mechanisms. Each of these mechanisms is a potential source for individual variation: For example, differences in cortical thickness and surface area in language-related brain structures have been associated with not only the natural variance in listeners' phonological proficiency when using word accent tones to interpret morphology [2, 4] but also with the listeners' extra-linguistic skills, such as pitch discrimination ability [5].

To investigate whether generic psychoacoustic abilities play a role in the different perception styles in phoneme context, we tested native Swedish listeners' pitch discrimination and vowel categorisation and discrimination ability. The aim was to gauge the heterogeneity of phonological proficiency and measure it in three aspects: i) *phonemic aptitude*, defined as how categorically the individual listeners perceive inter-category vowel sounds in a linguistic context, ii) *phonetic aptitude*, as in how accurately they judge differences between those inter-category vowel sounds as a function of acoustic distance, and iii) *acoustic aptitude*, measured as their pitch discrimination ability. We then assessed the relation between the performance at the different levels. Our study consisted of three experiments (Table 1).

Sixty native Swedish speakers (age: 19-40, mean: 27.13 years, 42 males) without any known hearing difficulties, language deficits, or neuropsychiatric diagnoses completed the experiments. All participants were recruited via the online service Prolific and were paid for their participation.

Single-subject analysis was done in base R. One outcome measure per participant was obtained from each experiment. Phonemic aptitude performance scores ranged from 5.2 to 187.6 (M = 64.8, SD = 52.2), phonetic – from 0.03 to 1.5 (M = 0.88, SD = 0.38), acoustic – from 1.4 Hz to 20.8 Hz (M = 5.86, SD = 4.11), indicating individual differences among the perceivers on all three performance levels. Group level analysis compared measurement outcomes from the phonemic, phonetic, and acoustic experiments. Since all three variables are random and contain error, the *lmodel2* library in R was used to fit model II ranged major axis (RMA) regression to compare the performance between the three levels. Phonemic aptitude was not predicted by phonetic aptitude (R^2 = 0.03, P_{perm} = 0.11), suggesting that these two aptitudes are separable aspects of phonological proficiency. Acoustic aptitude did not predict phonemic aptitude (R^2 = 0.01, P_{perm} = 0.19) but it did predict phonetic aptitude (R^2 = 0.12, P_{perm} = 0.01), showing that low-level frequency sensitivity may be involved in phonetic judgements.

To inspect the heterogeneity of phonological proficiency, a k-means cluster analysis was conducted using *cluster* and *factoextra* libraries in R. Based on their scaled phonemic and phonetic aptitude scores, the perceivers were grouped into clusters with maximum similarity within and maximum dissimilarity between them. Following the "elbow point" method, the best number of clusters was determined to be three, which explained 69% of the total variance. Clustering results showed the following types of perceivers (Figure 1): those that performed a) above average in both phonemic and phonetic aptitude tests, b) below average in both tests, and c) above average in the phonetic aptitude test but below average in the phonemic aptitude test. Thus, phonological proficiency appears to be more complex than a 1-dimensional axis between gradient and discrete perceivers. Interestingly, no participants exhibited the pattern opposite to c), which suggests that superior phonetic aptitude is a prerequisite for, or a side-effect of, developing and/or maintaining superior phonemic aptitude.

Table 1. Setup of the three experiments that measure phonemic, phonetic, and acoustic aptitude.

	Phonemic aptitude	**Phonetic aptitude**	**Acoustic aptitude**
Task	2-alternative forced choice, categorisation	2-alternative forced choice, discrimination	Transformed 1 up/2 down staircase [6]
Stimuli	Minimal CVC word pairs; V replaced with synthetic inter-category vowels	Synthetic inter-category vowels; acoustic distance (ΔF1+ΔF2)/2 varied from 0 to 0.375 Bark	Pure sine tones
Scope	6 inter-category continua (*kok—kåk*, *tår—tar*, *tar—tär*, *häl—hel*, *fyr—fur*, *tur—tör*)	6 inter-category continua (/u:/—/o:/, /o:/—/ɑ:/, /ɑ:/—/ɛ:/, /ɛ:/—/e:/, /y:/—/ʉ̟:/, /ʉ̟:/—/ø:/)	1/240th to 2 semitones (519-582 Hz, in steps of 0.25 Hz)
Trials	237 (3 blocks of 79)	432 (3 blocks of 144)	4 (2 runs in 2 directions)
Outcome measure	Slope of logistic regression (vowel ~ likelihood of choice), average of all 6 continua	Slope of linear regression (acoustic distance ~ % correct), average of all 6 continua	Average Just-Noticeable-Difference threshold (JND) at 70 % correct, Hz

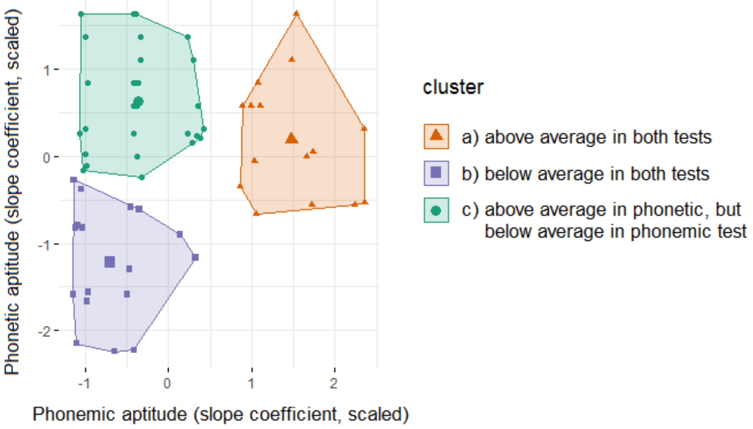


Fig.1 Results of clustering analysis. Perceivers were grouped by phonemic and phonetic aptitude scores.

References

[1] A. C. L. Yu and G. Zellou, “Individual Differences in Language Processing: Phonology,” *Annu. Rev. Linguist.*, vol. 5, pp. 131–150, 2019.

[2] A. Schremm, M. Novén, M. Horne, P. Söderström, D. van Westen, and M. Roll, “Cortical thickness of planum temporale and pars opercularis in native language tone processing,” *Brain Lang.*, vol. 176, pp. 42–47, 2018.

[3] J. Ou and A. C. L. Yu, “Neural correlates of individual differences in speech categorisation: evidence from subcortical, cortical, and behavioural measures,” *Lang. Cogn. Neurosci.*, vol. 37, no. 3, pp. 269–284, 2022.

[4] M. Novén, A. Schremm, M. Horne, and M. Roll, “Cortical thickness and surface area of left anterior temporal areas affects processing of phonological cues to morphosyntax,” *Brain Res.*, vol. 1750, no. December 2019, p. 147150, 2021.

[5] M. Novén, A. Schremm, M. Nilsson, M. Horne, and M. Roll, “Cortical thickness of Broca’s area and right homologue is related to grammar learning aptitude and pitch discrimination proficiency,” *Brain Lang.*, vol. 188, no. 188, pp. 42–47, 2019.

[6] H. Levitt, “Transformed Up-Down Methods in Psychoacoustics,” *J. Acoust. Soc. Am.*, vol. 49, no. 2B, pp. 467–477, 1971.

Non-native Articulatory Variability in English Phonological Rule Application: Evidence from Korean and Indian Learners

Gwanhi Yun[1] & Jae-Hyun Sung[2]

[1]Daegu University (Korea), [2]Kongju National University (Korea)
ghyun@daegu.ac.kr, jsung@kongju.ac.kr

Spoken words naturally exhibit variability in the acoustic and articulatory dimensions, and so do realizations of language-specific phonological rules [1, 2, 3]. While numerous studies have investigated and documented L1 and L2 speakers' cues to differentiate sound categories in their native and non-native languages, it calls for further investigation how various groups of learners, e.g., whether or not they are EFL (English as a Foreign Language) or ESL (English as a Second Language) learners, acquire and produce phonological rules in their target languages. This study examines how Korean EFL speakers and Indian ESL learners apply three different phonological rules in English (the target language) – palatalization, place assimilation, and word-final coronal deletion – to their production of coronal consonants, and compares articulatory variability produced by two groups using ultrasound imaging. The three research questions that are addressed in this study are: (1) Do EFL and ESL speakers exhibit phonological variation? (2) Do the spectrum of phonological variation differ in their magnitude according to individual rules? (3) Do EFL speakers show different patterns of phonological variation from ESL speakers?

Ultrasound tongue contours from four Korean EFL and eight Indian ESL speakers show that both groups of learners produce a wide range of phonological variation in their target language. Both EFL and ESL speakers produced palatalized and non-palatalized variants in the palatalization context (Figure 1), and speakers from both groups tended to produce non-palatalized variants more often than palatalized ones (EFL: 74% non-palatalized & 26% palatalized; ESL: 80% non-palatalized & 20% palatalized). Both groups' preference for non-palatalized variants in the palatalization context might have been resulted by the way palatalization rules in learners' native languages are applied, in which palatalization is optional in the colloquial register in Tamil [4], and obligatorily applied across morpheme boundaries in Korean [5].

In their production of place assimilation rules, both EFL and ESL speakers yielded three major variants: assimilated, non-assimilated and hyperarticulated variants (Figure 2). Contrary to the gestural patterns in the palatalization context, the assimilated variants were dominant in both groups, with a noticeable inter-group difference (EFL 38% vs. ESL 31%). Patterns of assimilation also exhibited variability across places, in which coronal-to-velar assimilation resulted in more variation than coronal-to-labial assimilation.

Word-final /t/ deletion rules, as shown in their gestural patterns of place assimilation rules, were realized by both groups in three variants: deletion, no deletion, and hyperarticulation (Figure 3). Their production of deletion rules also yielded inter-group variation, in which no deletion was dominant in the EFL group, and deletion in the ESL group. As reported in their production of place assimilation, both EFL and ESL speakers exhibited variation across different places. Hyperarticulation was favoured before a velar stop, e.g., *must cap*, than before a labial stop, e.g., *must pad*.

Articulatory patterns reported in this study confirm that both EFL and ESL speakers produce a substantial amount of phonetic variability as native speakers would do. The way phonological rules are applied is inherently gradient rather than categorical. Both Korean EFL and Indian ESL speakers show that the extent of the spectrum of pronunciation variants differ by individual learners and phonological rules, and the likelihood of each phonological variant varies across phonological contexts.

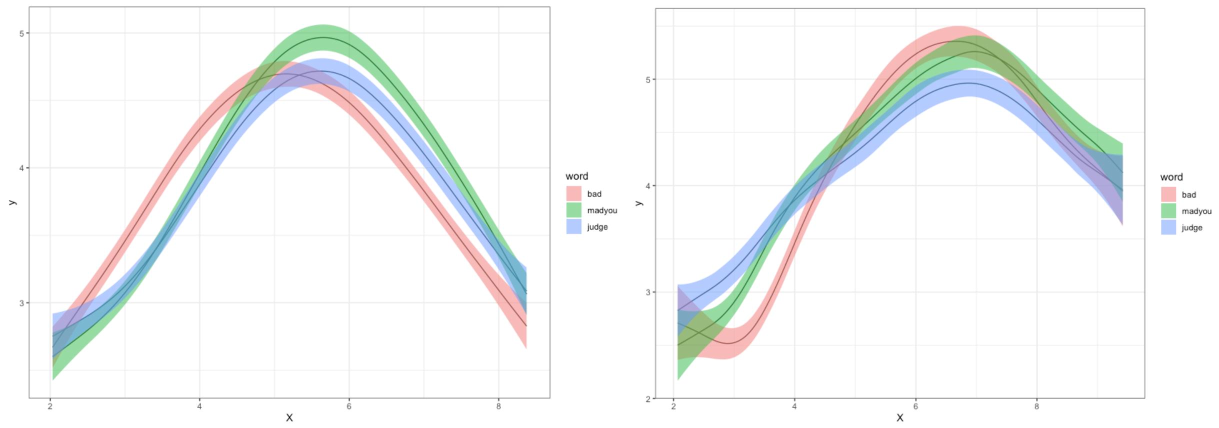


Fig.1 : Palatalization from EFL Speaker #1 (left) and ESL Speaker #1 (right)

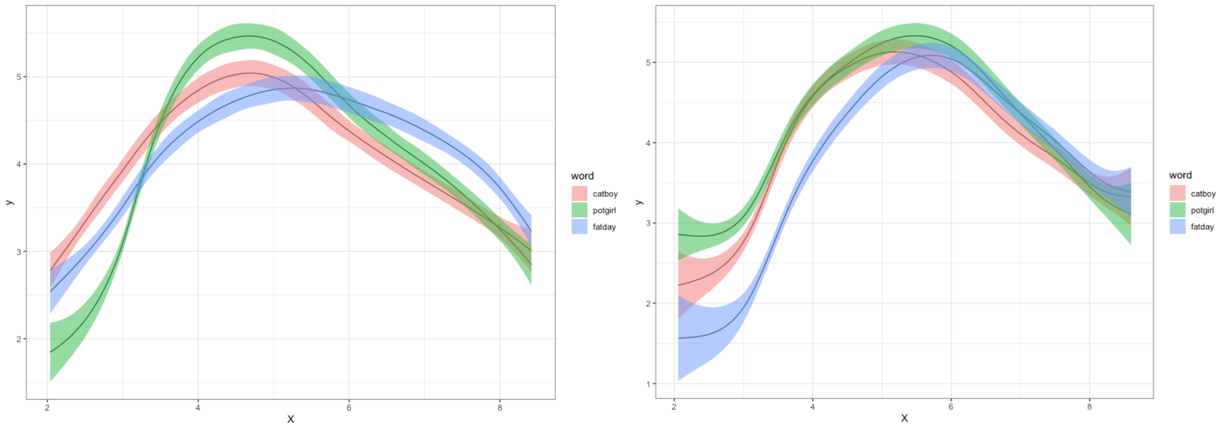


Fig.2 : Place Assimilation from EFL Speaker #2 (left) and ESL Speaker #8 (right)

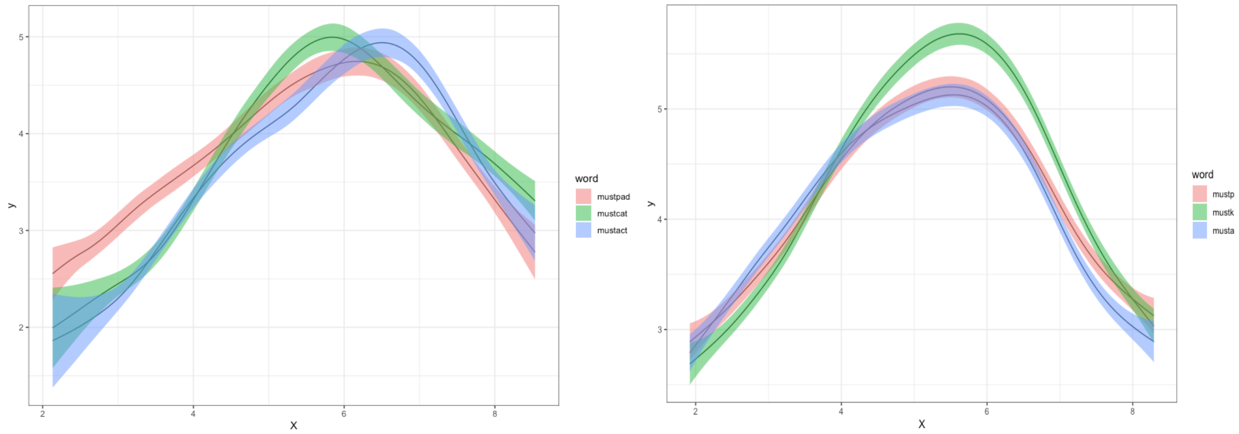


Fig.3 : Word-final /t/ Deletion from EFL Speaker #5 (left) and ESL Speaker #2 (right)

References

[1] Bush, N. (2001). Frequency effects and word-boundary palatalization in English. In J. Bybee & P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 244-280). Amsterdam, Netherlands: John Benjamins.

[2] Braver, A. (2013). *Degrees of incompleteness in neutralization : Paradigm uniformity in a phonetics with weighted constraints*. PhD thesis, Rutgers University.

[3] Yun, G. (2022). A mismatch in completeness between acoustic and perceptual neutralization in English flapping. *Korean Journal of English Language and Linguistics, 22,* 1133-1158.

[4] Schiffman, H. (1999). *A Reference Grammar of Spoken Tamil.* Cambridge, UK: Cambridge University Press.

[5] Sohn, H-M. (2001). *The Korean Language.* Cambridge, UK: Cambridge University Press.

In my humble opinion : The prosodic portrayal of the non-standard 1sg

Sophia BURNETT[1]

[1]*CY Cergy Paris Université (France)*
Sophia.Burnett@cyu.fr

Ongoing research into the re-emergence [1] of the non-standard 1sg in English ("i", herein NS1sg) shows [2] that the inclusion of the novel form in computer mediated communications variant spellings is indeed encoding sociophonetic variation [3]; that it is deployed for its attenuative effects in a language with only one 1sg and no speech levels. The attenuation afforded by the variant is displayed in both a graphemic and—as suggested here—an acoustic reduction; a strong graphemic "I" becomes a weak graphemic "i", and a diphthong /ɑɪ/ shifts to a monophthong /ə/ , /ɑː/ , or /ʌ/. This pairing of phoneme-grapheme [4] attenuation can prove useful on social media, in particular when conversing on potentially sensitive issues. The very sound of /ɑɪ/ is salient and self-assertive, the sound of /ə/ is not. The perception of this phonetic variation relies on endophasia [5][6], but where do these vocal depictions come from, and are they faithful to the grapheme?

This paper aims to demonstrate how vocal characterizations are being inscribed and enregistered in the NS1sg using personae, which are essentialized personifications of the imagined typical user [7]. Enregisterment has been characterized [8] as a social process whereby diverse behavioural signs (whether linguistic, non-linguistic, or both) are functionally re-analysed as cultural models of action, as behaviours capable of indexing stereotypic characteristics of incumbents of particular interactional roles, and of relations among them. As already established [9], some implementations of the NS1sg are visually iconic to the point of being purely indexical], used to signal by a young online demographic. This instant recognition bypasses the slower cognition that mobilizes reading on a deeper level and engages with subvocalization. I posit that some usages of the novel form—in particular those deployed for pragmatic, attenuative effect—do however rely on endophasia, and that the decision to be informed by a particular external voice relies heavily on the characterisation afforded by its tone and intonation [10], over all other idiosyncrasies. This inner voice will have been chosen as appropriate to the NS1sg grapheme from an 'archive' of aural representations, either of the reader's own voice, of a voice encountered during social interaction, or a voice heard via various forms of media—in any case externalized. These aural depictions of the standard and non-standard 1sg personae may come to us via news and information, fiction series, movies, video games, YouTube, Instagram, and advertising, etc.

The corpus was created from 47 main primary sources further divided into multiple tokens; all highly viewed English language material. 131 audio file segments containing standard or non-standard forms of the 1sg were isolated, and the tokens annotated as 'perceived as' 1sg or NS1sg. The high exposure is important for their memetic potential, but also for their impact potential, since the authority of the personae is enhanced by the implication of high numbers of views if not significant material resources. [11]

Segments were isolated first by auditory analysis and spectogram pattern, then their separation was fine-tuned using formant analysis. The F1 and F2 formants of the perceived standard and non-standard forms were measured at their optimal length [12] 1sg nuclei and offglides were measured (Fig.1), and NS1sg monophthongs (Fig.2) were treated in the same manner.

The depictions were then plotted to show saliences of the actor personae representations compared to formants of a standard plot of a 1sg (diphthong) and NS1sg (monophthong). Since target phonemes are representative of the intended utterance even in voiceless versions of diphthong offglides[13], the F2 were isolated in order to map frequencies and see whether the perceived 1sg and NS1sg from the mass media followed suit.

The results (Fig.3) show that of the 131 tokens (F2 Hz frequencies), 93 were perceived as 1sg, and 37 as NS1sg. The full corpus showed a frequency range of 1984Hz from min. 758Hz to 2742Hz max, with a mean 1827.04Hz and median of 1810Hz and SD of 1431Hz. Both data sets 1sg and NS1sg were normally distributed The results clearly show (p-value=<0.05) that the perceived

NS1sg sits in the frequency area of high F2, and that as expected the 1sg personae reside in the low F2. The frequency of the 'less self-asserting' voice correlates with the same personae from highly viewed mass media.

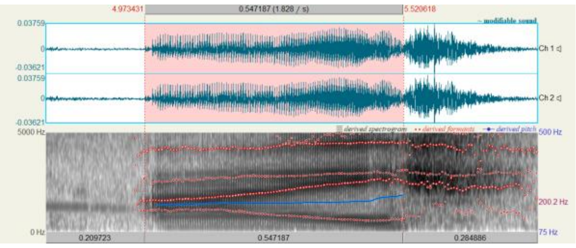


Fig.1 *Movie: Jurassic World, token n°3 (>$1,671,537,444 in box-office revenue). Long 1sg.*

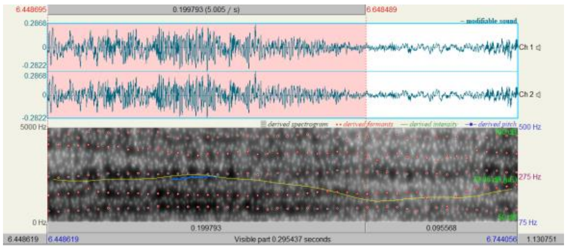


Fig.2 *Video game: COD Modern Warfare 2019, token n°4 (>26,500,000 sales) Breathy NS1sg.*

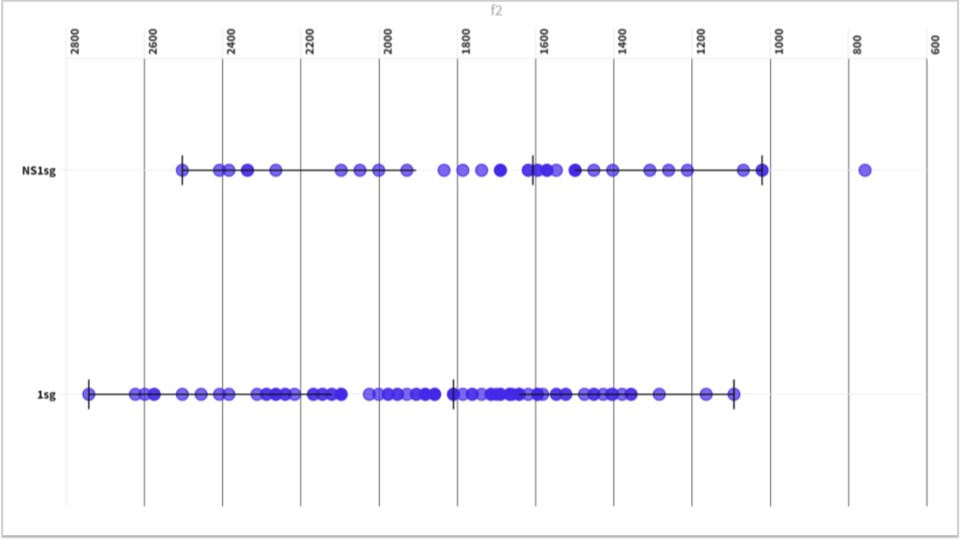


Fig. 3 *F2 formant frequencies in HZ of NS1sg and 1sg.*

References

[1] Burnett, S. (2022). Offsetting love and hate: The prosodic effects of the non-standard 1sg in tweets to Boris Johnson and Jeremy Corbyn over four days of the UK general election. *European Journal of Applied Linguistics*.

[2] Manfred, Markus (2021) Aphesis and Aphaeresis in Late Modern English Dialects (based on EDDOnline), *English Studies*, 102:1, 124-141

[3] Tatman, R. (2015). # go awn: Sociophonetic Variation in Variant Spellings on Twitter. *Working Papers of the Linguistics Circle*, 25(2), 97-108.

[4] Ernst Pulgram (1951) Phoneme and Grapheme: A Parallel, *WORD*, 7:1, 15-20.

[5] Bertolotti, M. (2018). A cognitive neuroscience view of inner language : to predict and to hear, see, feel. In Langland-Hassan, P. & Vicente, A. (eds.), *Inner Speech: nature, functions, and pathology*. Oxford University Press.

[6] Jaffe, A. and Walton, S. (2000), The voices people read: Orthography and the representation of non-standard speech. *Journal of Sociolinguistics*, 4: 561-587.

[7] Ilbury, C. (2020). "Sassy Queens": Stylistic orthographic variation in Twitter and the enregisterment of AAVE. *Journal of sociolinguistics*, 24(2), 245-264.

[8] Agha, A. (2015). Enregisterment and communication in social history. *Registers of Communication*, 18(1), pp.27-53.

[9] Burnett, S. (2022). The non-standard 1sg and the lessened Self. *Proceedings of the 13th International Symposium on Iconicity in Language & Literature*, Sorbonne University, Paris.

[10] Cruttenden, A. (1997). Intonation. Cambridge University Press.

[11] Foucault, Michel. 1981. The order of discourse. In *Untying the Text: A Post-Structuralist Reader*. Edited by Robert J. C. Young. Translated by Ian McLeod. London: Routledge.

[12] Jacewicz, E., Fujimura, O., & Fox, R. A. (2003). Dynamics in diphthong perception. In *Proceedings of the XVth International Congress of Phonetic Sciences*, Barcelona, Spain (pp. 993-996).

[13] Chládková K, Hamann S, Williams D, Hellmuth S. F2 slope as a Perceptual Cue for the Front-Back Contrast in Standard Southern British English. *Lang Speech*. 2017 Sep ;60(3) :377-398.

Effects of focus and lexical tones on preboundary lengthening and its kinematic characteristics in Mandarin Chinese: A preliminary report

Hongmei Li[1,3], Sahyang Kim[2], and Taehong Cho[3]

[1]Yanbian University (China), [2]Hongik University (Korea),
[3]Hanyang Institute for Phonetics & Cognitive Sciences of Language, Hanyang University (Korea)
angela.hongmeili@gmail.com, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Preboundary lengthening (henceforth PBL) is a temporal expansion of domain-final phonological units before a prosodic boundary, which is observed cross-linguistically [1,2]. Within the framework of Articulatory Phonology, PBL is considered to be modulated by the π-gesture [3,4,5], a non-tract variable prosodic gesture that locally slows down the constriction gestures. Although PBL is deemed physiologically and biomechanically driven, the detailed articulatory implementation of PBL and its scope have been found to be fine-tuned by higher-order linguistic structures [6,7,8,9,10], such as the prominence system of a language. For example, in English, PBL was not only realized on the phrase-final syllable regardless of stress, but also on a non-final stressed syllable [6]. PBL was further regulated by phrase-level prominence in English, being modulated by the degree of prominence [7]. The interaction between PBL and language-specific prominence was also observed with Japanese and Korean [8,9,10], but no further interaction between PBL and phrase-level prominence was detected in either language. The current study extends this cross-linguistic investigation of PBL to Mandarin Chinese in order to observe how it manifests and its scope extends in relation to the unique system of lexical tone in Mandarin.

We investigate the kinematic characteristics and the scope of PBL in Mandarin, by examining the lip closing and opening gestures during the production of monosyllabic CV words at the IP-final and the IP-medial positions. In order to further understand the fine-grained phonetic details of PBL and its interaction with language-specific prominence system, we also vary phrase-level prominence and lexical tones. Mandarin has four lexical tones that are specified with different tonal targets: a high-level tone (T1), a rising tone (T2), a low-dipping tone (T3), and a falling tone (T4). Considering that each lexical tone has distinctive tonal targets and intrinsic temporal structure [11], a specific question that arises as to how PBL may be modulated by lexical tones in Mandarin and whether and how the presumed interaction between lexical tones and PBL may be further conditioned by prominence.

Two CV sequences (/pa/, /ma/) across four lexical tones were produced by 12 speakers (6F,6M) in an EMA (Electromagnetic Articulograph) experiment. Each target word was embedded in a carrier sentence that was an answer to a question in a mini dialogue in which Boundary (IP-medial vs. IP-final) and Focus (UnFoc vs. Foc) conditions varied, as shown in Table 1. Five kinematic measures for lip closing and opening gestures of CV were taken in MATLAB, including: (a) *movement duration* (onset-target); (b) *formation duration* (onset-release); (c) *time-to-peak velocity* (onset-pkvel); (d) *movement displacement* (onset-target); (e) *peak velocity*.

The results showed no sign of PBL for the lip closing gesture of CV words, although its movement was *slower* in velocity and *smaller* in displacement phrase-finally than phrase-medially. As for the lip opening gesture, however, clearly exhibited the PBL effect with *longer*, *larger* and *slower* phrase-final movement, which could be accounted for by the theory of π-gesture [3,4,5]. *Time-To-Peak velocity*, however, was not necessarily longer associated with PBL. Furthermore, PBL interacted with focus-induced prominence (Fig.1). Under focus, PBL came with slower velocity with no spatial expansion; however, without focus, PBL came with spatial expansion but no slowing-down. This suggests that in the absence of focus-induced hyperarticulation, PBL generates both temporal and spatial expansion, possibly counteracting a slowing-down effect. PBL also interacted with lexical tones (Fig.2). Compared to simplex Tone1, spatiotemporal realization of PBL was much more robust for Tone3 (low-dipping) and Tone4 (falling) with further augmented PBL for Tone3 under focus, probably to make sufficient room for the realization of their tonal complexity. These results indicate that although PBL in Mandarin follows the cross-linguistically applicable patterns, it is modulated by the phonetically-driven phonological requirements for maximizing tonal contrast when it is licensed by prosodic structure.

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

	IP-medial	IP-final
UnFoc	A: [**mɑʊ1** mi1 pa1 pi4 ʂɤŋ4 ma?] Does **Cat** EIGHT win? B: [pu4] # [**ma1** mi1 *pa1* pi4 ʂɤŋ4.] No. **Mommy** *EIGHT* wins.	A: [ni3 ʈʂʰu1 **mɑʊ1** mi1 pa1 ma?] Do you play **Cat** EIGHT? B: [pu4] # [uɔ3 ʈʂʰu1 **ma1** mi1 *pa1*] # [pi4 ʂɤŋ4 pa?] No. I play **Mommy** *EIGHT*. Must win, right?
Foc	A: [ma1 mi1 **ta1** pi4 ʂɤŋ4 ma?] Does Mommy **BUILD** win? B: [pu4] # [ma1 mi1 ***pa1*** pi4 ʂɤŋ4.] No. Mommy ***EIGHT*** wins.	A: [ni3 ʈʂʰu1 ma1 mi1 **ta1** ma?] Do you play Mommy **BUILD**? B: [pu4] # [uɔ3 ʈʂʰu1 ma1 mi1 ***pa1***] # [pi4 ʂɤŋ4 pa?] No. I play Mommy ***EIGHT***. Must win, right?

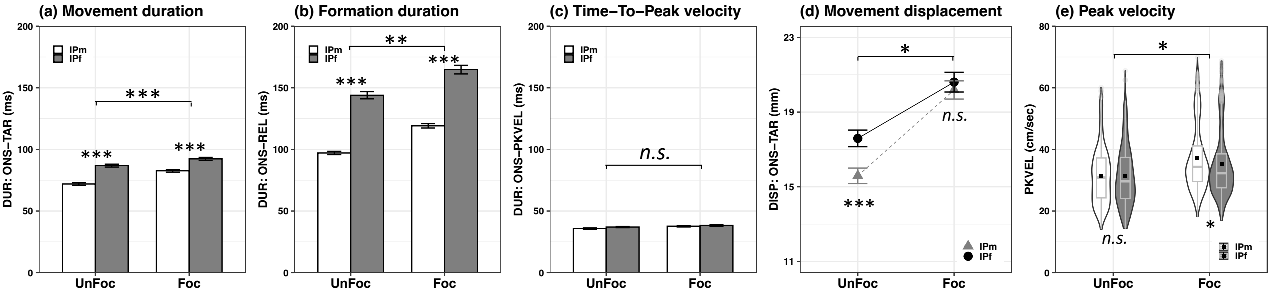


Fig.1 PBL x Prominence interactions for the lip opening gesture. Error bars show standard errors.

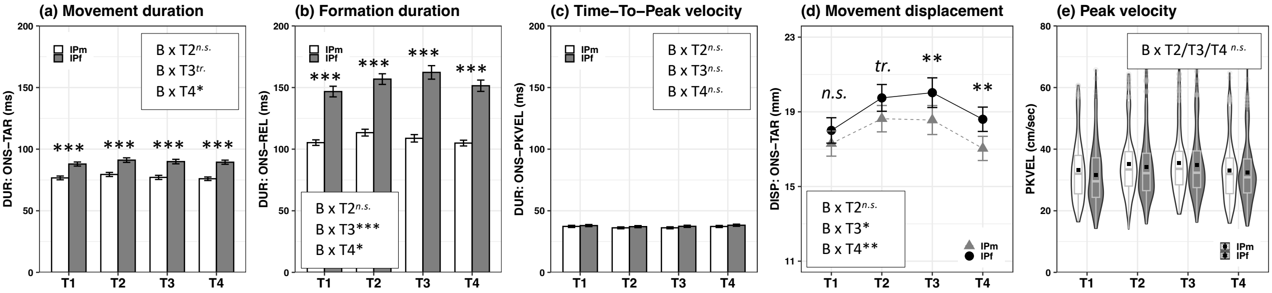


Fig. 2 PBL x Tone interactions for the lip opening gesture. Error bars show standard errors.

References

[1] Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208-1221.

[2] Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *the Journal of the Acoustical Society of America*, *89*(1), 369-382.

[3] Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149-180.

[4] Byrd, D., & Krivokapić, J. (2021). Cracking prosody in articulatory phonology. *Annual Review of Linguistics*, *7*(1), 31-53.

[5] Iskarous, K., & Pouplier, M. (2022). Advancements of phonetics in the 21st century: A critical appraisal of time and space in Articulatory Phonology. *Journal of Phonetics*, *95*, 101195.

[6] Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445-472.

[7] Kim, S., Jang, J., & Cho, T. (2017). Articulatory. characteristics of preboundary lengthening in interaction with prominence on tri-syllabic words in American English. *The Journal of the Acoustical Society of America*, *142*(4), EL362-EL368.

[8] Seo, J., Kim, S., Kubozono, H., & Cho, T. (2019). Preboundary lengthening in Japanese: To what extent do lexical pitch accent and moraic structure matter? *The Journal of the Acoustical Society of America*, *146*(3), 1817-1823.

[9] Kim, J. J., Baek, Y., Cho, T., & Kim, S. (2019). Preboundary lengthening and boundary-related spatial expansion in Korean. In *Proceedings of the 19th International Congress of Phonetic Sciences*.

[10] Jang, J., & Katsika, A. (2020). The amount and scope of phrase-final lengthening in Seoul Korean. In *Proceedings of the 10th International Conference on Speech Prosody, Tokyo, Japan* (pp. 270-274).

[11] Zhang, J. (2013). The effects of duration and sonority on contour tone distribution: A typological survey and formal analysis. Routledge.

A dynamical systems approach to F0 hard-landing downtrends in Embosi

Yubin Zhang[1], Annie Rialland[2] & Louis Goldstein[1]

[1]University of Southern California [2]Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne-Nouvelle, 4 rue des Irlandais, 75005 Paris, France

yubinzha@usc.edu, annie.rialland@sorbonne-nouvelle.fr, louisgol@usc.edu

Dynamical systems have increasingly come to be employed in the representation of speech, leveraging the lawful relation they provide between the discrete, context-free parametric specification of a system (for example a phonetic unit) and the continuous, context-dependent change in some measurable quantity [1]. Articulatory Phonology [2], [3] initially advanced this dynamical perspective by proposing that the primitive units of phonological representations and speech production are dynamically specified vocal tract constriction gestures. Later developments of this particular dynamical framework began to model prosody—both phasal boundaries and intonation—by incorporating prosodic planning dynamics [4]. The current work extends the dynamical systems approach to the investigation of utterance-level intonational events.

We focus on the utterance-level f0 downtrend phenomena, where the f0 of speech declines over an utterance, especially for declaratives. F0 downtrends have been theorized to arise from different intonational events, such as more global components like f0 downstep (downdrift) and declination, and more localized f0 final lowering [5]. Existing quantitative models of downstep, built under the assumption of separation of phonological units from quantitative f0 measurements, have included an exponential decay function to model the f0 patterns (e.g, Liberman & Pierrehumbert, 1984; Myers, 1996). On a dynamical view, these models are equivalent to a first-order system with point-attractor dynamics (whose analytic solution is an exponential decay). One problem for this model is that not all languages exhibit global intonational trends that can be readily characterized by point-attractor dynamics alone. Laniran & Clements (2003) use the terms soft landing versus hard landing to describe two types of f0 global trends with different kinematic profiles. In soft-landing like the attractor dynamics, the f0 approaches an asymptote smoothly with near zero velocity, whereas, in hard landing, f0 drops quickly and the negative velocity is relatively large utterance-finally. Laniran & Clements (2003) postulate a linear model to account for some of the hard-landing patterns found in their Yoruba data. An even more extreme pattern of hard landing seems to be present in Embosi, a Bantu language (C25) spoken in the Republic of Congo [8]. Initial observations suggest that the f0 contour in declarative sentences in Embosi exhibits initial rising and final hard landing. This suggests a distinct type of dynamical system, namely the free-fall dynamics that models the parabolic flight trajectory of gravitational fields. The current study analyzes f0 downtrends in a corpus of Embosi data and shows that it can be well modelled using primarily a free-fall-style dynamical system.

The utterances analyzed in the current study were taken from an Embosi corpus recorded by three native speakers of Embosi [8], [9]. The F0 contour of each utterance was extracted using the auto-correlation algorithm implemented in `Praat`. The mean f0 of each target moraic interval was calculated. The initial and final f0 events were analyzed separately using mixed-effects models. The statistical results reaffirm that the utterance-level f0 contour in Embosi declaratives resembles a parabolic curve, exhibiting initial rising and final 'hard-landing' patterns (See Fig. 1). We also found evidence for tone-specific f0 downtrends.

The proposed dynamical f0 hard-landing model includes two global intonational units for H and L, and one dynamical boundary L% unit. The global H and L units are two free-fall-style dynamical systems with three parameters each—initial height, initial velocity and acceleration. The L% has a blending mechanism, where the final f0 is the average of the f0 of L% and the original f0 of their own tone-specific dynamics. The fitted results are consistent with the observed patterns of initial f0 rising, final f0 hard-landing, and the tone-specific downtrends (See Fig.2).

The current work offers a dynamical treatment of intonational events, which can be incorporated into the prosodic planning dynamics component of Articulatory Phonology [2], [3]. The goal of individual lexical tones or pitch accent tones is hypothesized to be modulated by the proposed intonational dynamical units like global H/L downtrend units and final L%. Moreover, the current modelling work reconceptualizes the proposal of 'soft landing' versus 'hard landing' by Laniran & Clements (2003) in the framework of dynamical systems. A fuller picture of the dynamics of intonational units and their coordination patterns across languages remain to be investigated by future studies.

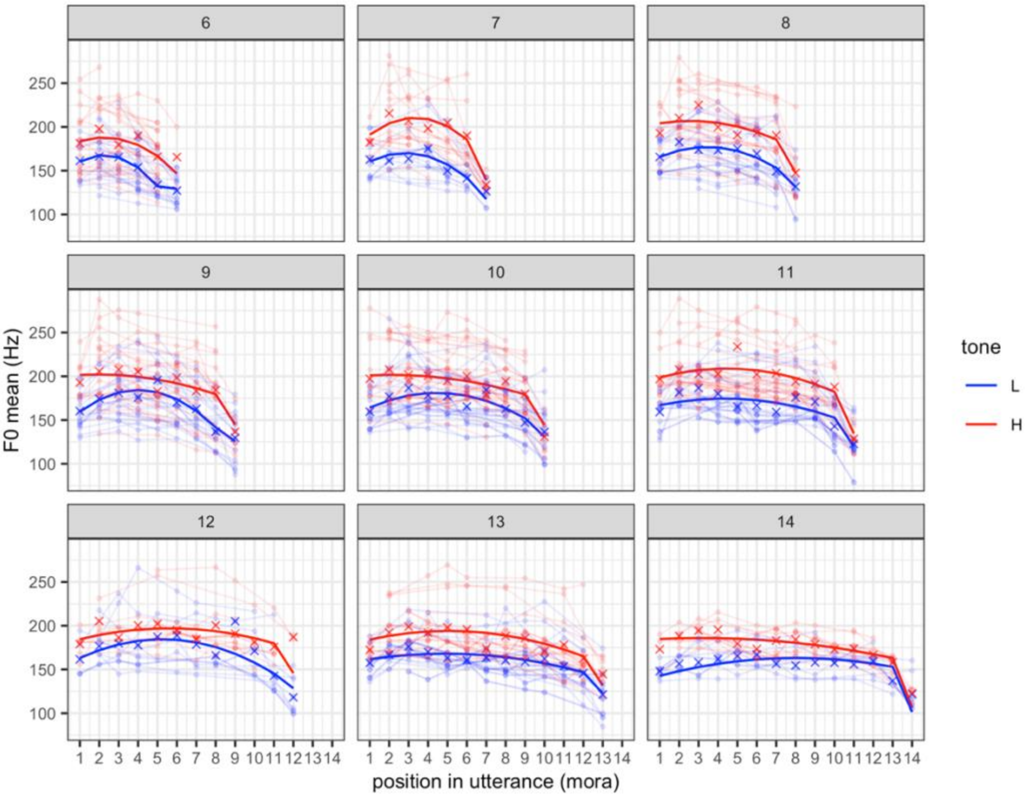


Fig.1 The fitted results based on the model versus the real data. The solid line—model predictions; The cross—averaged raw f0 values; The transparent dots and lines—individual data points.

[1] K. Iskarous, "The relation between the continuous and the discrete: A note on the first principles of speech dynamics," *J. Phon.*, vol. 64, pp. 8–20, 2017, doi: 10.1016/j.wocn.2017.05.003.

[2] E. Saltzman and K. Munhall, "A dynamical approach to gestural patterning in speech production," *Ecol. Psychol.*, vol. 1, no. 4, pp. 333–382, 1989, doi: 10.1207/s15326969eco0104_2.

[3] C. P. Browman and L. M. Goldstein, "Towards an articulatory phonology," *Phonol. Yearb.*, vol. 3, pp. 219–252, 1986, doi: 10.1017/s0952675700000658.

[4] D. Byrd and J. Krivokapić, "Cracking prosody in articulatory phonology," *Annual Review of Linguistics*, vol. 7. Annual Reviews Inc., pp. 31–53, Jan. 04, 2021. doi: 10.1146/annurev-linguistics-030920-050033.

[5] L. J. Downing and R. Rialland, "Introduction," in *Intonation in African Tone Languages*, 2016, pp. 1–16. doi: 10.1515/9783110503524.

[6] M. Y. Liberman and J. B. Pierrehumbert, "Intonational invariance under changes in pitch range and length," in *Language Sound Structure*, 1984, pp. 157–233.

[7] Y. O. Laniran and G. N. Clements, "Downstep and high raising: Interacting factors in Yoruba tone production," *Journal of Phonetics*, vol. 31, no. 2. pp. 203–250, 2003. doi: 10.1016/S0095-4470(02)00098-0.

[8] A. Rialland and M. E. Aborobongui, "How intonations interact with tones in Embosi (Bantu C25), a two-tone language without downdrift," in *Intonation in African Tone Languages*, 2016, pp. 195–222. doi: 10.1515/9783110503524-007.

[9] A. Rialland *et al.*, "Parallel corpora in Mboshi (Bantu C25, Congo-Brazzaville)," in *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 2019, pp. 4272–4276.

Acoustic Correlates of Emphatic Accent in French Vowels /i, a, u/

Hye-Sook Park[1], Sunhee Kim[2]

[1, 2]Seoul National University (Korea)
{cielcine ; sunhkim}@snu.ac.kr

The term "emphasis" refers to the "relative prominence given to certain syllable or word compared to others" [1]. An emphatic accent, an accent resulting from emphasis, affects both prosodic (suprasegmental) and segmental aspects of the syllable produced with an emphatic accent. From a prosodic point of view, a syllable with an emphatic accent is generally realized with a longer duration, higher F0, and greater intensity than a syllable without an empathic accent [2]. From a segmental point of view, the F1 values of the vowel with an emphatic accent are smaller in closed vowels and larger in open vowels, and the F2 values are larger in front vowels and smaller in back vowels [3]. Thus, the vowel space (F1*F2) of accented vowels results in an expansion, which is reported as speech enhancement phenomenon observed in clear speech and in different prosodically strengthened positions [4, 5]. The French language is known to have a "tonic accent (*accent tonique*)," which is realized on the final vowel of a rhythmic group, unless the final vowel is a schwa, with a longer duration and higher or lower pith [6]. Its primary function is to mark phrasing boundaries in a given utterance [7]. On the other hand, an *emphatic accent* in French is known to appear in the initial syllable of a word and is called by different names such as "*accent initial* (initial accent)" [8], "*additional accent*" [9], "*accent d'insistance* (accent for insistence)," and "*accent didactique* (didactic accent)" [6]. According to [10], the vowel space in the first vowel of French words is reduced compared to that of the final vowels of such words.

The objective of this study is to examine the acoustic properties of three French peripheral vowels /i, a, u/ with an emphatic accent based on an analysis of the words spoken in isolation (citation forms). We extracted 405 disyllabic words from the NAVER French-Korean dictionary, which is designed for Korean learners of French. Each word in the dictionary is provided with a sound file recorded by a local voice actress who speaks standard French. The words used for this study are each of a C1V1.C2V2 structure with stop consonant as the preceding consonant. A total of 241 words with an emphatic accent on V1 are selected (/ˈi/ 71 ***pi**pa*, /ˈa/ 101 ***ta**pa*, /ˈu/ 69 ***cou**per*) and 164 words without an emphatic accent on V2 are selected (/i/ 58 *pap**i***, /a/ 64 *cat**a***, /u/ 42 *couc**ou***). Acoustic measurements for duration, F0, intensity, F1, F2, and F3 were performed using Praat. Statistical analysis was conducted by using SPSS.

Table 1 presents the results of a t-test for prosodic features with and without an empathic accent for /i, a, u/. In Table 1, F0 and intensity show statistically larger values in all three vowels, while the duration of V1 is significantly shorter than the duration of V2 in all three vowels. A regression analysis on the prosodic features indicate that the most significant prosodic feature in V1 was intensity (β=0.660), followed by duration (β=-0.224), and F0 (β=-0.089). Table 2 reveals the results of the t-test for segmental features with and without an empathic accent. The F1 values of V1 are significantly higher than that of V2 for high vowels /i, u/, while F2 values of the low vowel /a/ show significantly higher values in V1. As for F3, the front high vowel /i/ is higher in V2, while the back high vowel /u/ is significantly higher in V1. This means that /i/ becomes more round and /u/ becomes less round in V1. Finally, Figure 1 illustrates that the vowel space of V1 (24.83kHz^2) and V2 (25.74kHz^2).

In this study, we have examined the acoustic properties of three French peripheral vowels /i, a, u/ with an emphatic accent in prosodic and segmental aspects. The results show that compared to unaccented vowels the duration of the accented vowels is shorter and the vowel space smaller. The results are similar to studies done on French [10] but do not conform to previous works on English, Croatian or Korean [2, 3, 4, 5]. Based on the results, it may be asserted that the French words have different acoustic and articulatory behavior compared to other languages such as English, Croatian or Korean. Further research will be required to confirm this assertion and to identify their causes.

Table 1. Results of t-test for prosodic features with and without an empathic accent for /i, a, u/

Prosodic features		M(SD)	*t*(*p*)
Duration(ms)	[ˈi]	129.8(30.5)	**-5.194(0.000)*****
	[i]	157.2(28.7)	
	[ˈa]	126.6(23.6)	**-5.663(0.000)*****
	[a]	147.3(21.7)	
	[ˈu]	118.8(24.2)	**-3.341(0.001)****
	[u]	134.1(22.0)	
F0(Hz)	[ˈi]	320.1(34.1)	**36.154(0.000)*****
	[i]	154.5(16.4)	
	[ˈa]	300.6(29.2)	**48.954(0.000)*****
	[a]	137.7(13.0)	
	[ˈu]	323.4(29.5)	**36.540(0.000)*****
	[u]	159.5(17.8)	
Intensity(dB)	[ˈi]	78.7(3.5)	**15.027(0.000)*****
	[i]	69.6(3.2)	
	[ˈa]	81.2(3.7)	**16.700(0.000)*****
	[a]	72.8(2.8)	
	[ˈu]	81.5(2.7)	**19.530(0.000)*****
	[u]	71.4(2.5)	

Table 2. Results of t-test for prosodic features with and without an empathic accent on /i, a, u/

Segmental features		M(SD)	*t*(*p*)
F1	[ˈi]	362.8(32.9)	**7.917(0.000)*****
	[i]	308.6(42.7)	
	[ˈa]	781.1(76.6)	-1.256(0.211)
	[a]	796.6(78.5)	
	[ˈu]	388.3(52.6)	**3.613(0.000)*****
	[u]	352.5(47.1)	
F2	[ˈi]	2327.3(131.6)	1.696(0.092)
	[i]	2295.3(81.0)	
	[ˈa]	1810.4(230.0)	**7.098(0.000)*****
	[a]	1572.4(173.1)	
	[ˈu]	1108.4(162.9)	-1.148(0.256)
	[u]	1175.2(355.2)	
F3	[ˈi]	3199.8(222.4)	**-8.696(0.000)*****
	[i]	3448.6(84.1)	
	[ˈa]	2922.1(209.1)	1.070(0.286)
	[a]	2895.8(105.3)	
	[ˈu]	2867.2(196.1)	**4.078(0.000)*****
	[u]	2719.9(163.5)	

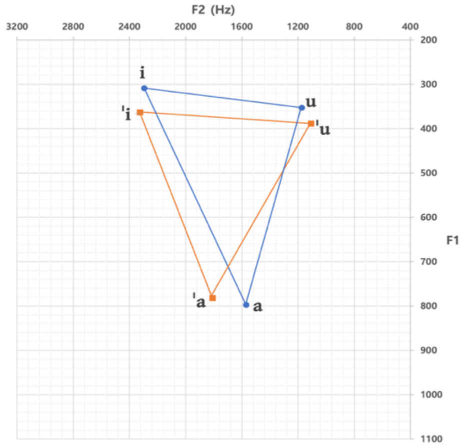


Figure 1. Vowel space (F1*F2) of /i, a, u/ with (red) and without (blue) an emphatic accent

References

[1] Bean, C., Folkins, J. W., & Cooper, W. E. (1989). The effects of emphasis on passage comprehension. *Journal of Speech, Language, and Hearing Research*, 32(4), 707-712.
[2] Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: The MIT Press. 110, 131.
[3] Lindblom, B., Agwuele, A., Sussman, H. M., & Cortes, E. E. (2007). The effect of emphatic stress on consonant vowel coarticulation. *The Journal of the Acoustical Society of America*, *121*(6), 3802-3813.
[4] Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. Journal of Phonetics, 39(3), 344-361.
[5] Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. Journal of Phonetics, 36(1), 91-113.
[6] Carton, F. (1974). *Introduction à la phonétique du français*. Bordas. 117-122.
[7] Vaissière, J. (1991). Rhythm, accentuation and final lengthening. *Music, language, speech and brain*, 59, 108-120.
[8] Astésano, C., Bard, E. G., & Turk, A. (2007). Structural influences on initial accent placement in French. *Language and speech*, 50(3), 423-446.
[9] Dahan, D., & Bernard, J. M. (1996). Interspeaker variability in emphatic accent production in French. *Language and speech*, *39*(4), 341-374.
[10] Gendrot, C., & Adda-Decker, M. (2006). Analyses formantiques automatiques en français : périphéralité des voyelles orales en fonction de la position prosodique. *Proc. Journées d'Étude sur la Parole*.

Edgy articulation: the kinematic profile of Accentual Phrase boundaries in Seoul Korean

Jiyoung Jang[1] & Argyro Katsika[1]

[1]*University of California, Santa Barbara (USA)*
jiyoung@ucsb.edu, argyro@ucsb.edu

Korean is an edge-prominence language, in which phrasal prominence is marked by the means of prosodic phrasing. For instance, the focused word consistently starts an Accentual Phrase (AP) or a higher phrase, and any following AP boundaries up to the end of the Intonational Phrase (IP) is known to undergo elimination, or possibly attenuation, referred to as dephrasing [1]. APs, with a proposed underlying tonal pattern of THLH (where the type of T depends on the AP-initial segment [2]), serve as the basic intonational unit in Korean (see [3, 4]). Despite this functional load of APs in the language's prosody, findings are scarce on the phonetic dimensions of these prosodic boundaries. The present study aims to examine the kinematic manifestation of constriction gestures at the left edge of APs encoding different types of focus information. Specifically, test words vary with respect to focus status. They are 1) focused, being initial in a focused AP, 2) unfocused, but not dephrased, being initial in a pre-focal AP, or 3) unfocused and expected to be dephrased, by virtue of following a focused AP. We predict gestures under focus to be longer, larger, and faster, based on previous articulatory research on phrasal prominence mainly in head-prominence languages [5] and limited work on Korean ([7], see [8] for a review). However, the stretch of speech affected by the focus effect as well as the kinematic manifestation of dephrasing are still unclear issues. One hypothesis is that the tonal attenuation observed in dephrasing is accompanied by articulatory attenuation, i.e., shorter and smaller AP-initial gestures as compared to their unfocused counterparts, i.e., effects similar to those of de-accentuation in stress languages (e.g., [9]).

Seven native Seoul Korean speakers (5F, 2M) participated in an Electromagnetic Articulography study. Test word /minami/ was embedded in stimuli sentences (Table 1). Position of (contrastive) focus was varied, as prompted by mini dialogues, so as to yield the following focus types on the test word: 1) focused, when focus is on the test word, 2) unfocused, when focus is on the AP following the test word, and 3) dephrased, when focus is on the initial AP. Example of the dephrased condition is shown in Table 1. Eight repetitions of each condition, randomized along with other stimuli examining other aspects Korean prosody, were collected, except for one speaker, by whom five repetitions were recorded due to technical reasons. The acquired data were checked for their prosodic rendition, which, among other dimensions, confirmed tonal attenuation due to dephrasing. Consonant (C) gestures of the test prosodic word /minami/ was measured. Lip aperture was used for measuring /m/ and tongue tip vertical displacement for /n/. A semi-automatic procedure was used to detect the points in time in which each C gesture reached its onset, peak velocity, maximum position, and release. The following kinematic measures were calculated based on these time-points: formation duration (interval between onset and release), displacement (spatial difference between max and onset), and formation's peak velocity. The retrieved data were normalized within each C gesture and analyzed by linear mixed effects analysis with Type (Focused, Unfocused, Dephrased) as a fixed factor and Speaker as a random factor using the *lme4* package in R. Pairwise comparison was done by the means of the *emmeans* package.

Results are summarized in Figure 1. There was a main effect of Type in all three kinematic dimensions—formation duration, displacement, peak velocity. Pairwise comparisons confirmed that gestures under focus were longer, larger, and faster than their unfocused and dephrased counterparts, as predicted based on previous findings (e.g., [5, 7]). The focus-induced prominence effect spanned over the measured C gestures, indicating that the scope of the effect is not local to the boundary but extends at least three syllables away, possibly affecting the whole AP-initial word. With respect to dephrasing, initial (C1) and medial (C2) C gestures differentiated between the unfocused and dephrased conditions: C1 was larger and C2 was faster in dephrased condition as opposed to unfocused ones (Figure 1). These results are contrary to our prediction of articulatory attenuation in dephrasing, and is possibly attributable to a spillover effect from focus on the immediately preceding AP.

Overall, findings indicate that the kinematic profile of phrasal prominence marking in an edge-prominence language like Korean is similar to that of head-prominence languages: prominent gestures are larger, longer and faster with the scope of the effect expanding beyond the boundary-adjacent and stressed syllable respectively [10]. Also, results suggest that dephrasing is more of a tonal attenuation than articulatory attenuation. Instead, dephrased gestures might undergo some focus-induced spillover strengthening effect.

Table 1. Example dialogue reflecting the dephrased condition, i.e., focus on the initial AP. Focused words are in bold and measured intervals are underlined.

Type	*Example dialogue*
Dephrased	(Participants were asked to imagine a situation where there are two Uncle Minams, one in a magic club and another in a secrecy club.) Prompt sentence: 'It's not the one in the magic club.' Test sentence: [$_{\text{AP}}$ **pimilpu**] [$_{\text{AP}}$ minami gomopuga] [$_{\text{AP}}$ nɛmaŋminam]? 'Uncle Minam of **the secrecy club** is the handsome guy from Nemang?'

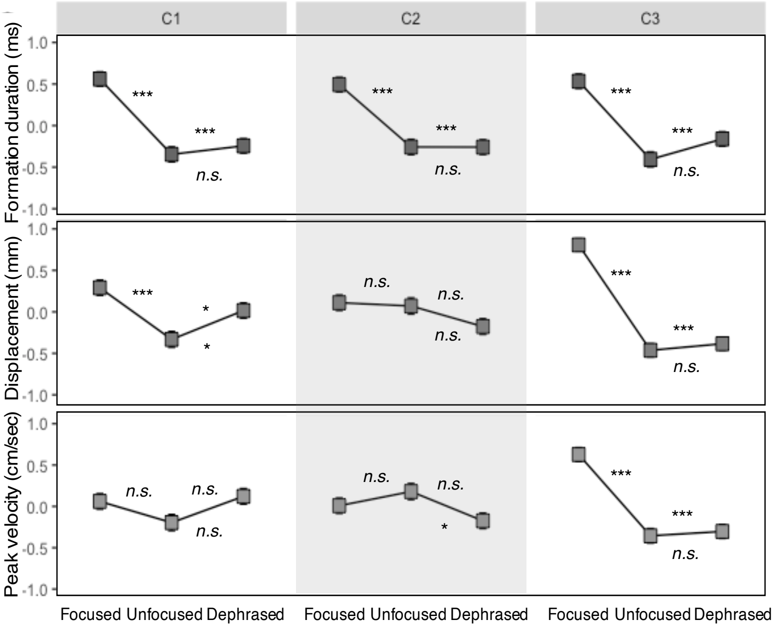


Fig. 1 Normalized duration, displacement, and peak velocity of C gestures by AP Type. Asterisks indicate results of pairwise comparisons: ***='*p*<0.001', *='*p*<0.05', *n.s.*='*p*>0.05'.

References

[1] Jun, S. A. (1993). *The Phonetics and Phonology of Korean*. PhD dissertation, The Ohio State University.

[2] Jun, S. A. (2000). K-tobi (korean tobi) labelling conventions. *Speech Sciences*, *7*(1), 143-170.

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

[4] Jun, S. A. (2005). Korean intonational phonology and prosodic transcription. *Prosodic typology: The phonology of intonation and phrasing*, 1, 201.

[5] Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Papers in Laboratory Phonology VIII: Varieties of Phonological Competence (Phonology and Phonetics)* (pp. 519–548). Berlin, Germany: Mouton de Gruyter.

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

[7] Shin, S., S. Kim, and T. Cho. (2015). What is special about prosodic strengthening in Korean: evidence in lingual movement in V#V and V#CV. *Proc. 17th ICPhS*, Glasgow, U.K.

[8] Cho, T. (2022). The phonetics-prosody interface and prosodic strengthening in Korean. *The Cambridge Handbook of Korean Linguistics Cambridge Handbooks in Language and Linguistics* (pp. 248-293). Cambridge University Press Cambridge.

[9] Katsika, A., Jang, J., Goldstein, L., Krivokapic, J., & Saltzman, E. (2019). The kinematics of prominence in American English. *The Journal of the Acoustical Society of America*, *146*(4), 3084-3084. https://doi.org/10.1121/1.5137712

[10] Katsika, A., & Tsai, K. (2021). The supralaryngeal articulation of stress and accent in Greek. *Journal of Phonetics*, 88, 101085.

Some Asymmetrical Pre- Versus Post-focal Effects on Articulatory Realization of Prominence Distribution in Korean: A Preliminary Report

Suyeon Im[1], Sahyang Kim[2] & Taehong Cho[3]
[1]Soongsil University (Korea), [2]Hongik University (Korea), [3]Hanyang University (Korea)
sim@ssu.ac.kr, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

In the present study, we investigate how CVC words are reduced in non-prominent (non-focal) contexts adjacent to the focal context in Korean, relative to when they are focused. That is, we explore how a language which does not employ lexical stress in the prominence system expresses prominence distribution in relative terms and how the results may compare to existing prominence-related data available in other languages such as English (e.g., [1, 2]) that uses a typologically different prominence system. Moreover, we use an Electromagnetic Articulograph, so that we can examine how reduction in non-focal contexts versus hyperarticulation in focal contexts are kinematically expressed in both spatial and temporal dimensions.

A specific question to be addressed is how articulatory reduction of pre-focal versus post-focal words may be similar to or different from each other, as compared to when the same words are focused. Both the pre- and post-focal conditions are expected to induce articulatory reduction, but given the directional asymmetry of pre- versus post-focal conditions, we expect that the degree of reduction will also be asymmetrical. It is hypothesized that the degree of articulatory reduction will be greater in the pre-focal than in the post-focal context, assuming that the auditory-perceptual impacts are likely to be greater when there is a drastic and rapid increase in articulatory force from a non-focal gesture on the following focal gesture rather than the other way around (e.g., [3]).

Twelve speakers of Seoul Korean participated in an articulatory experiment using an Electromagnetic Articulograph (Carstens AG501). The participants read provided written sentences on a computer screen in response to question sentences, which were designed to elicit corrective focus in the answer sentences on (a) the target word (focal condition), (b) the preceding word (pre-focal condition) and (c) the following word (post-focal condition). Movement duration (ms; DUR), displacement (mm; DISP), and peak velocity (cm/sec; PKVEL) were measured and were submitted to a linear mixed-effects model which was run for each movement (C1 closing, C-to-V opening, C2 closing) with Focus (focal, pre-focal, post-focal) and Word (/pap/, /pam/) as fixed factors including their interaction. The random structure included by-subject intercept and slope for all the fixed factors.

A basic finding was that non-focal gestures were much more reduced in both spatial and temporal dimensions than focal gestures that received a corrective contrastive focus (Figure 2). In other words, focal gestures were hyperarticulated (being larger, longer and faster than non-focal gestures), being ‘prominent’ above the surrounding non-focal words. This hyperarticulation pattern in Korean is largely consistent with the hyperarticulation pattern generally reported in English [1, 2, 4].

As for the specific research question of how pre-focal versus post-focal effects may differ from each other relative to focal effects, our results indicated that pre-focal gestures tended to be reduced more than post-focal gestures. This asymmetry was evident in two cases. For one thing, C1-closing gesture when in the pre-focal condition was reduced (in displacement and peak velocity) compared to when in the focal condition, whereas the same C1-closing gesture of the post-focal word showed no such reduction (Figure 1). For another, while C2-closing gesture was substantially reduced when in both pre- and post-focal conditions, it was the pre-focal C2-closing gesture that was reduced more (as evident in displacement and peak velocity), relative to the post-focal C2-closing gesture (Figure 3).

These results indicate that the nature of reduction differs depending on the directionality of prominence distribution—i.e., whether it occurs in the pre-focal or post-focal context. On the one hand, the pre-focal word was reduced as a whole from the beginning C1-onset gesture to the final C2-closing gesture which was immediately adjacent to the focal word. On the other hand, the post-focal word was reduced in a rather progressively gradient way.

The present study was the first to explore how prominence distribution would be reflected in kinematic terms in Korean, especially with respect to reduction patterns of pre-focal versus post-focal gestures as compared to focal gestures. The general differences in the focal versus the non-focal contexts illuminate that relative prominence is kinematically realized in a form of ‘hyperarticulation’ in much the same way across languages. Furthermore, the directional asymmetry of pre-focal versus post-focal effects further implies that although prominence may be defined differently in the phonology of a given language, articulation of prominence is fine-tuned by the production system of the language that optimizes prominence distribution taking into account both the listener-oriented auditory-perceptual saliency and the speaker-oriented motor efficiency.

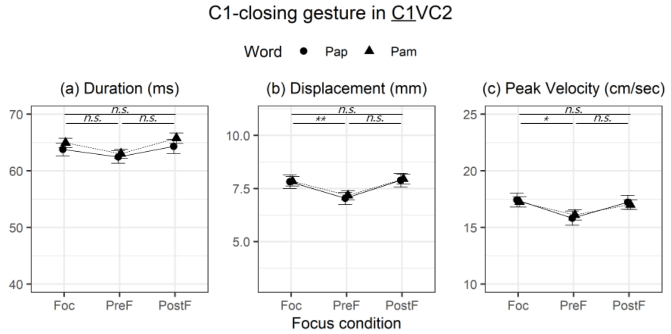


Figure 1. C1 lip-closing duration (left), displacement (middle), and peak velocity (right)

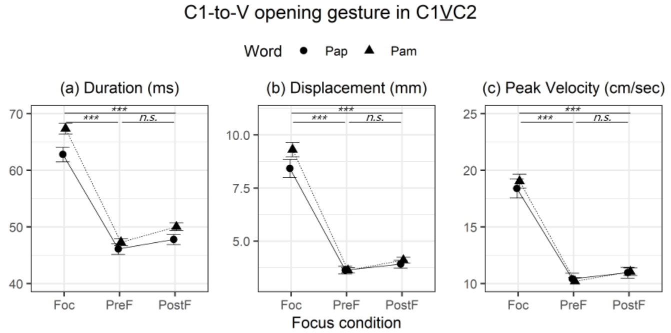


Figure 2. The V lip opening duration (left), displacement (middle), and peak velocity (right)

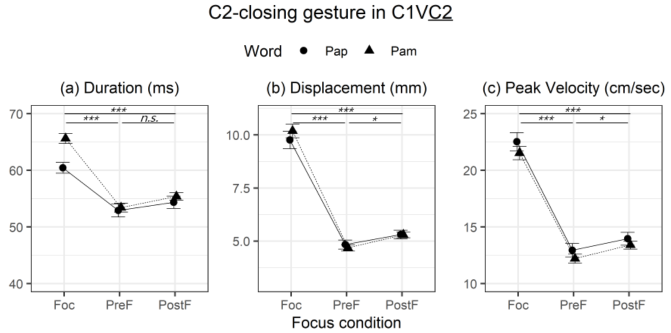


Figure 3. The C2 lip closing duration (left), displacement (middle), and peak velocity (right)

References

[1] de Jong, K. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Am. 97,* 491–504.

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

[3] Wright, R. 2004. A review of perceptual cues and cue robustness. *Phonetically Based Phonology, 34*, 57.

[4] de Jong, K. 1991. An articulatory study of consonant-induced vowel duration changes in English. *Phonetica, 48,* 1–17.

Bottom-up learning of a phonetic system using an autoencoder

Frank Lihui Tan & Youngah Do
The University of Hong Kong
tt115889@connect.hku.hk, youngah@hku.hk

Human learners tend to perceive and acquire sound input categorically, even if the input is of a gradient nature [1, 2, 3]. This tendency results in the acquisition of a "prototypical phonetic system" [4], in which sounds near phonetic category centers are perceived as closer to their centers than they actually are. It is unclear how much of this prototypical system is learned through language experience and how much can be attributed to humans' innate cognitive properties, as infants as young as one month old already show categorical perception [1].

The current study aims to investigate whether 'naïve' learners can still acquire a prototypical phonetic system at the very initial stage of phonetic learning, before they have acquired any knowledge about their language's lexicon or structure. To answer this question, we test an autoencoder model that is trained purely in a bottom-up fashion, without assuming any abstract featural system. Unlike previous models, we evaluate the model's learning outcome directly on the hidden representations of phones, instead of on the basis of distinctive features.

Our results show that the autoencoder model is able to capture the phonetic system in the form of phone distributions in the hidden space, which is a close analogy to the prototypical category learning observed in human learners [4]. We selected two typologically unrelated languages, English and Mandarin, and used 38 hours of recordings from 40 English speakers (Buckeye Speech Corpus [5]) and a comparable size of recordings from Mandarin speakers (AISHELL-3 [6]) as training data. The recordings were in wave format, and we extracted mel-frequency cepstral coefficients before training. We then randomly segmented the data to ensure that no or minimal phonological cues and segmental boundary information was provided. No ground truth labels were incorporated. For each language, we built an autoencoder to encode the input information into a hidden space and reconstruct the hidden representation back to the input with least distortion [7]. The encoder simulated the complex, layered neural transformations underlying speech perception, which ultimately converted the sensory receptor signal to the underlying neural code of segments [8]. The decoder simulated the reverse process of generating sounds from internal representations, but excluded articulation since there is no simulation of articulators. The training was unsupervised, without external feedback, and no segment boundary was provided, which is analogous to the early stage of infants' phonetic acquisition [9].

The preliminary clustering task revealed that phonetic knowledge successfully emerged in the hidden space for both English and Mandarin languages. The models' hidden representations yielded significantly higher homogeneity, completeness, and V-measure scores than random clustering (e.g., $V_{\_Random\_English}$=0.006 vs. $V_{\_English}$=0.289), indicating that the autoencoder was able to reproduce the input sounds and identify phonetic category centers, even without phonological context or segmental boundary information. Further evaluations of the hidden representations showed that the model was able to project tokens of the same phone to similar areas and tokens of different phones to different areas in the hidden space, while successfully learning feature-based contrasts such as [±back], [±high], [±strident], and [±voice]. The current model trained solely on phonetic cues was able to construct a phonetic system that distinguishes sounds, implying that the model was able to project an acoustic token to its correct absolute position, rather than merely achieving paired phone contrasts.

Although the model achieved significant success, it did not capture the acquisition of allophones. Unlike top-down models [10, 11], which take into account human's different perceptual sensitivity and learnability of phonemes and allophones [12, 13], the current bottom-up model did not show a significant difference between phoneme and allophone projections. For example, the distribution of allophones of Mandarin /i/ (/i, ɹ̩, ɻ̩/) was similar to that of phonemes (μ (i,ɹ,ɻ dists)=2.978; μ (y, ɚ, a, ɤ, u dists)=3.194; p=0.817).

In summary, this study demonstrates that an autoencoder model can learn phonetic knowledge from contextless acoustic input without supervision or explicit segment boundary. The model was able to project different phones to different areas in the hidden space, similar to the way human infants acquire phonetic knowledge. This suggests that infants' phonetic knowledge may not be innate but can be acquired based purely on acoustic information, without relying on language-specific learning facilities. However, the model could not reach phonological knowledge without training on phonological cues, which is acquired by infants at around 8-10 months [14]. This implies that phonological information plays an indispensable role in the language-specific refinement of learners' knowledge on phonetics and phonology.

References

[1] Eimas, P. D. et al. (1971). Speech Perception in Infants. *Science*, *171*(3968), 303–306.

[2] Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant behavior and development*, *10*(3), 279-293.

[3] Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672–683.

[4] Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107.

[5] Pitt, M. A. et al. (2007). *Buckeye Corpus of Conversational Speech (2nd release).*

[6] Shi, Y. et al. (2021). *AISHELL-3: A Multi-speaker Mandarin TTS Corpus and the Baselines* (arXiv:2010.11567).

[7] Bank, D. et al. (2021). *Autoencoders.*

[8] Eggermont, J. J. (2001). Between sound and perception: reviewing the search for a neural code. *Hearing Research*, *157*(1–2), 1–42.

[9] Räsänen, O. (2014). Basic cuts revisited: Temporal segmentation of speech into phone-like units with statistical learning at a pre-linguistic level. *Annu. COGSCI.*

[10] Kolachina, S., & Magyar, L. (2019). What do phone embeddings learn about Phonology? *Proceedings of SSIGMORPHON*, 160–169.

[11] Silfverberg, M., Mao, L. J., & Hulden, M. (2018). Sound analogies with phoneme embeddings. *Proceedings of SCiL*, 136–144.

[12] Martin, A. et al. (2013). Learning Phonemes With a Proto-Lexicon. *Cognitive Science*, *37*(1), 103–124.

[13] Peperkamp et al. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition*, *101*(3), B31–B41.

[14] Hayes, B. (2004). Phonological acquisition in Optimality Theory: The early stages. In R. Kager et al. (Eds.), *Constraints in Phonological Acquisition* (pp. 158–203). CUP.

Russian Prosody as a Special Case of the Mobile Stress System in Indo-European Languages

Lindy Comstock[1,2]

[1]*University of California, Los Angeles (USA),* [2]*HSE University (Russia)*
lbcomstock@ucla.edu

Redundancy is an element of linguistic systems which constrains variation in features that convey an important function, and which counteract phonological change and preserve the intelligibility of speech. Derivational processes are a means by which dual representational systems may be acquired, particularly in unmarked structures. Highly regular and resistant to phonological change [1,2], Russian adjectival suffixes are typically unstressed and pronounced with phonetic vowel reduction, yet they are perceived to bear pitch accents [3]. This paper proposes that the intersection of phonological and morphological features found in Russian adjectival suffixes (i) forms a dual representational system that preserves word-internal metrical structure, and (ii) sheds light on unexplained features of Russian prosody.

There is currently no theoretical model that can fully explain the prosodic features observed in Russian [4-7]. Scholars agree that Russian prosody is notable for its strong macro-rhythm and inventory of bitonal pitch accents [6,7], yet existing models exclude accentual phrases and edge tones because researchers have found it difficult to reconcile the perception of additional tones on syllables that bear no stress with a traditional understanding of how prominence is marked. Halle's [8] work on stress and accent in Indo-European (IE) languages offers a possible solution based on Idsardi's [9] AM theory. Only one stress may be assigned per Russian word [10], whereas the morphemes are inherently accented [8]. This conflict is resolved in IE languages with mobile stress by means of the LLL edge-marking rule: a L(eft) parenthesis placed to the L(eft) of the L(eftmost) element in the string. Word stress is realized on the syllable that is the word head. This system is illustrated with the word "democratic" (demokratičeskaâ) in Figure 1. Every syllable that could bear stress is marked with * in line 0, and a bracket denotes the onset of a foot. Applying the rule, word stress is calculated (see line 2). While Halle did not study prosody, he notes that word stress is accompanied by a high tone [8,11]. The fact that low tones are assigned to all syllables except the one bearing stress suggests that a salient difference is perceived between high and low tones.

The influence of mobile stress in Russian is seen throughout its morphological and phonological paradigms. Therefore, the lexical accent assigned to morphemes may persist as a representation within the prosodic structure of Russian words. There is no evidence that every syllable bears a tone; instead, the salient difference Halle perceived likely reflects the H+L bitonal pitch accent on stressed syllables. The second high tone and the subsequent fall perceived by researchers on non-stressed morphemes may represent boundary tones between morphological units, particularly those serving an important grammatical function, like derivational suffixes (Figure 2). Phonetic variation also serves as a cue to the internal structure of words in a wide variety of languages [12-13]. Therefore, additional support for our hypothesis lies in the systematic use of a subset of vowels in adjectival suffixes and the properties of vowel reduction in Russian. Suffixes encode gender, number, and case, such that the perception of an internal morphological boundary between the root and suffix may be critical for deciphering word meaning. Russian generally observes two patterns of vowel reduction. The second concerns vowels in all positions other than the pretonic and tonic position. The examples utilize the feminine suffix in nominative case, which comprises two vowels after the initial consonant: unpalatalized /a/ and palatalized /j+a/. During reduction, the initial /a/ is raised /ə/, and we would expect /ja/ to raise to /i/ [15]. However, vowel reduction after palatalized consonants in Russian declensions is not fully realized [15]. This phenomenon reflects the need to preserve critical paradigmatic information that is encoded in Russian derivational suffixes [15]. Therefore, we would anticipate only the first vowel to be raised in the given phonetic environment.

This interpretation provides a justification for the H to L fall in pitch that is perceived in Russian adjectival suffixes when in a non-phrase final position (Figure 3). The high tone, represented by the reduced initial vowel, is followed by a palatalized vowel without full reduction, and therefore is perceptually lower. To investigate this hypothesis, Russian pitch accents and the F1/F2 values for vowels were analyzed in adjectives: the feminine singular nominative ending preceded by (i) /k/ (any – вся́кая, democratic – демократи́ческая), and (ii) /n/ (stupid – глу́пая, smart – у́мная). The sentences were produced by four native speakers of Russian (2 female). The dataset comprised between 40-50 instances of each suffix vowel (/a/,/ja/) per participant. ANOVA and regression analyses was performed to identify if (i) vowel type or (ii) formant values would predict perception of a high or low tone. Position was found to be a significant predictor of the mean vowel pitch.

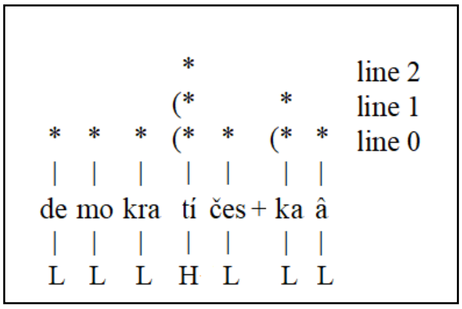


Fig.1 Isardi’s Model

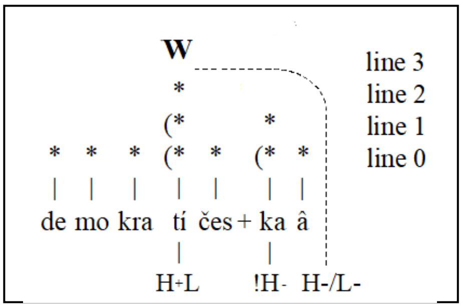


Fig.2 Proposed Model

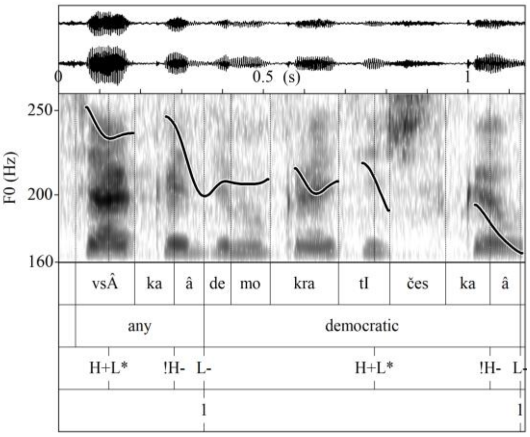
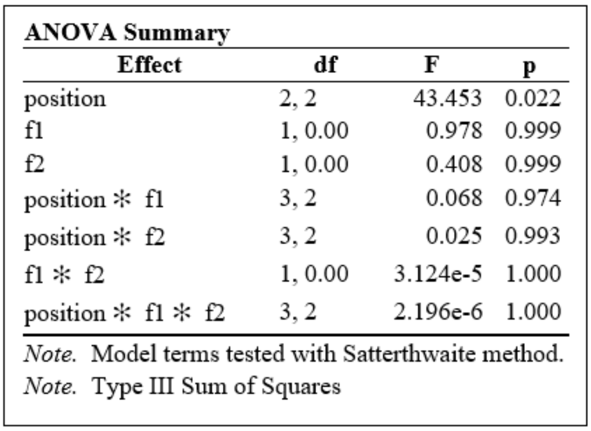


Fig.3 Tones in Russian Adjectives



ANOVA Summary

Effect	df	F	p
position	2, 2	43.453	0.022
f1	1, 0.00	0.978	0.999
f2	1, 0.00	0.408	0.999
position * f1	3, 2	0.068	0.974
position * f2	3, 2	0.025	0.993
f1 * f2	1, 0.00	3.124e-5	1.000
position * f1 * f2	3, 2	2.196e-6	1.000

Note. Model terms tested with Satterthwaite method.
Note. Type III Sum of Squares

Fig.4 Statistical Analysis

References

[1] Hamilton Jr, W. S. (1976). Vowel power versus consonant power in Russian morphophonemics. *Russian linguistics*, *3*(1), 1-18.

[2] Kiparsky, P. (2003). The phonological basis of sound change. *The handbook of historical linguistics*, 311-342.

[3] Yokoyama, O. T. (2001). Neutral and non-neutral intonation in Russian: A reinterpretation of the IK system ('intonational constructions','intonatsionnye konstruktsii', or IKs). *WELT DER SLAVEN-HALBJAHRESSCHRIFT FUR SLAVISTIK*, *46*(1), 1-26.

[4] Odé, C. (1989). *Russian intonation: a perceptual description* (Vol. 3). Rodopi.

[5] Odé, C. (2003). Description and transcription of Russian intonation (ToRI). *Studies in Slavic and General Linguistics*, *30*, 279-288.

[6] Igarashi, Y. (2005). Phonology of Russian intonation. [Doctoral dissertation, Tokyo University of Foreign Studies]. Accessed online at http://repository.tufs.ac.jp/handle/10108/35618.

[7] Yokoyama, O. T. (2001). Neutral and non-neutral intonation in Russian: A reinterpretation of the IK system. *WELT DER SLAVEN-HALBJAHRESSCHRIFT FUR SLAVISTIK*, *46*(1), 1-26.

[8] Halle, M. (1997). On stress and accent in Indo-European. *Language*, 275-313.

[9] Idsardi, W. J. (1992). *The computation of prosody* (Doctoral dissertation, Massachusetts Institute of Technology).

[10] Kiparsky, P, & Halle, M. (1977). Towards a reconstruction of the Indo-European a Studies in stress and accent. (Southern California Occasional Papers in Linguistics, 1.) Los Angeles: University of Southern California.

[11] Bethin, C. Y. (2006). Stress and tone in East Slavic dialects. *Phonology*, *23*(2), 125-156.

[12] Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(3), e1547.

[13] Vaysman, O. (2009). *Segmental alternations and metrical theory* (Doctoral dissertation).

[14] Halle, M. (1971). *The sound pattern of Russian: A linguistic and acoustical investigation*. De Gruyter Mouton.

[15] Bethin, C. Y. (2012). On paradigm uniformity and contrast in Russian vowel reduction. *Natural Language & Linguistic Theory*, *30*, 425-463

A Perceptual Study on the Distinctive Features of Entering Tones in Guangzhou Cantonese

Dong Han[1], Mingyang Yu[2]

[1]Hong Kong Shue Yan University (Hong Kong), [2]City University of Hong Kong (Hong Kong)
dhan@hksyu.edu, mingyanyu5-c@my.cityu.edu.hk

Background: Acoustic experiments indicate that in Cantonese the monosyllables of the high-level tones (HLT) and high-level entering tones (HLET) share pitch values, but differ in duration, as do the mid-level tones (MLT) and mid-level entering tones (MLET), and the low-level tones (LLT) and low-level entering tones (LLET) [1]. However, various studies have reported differences in the contribution of the features of the stop coda and the short duration to the perception of entering tones in Guangzhou Cantonese [2,3,4]. To better investigate the role of the two features in the perception of entering tones, the present study conducted three perceptual experiments on 12 native Guangzhou Cantonese speakers (6 male, 6 female) by manipulating the duration of three pairs: namely, "HLT - HLET", "MLT - MLET", and "LLT - LLET", respectively.

Stimuli and experimental procedures: As claimed by [5], [iː] is a long vowel when pronounced as HLET, MLET, and LLET. Hence, we selected three pairs of monosyllables with the rhymes [iː] and [iːt] from the entering and non-entering tone pairs. The three pairs of original sounds were taken from [6] to prepare the stimuli: (1) 枝[tsiː55] (branch) and 滶[tsiːt5] (squirt), with durations of 430 ms and 280 ms, respectively; (2) 智[tsiː33] (wise) and 节[tsiːt3] (festival), with durations of 450 ms and 300 ms, respectively; and (3) 字[tsiː22] (character) and 截[tsiːt2] (cut), with durations of 430 ms and 320 ms, respectively.

A total of 66 stimuli were prepared by manipulating the duration of each syllable in 10 steps with the Time-Domain Pitch-Synchronous Overlap-and-Add function in Praat [7]. The duration differences between stimuli in the same pairs were the same. In each experiment, the duration differences between stimuli were calculated from the duration differences between the two ends of the pair. For example, the durations of 枝[tsiː55] and 滶[tsiːt5] are 430 ms and 280 ms, respectively. A total of 20 stimuli were created, with 10 stimuli produced from 枝[tsiː55] and 10 stimuli produced from 滶[tsiːt5], each with a duration difference of (430 - 280) / 10 = 15 ms.

The experiments were conducted through Gorilla [8], an online experiment platform. The order of the three experiments was randomly determined by Gorilla, and each experiment took approximately five minutes to complete, with a five-minute break between experiments. During each experiment, the Gorilla system would randomly play 110 stimuli (22 stimuli, each repeated 5 times), totalling 330 trials of the 66 stimuli from the three experiments.

In each trial, the subjects first saw a "+" in the centre of the screen, which would disappear quickly. Then, a stimulus sound was played automatically, and two characters were presented on the page. The character on the left side was with a non-entering tone, and the one on the right was with an entering tone. After hearing the sound, the subjects were forced to choose between the two options. If the subjects thought they heard the character with a non-entering tone, they needed to press the "F" key with their left hand; otherwise, they should press the "J" key with their right hand.

Data processing and analysis: For each trial, "F" key selection was counted as 20% and "J" key selection as 0%. Since each stimulus was repeated 5 times, the result ranged from 0% to 100%. The data were analysed with a two-way ANOVA in R [9].

Results and discussion: In the three experiments, it was found that stop coda had a significant effect on the participants' judgments: F (1,242) = 5434.92, p < .001; F (1,242) = 5173.42, p < .001; F (1,242) = 13076.38, p < .001**,** while duration had no significant effect: F (10,242) = 1.72, p > .05; F (10,242) = 1.18, p > .05; F (10,242) = 1.10, p > .05. As illustrated in Figure 1, most of the results for the non-entering tones continuum were close to 100%, and most of the results for the entering tones continuum were close to 0%, suggesting that the perception of Guangzhou Cantonese monosyllabic entering tones was made according to the stop coda, regardless of the tone height. Only a small percentage of participants' judgments were influenced by duration. For instance, the first nine stimuli for 智[tsi:33] were above 90%, and those of the tenth and eleventh stimuli were below 90%, indicating that when the duration of 智[tsi:33] is shortened, participants are more likely to perceive it as 节[tsi:t3]. Therefore, there are two findings on the perception of monosyllables with entering tone in Guangzhou Cantonese: (1) the stop coda is the distinctive feature; (2) the short duration is not a necessary feature, but it will influence perception. The findings suggested that the perception of entering tone in Guangzhou Cantonese is similar to that of Shanghai dialect [10] and Wu dialect [11], in which the stop coda is the distinctive feature. In addition, our findings provide experimental evidence supporting [2]'s suggestion that the stop coda is the distinctive feature of the entering tone perception and that the short duration is not necessary. Thus, the findings of this study contribute to a deeper understanding of the entering tones and phonological system of Cantonese.

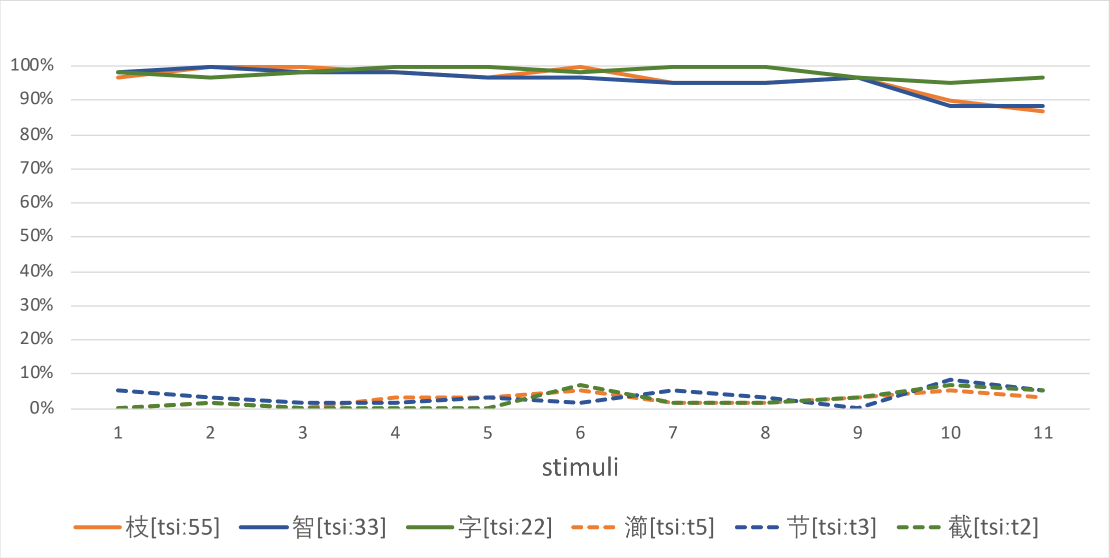


Fig.1 Means of 12 participants' judgements on stimuli.

References

[1] Liu, Y., Shi, F., Rong, R., & Sun, X. (2011). Xianggang yueyu shengdiao de fenzu fenxi [The tonal analysis of Hong Kong Cantonese in age groups]. *Yuyan Yanjiu,* (04), 98-106.

[2] Xia, Z. Y. (2006). Lun Rusheng duancu jishoucang - Rusheng lun zhi jiu [The 9th discussion of the Rusheng tone: Discussion of the feature of the Rusheng tone]. *Chengdu daxue xuebao - Sheke Ban,* (03), 88-90.

[3] Zhao, H. (1997). Qiantan hanyu Rushengyun seyinwei xiaoshi de yuanyin [A brief discussion on the reasons for the disappearance of final stop consonants in entering tone syllables in Chinese.]. *Guizhou Minzu Xueyuan Xuebao - Sheke Ban,* (02), 62-65.

[4] Zhu, X. N., Jiao, L., Yan, Z. C., & Hong, Y. (2008). Rusheng yanhua santu [Three ways of Rusheng sound change]. *Zhongguo Yuwen.* (4), 324-338.

[5] Shen, R. Q. (2015). Cong duanchang dao gaodi: Guangfupian yueyu Rusheng de shengxue xingzhi ji yanhua lujing [From short-long to high-low: The evolution of Cantonese stopped tones]. *In proceedings of the 18th International Conference on Yue Dialects.*

[6] Liu, X. Z. (2014). *Guangzhouhua danyinjie yutuce [Visible Cantonese: The spectrographic album of monosyllables of Guangzhou Cantonese].* World Publishing Corporation.

[7] Boersma, P. & Weenink, D. J. M. (2001). PRAAT, a system for doing phonetics by computer. *Glot International, 5*, 341-345.

[8] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388-407. https://doi.org/10.3758/s13428-019-01237-x

[9] R Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* URL https://www.R-project.org/.

[10]Gao, Y. F. (2004). *A study on tone perception* [Doctoral dissertation, Shanghai Normal University]. China National Knowledge Infrastructure.

[11]Yuan, D. (2015). *A study on phonetic variation in Wu dialect based on experimental analysis.* Shanghai People's Press.

The roles of talker variability, lexical frequency, and listener characteristics in second language speech perception

Donghyun Kim[1], Andrew Lee[2] & Ron Thomson[3]

[1]Kumoh National Institute of Technology (Korea), [2]Brock University (Canada), [3]Brock University (Canada)
heydonghyun@gmail.com, Andrew.Lee@brocku.ca, rthomson@brocku.ca

Accurately perceiving non-native speech sounds is a major challenge for second language (L2) learners [1,2]. Previous studies [3,4] have identified several factors affecting L2 speech perception, including cross-linguistic influence, age, length of residence, and orthographic effects. Despite this line of research, there is still a need for further investigation of factors related to auditory input. The present study aims to investigate the influence of talker variability and lexical frequency on L2 speech perception, and how these factors interact with individual differences, including working memory, L2 receptive vocabulary knowledge, and L2 proficiency.

To this end, 120 Korean learners of English participated in a series of experiments. Using 28 English words (14 minimal pairs) of varying lexical frequency, the current study targeted the /i/-/ɪ/ vowel contrast in English, which is notoriously difficult for Korean learners of English [5,6]. Each participant was assigned to one of three AX discrimination tasks (i.e., 40 participants per task). These tasks involved 336 target trials using the same 28 words, presented by 2, 6, or 12 different talkers. Participants also completed forward and backward digit span tasks to measure their working memory capacity and the Lexical Test for Advanced Learners of English (LexTALE) to measure their receptive vocabulary knowledge. Their general English proficiency was assessed using their Test of English for International Communication (TOEIC) scores. Forty-seven native speakers of English also participated in the current study as baseline participants.

Results revealed that individuals with higher working memory were better able to discriminate between the two target vowels in the 12-talker condition, suggesting that working memory helps non-native listeners cope with variability in speech. The results also indicated an interplay between working memory and L2 proficiency as well as between working memory and L2 vocabulary size in the discrimination of the non-native vowel contrast. Specifically, proficient L2 learners with high working memory were less affected by talker variability, while those with low working memory were significantly impacted by it (as shown in Figure 1). This pattern also held for L2 vocabulary size, with only those with low working memory being influenced by talker variability, especially in the 12-talker condition, regardless of their L2 vocabulary size (as shown in Figure 2). These findings suggest that individuals with higher working memory, combined with higher L2 proficiency and larger L2 vocabulary size, were able to overcome the variability caused by multiple talkers and thus discriminate between the two target vowels more effectively.

Overall, the present study highlights the importance of working memory capacity in L2 speech perception and the role it plays in overcoming the variability induced by multiple talkers. It also emphasizes the interplay between working memory, L2 proficiency, and L2 vocabulary size, showing that a combination of these factors is necessary for effective L2 speech perception. The current study has important implications for L2 speech learning and provides a useful empirical foundation for individualized L2 pronunciation training.

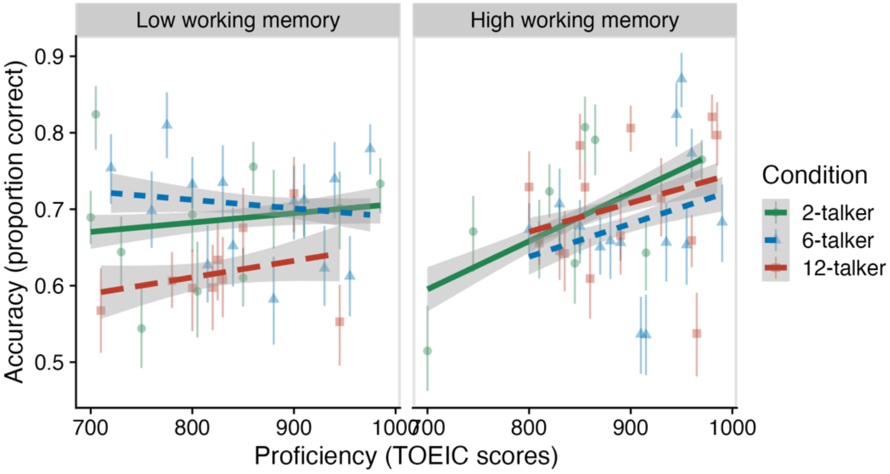


Figure 1. Discrimination accuracy by working memory and L2 proficiency

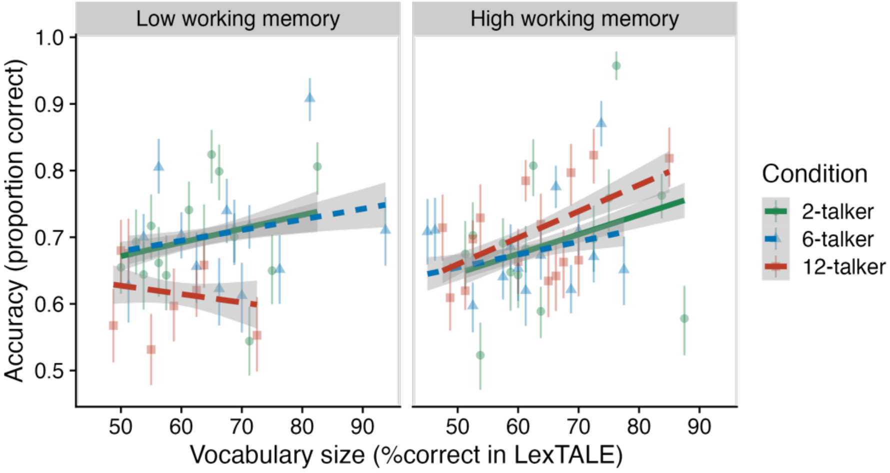


Figure 2. Discrimination accuracy by working memory and L2 vocabulary size

References

[1] Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). John Benjamins.

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

[3] Derwing, T. M., Munro, M. J., & Thomson, R. I. (2022). *The Routledge handbook of second language acquisition and speaking*. Routledge.

[4] Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). John Benjamins Publishing.

[5] Kim, D., Clayards, M., & Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics, 67*, 1–20.

[6] Lee A. H., & Lyster R. (2016) Effects of different types of corrective feedback on receptive skills in a second language: A speech perception training study. *Language Learning, 66*(4), 809–833.

Perceptual stability of sibilants undergoing acoustic variation: Interplay between acoustic processing versus influences of articulatory and/or motor patterns

Daniel Pape[1]

[1]*McMaster University (Canada)*
paped@mcmaster.ca

One of the most interesting, long debated and classical questions in phonetics/phonology is whether phoneme perception is based on apparent (surface) acoustic forms or rather underlying representations, be it motor patterns or articulatory gestures, or a combination of both. For this project, the idea was tested that manipulating the surface forms of the two underlying sibilants /s ʃ/ into their *opposite* acoustic form and spectral shape[1] would either generate perceptual identification differences for listeners (thus pointing to the importance of underlying articulatory differences) or not (pointing to the sole importance of the presented acoustic signal without taking into account underlying articulatory information during the perception process). In other words, if an underlying /ʃ/ sibilant is acoustically changed so its new spectral shape is completely identical to the spectrum of /s/, does this acoustic manipulation generate perceptual differences in dependence of whether the original *articulatory* shape was either /s/ or /ʃ/, although the perceptually judged *acoustic* stimulus is identical?

To generate the stimuli, we recorded a number of prototypical [s] and [ʃ] items in /aCa/ context from 4 female Canadian English speakers. First, stimuli were high-pass filtered above 1000Hz to exclude possible influences of irrelevant low frequency components (with respect to sibilant perception). Then we manipulated the acoustic signals (see figure 1 left panel for example) in a stepwise manner until the [s] spectral shape was acoustically completely identical to [ʃ] (i.e. amplifying frequencies stepwise below 3000 Hz, and attenuating frequencies above 3000 Hz). We used the same procedure to change the [ʃ] spectral shape to /s/ (i.e. amplifying frequencies above 3000Hz and attenuating below 3000Hz, also in a stepwise manner). As a result of these manipulations we generated a continuum of perceptual stimuli /s ʃ/ that were (1) identical in *acoustic* spectral shape distributions (see figure 1 right panels) but (2) significantly differed in the underlying original *articulatory* configuration and thus motor configurations. If these acoustically identical, but underlyingly articulatorily different stimuli would be judged as identical by listeners, then this would be evidence of purely acoustic processing of speech stimuli in speech perception. However, if these spectrally identically stimuli would be judged differently, then this would give evidence that underlying articulatory differences (apparent before manipulation and thus supposedly still extractable after manipulation) are indeed used by listeners during the perception experiment, thus pointing to more evidence that articulatory configurations and/or gestures are processed, in addition to the acoustic surface form alone.

32 Canadian English listeners participated in the experiment for course credit, their task was a forced choice identification of the presented sibilants (/s/ or /ʃ/) for each presented audio stimulus from the four speakers. The original and manipulated stimuli were either presented in isolation (C) or embedded in their original vocalic context (/aCa/)[2]. Listeners were not allowed to repeat stimuli, and 8 repetitions of each audio file was randomly presented throughout the experiment. A 10 stimuli practice session was carried out before the main experiment. We also excluded responses above 2.5 standard deviations of each listener's mean reaction time from further analysis.

Figure 2 shows the obtained results over all listeners and all four speakers (who provided the original stimuli). It can be seen that an underlying, but acoustically manipulated /ʃ/ is completely perceived as /s/, whereas the underlying (and acoustically manipulated) sibilants /s/ are not changing sibilant perception to its opposite, but rather generate chance responses. Response changes (to the opposite sibilant) are not linear but rather follow a categorical distribution, but with only underlying /ʃ/ changing completely into the opposite sibilant perception. Significance tests performed on step 7 in figure 2 showed that the perceived differences between underlying /s/ and underlying /ʃ/ are highly significant ($t(31) = -6.211, p < .001$)[3], even though their acoustic spectral shape was completely identical. We did not find perceptual differences between the 4 originally recorded speakers, and there was no effect of sibilant presentation mode (i.e. whether sibilants were presented in isolation or embedded in vowels).

[1] Thus [s] is stepwise acoustically manipulated to have a matching [ʃ] acoustic shape, and vice versa.

[2] In order to test for the effect of neighboring vowels (versus presentation in isolation) and formant transitions

[3] Step 7 (the manipulated final stimulus result) was compared to the original (opposite) sibilant stimulus perception result.

To conclude, the two fricatives /s ʃ /examined in this study behave differently in the conducted perception experiment: only the acoustically manipulated postalveolar sibilant completely changes perception to an alveolar sibilant, whereas the underlying (but also acoustically manipulated) alveolar sound resists a perceptual class change, even though the acoustic spectral shape of the manipulated /s ʃ/ sounds is almost completely identical to their sibilant counterpart. Since the acoustic shapes to be judged are almost identical but perception results show clear differences in sibilant identification our interpretation of these results is that articulatory and/or motor patterns are indeed additionally used for phoneme processing and identification in acoustically challenging conditions, in this case to help with robust perceptual identification of sibilants.

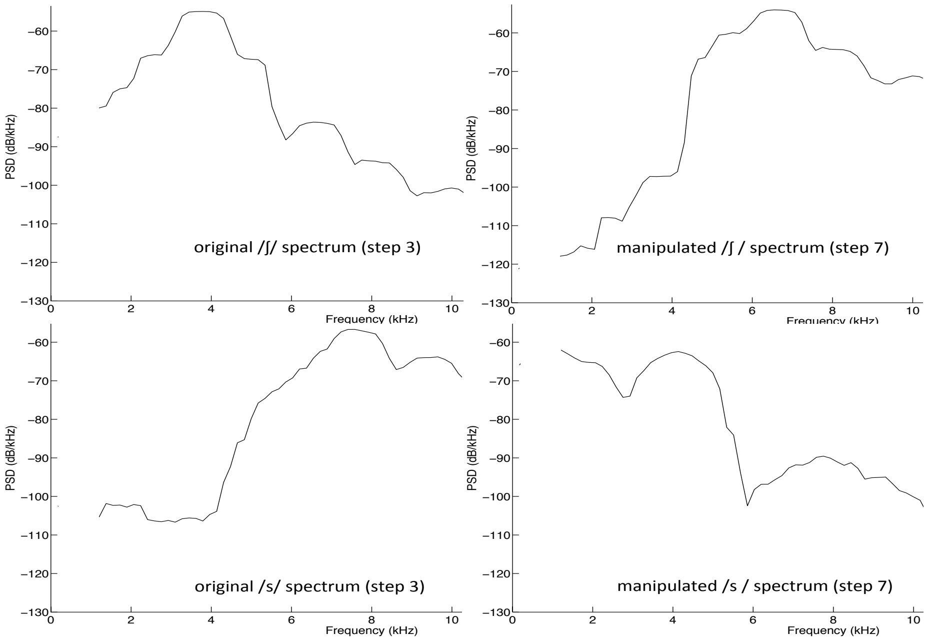


Fig. 1. Comparison between an example of the original recorded stimulus for one speaker (/ʃ/ and /s/, left panels) and the acoustically manipulated stimuli with +48dB amplification of the relevant frequencies of the opposite sibilant (right panels), i.e. *high frequency amplification for underlying / ʃ/ and low frequency amplification for underlying* /s/).

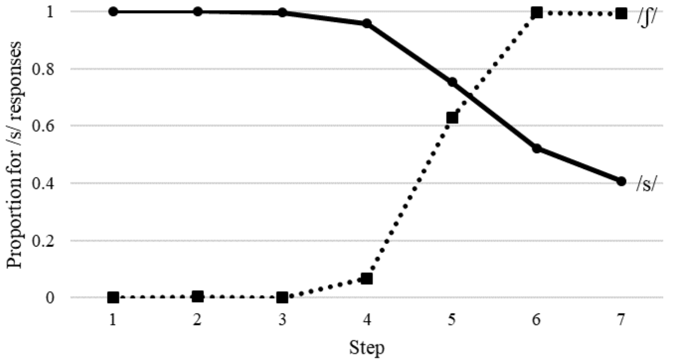


Fig. 2. Results: Probabilities of /s/ identifications (y-axis) against presented stimulus continuum (x-axis). The original recorded sibilant (/s/ or /ʃ/) is shown at step 3 (see figure 1 left panel). The stepwise increases of amplification of frequency regions relevant to the *opposite* sibilant, and thus acoustic manipulations leading to a perceptual change towards to opposite sibilant, are shown in steps 4-7. Step 7 shows the complete acoustic change where underlying /s/ would be identical in acoustic shape to /ʃ/ and vice versa (see figure 1 right panels). For completeness, steps 1-2 show amplification of *frequency regions relevant to the stimulus in question (high frequencies for /s/ and low frequencies for* /ʃ/).

Acoustic Evidence for Gestural Alignment: Vowel Devoicing in Malagasy

Jake Aziz[1]

[1]*University of California, Los Angeles (USA)*
jakeaziz@g.ucla.edu

Introduction: Vowel devoicing has been variably described as a phonological process (e.g., [1]) or as a phonetic consequence of overlapping gestures (e.g., [2] for Korean; [3] for Turkish). Research on devoicing puts itself at the center of a debate on the nature of such sound processes, as each of these accounts makes a different assumption about the role of phonetics and phonology in the grammar. In this paper, I will show acoustic data (Center of Gravity) from Merina Malagasy (Austronesian, Madagascar) that serves as evidence for the gestural account. These data are then modelled in a variant of Articulatory Phonology [4,5] that uses Alignment constraints to regulate the relative timing of gestures. This result shows that an account of vowel devoicing in Malagasy must make reference to both the articulators involved and the phonological constraints that modulate them, indicating that the phonology must have access to phonetic information.

Data: High vowels are frequently devoiced in unstressed utterance-medial syllables in the Merina dialect of Malagasy, but the precise realization and distribution of these vowels has not been investigated. Here, I present data collected from two speakers of Merina who produced a total of 319 tokens targeting unstressed /a/, /i/, and /u/ in various segmental environments. The acoustic analysis reveals that vowels in the devoicing environment may be realized as co-articulated or deleted.

Of interest to us are co-articulated vowels: these vowels are realized concurrently with the preceding consonant, typically a fricative. Acoustically, the result is extended high energy frication whose Center of Gravity reflects the underlying vowel. Compare Figure 1, which shows the spectrogram for /si/, with Figure 2, /su/: for /su/, CoG lowers, indicative of a rounding gesture associated with /u/; this is not present for /si/. In both cases, no voiced vowel is realized.

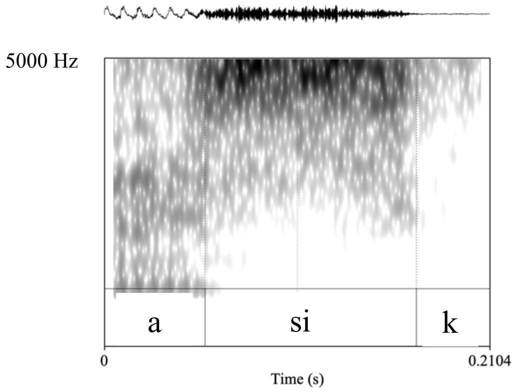


Fig.1 Co-articulated /si/

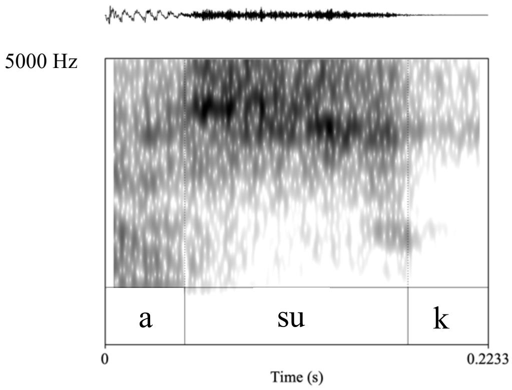


Fig.2 Co-articulated /su/

Analysis: These acoustic data can be explained by a theory of gestural overlap: in sum, before a consonant gesture ends, the vowel gesture begins, causing the observed effects on CoG of the consonant. In these cases, the vowel's glottal gesture is completely overlapped by the preceding voiceless consonant's, and thus no audible voiced vowel is observed.

I analyse the Malagasy data using Gafos's [5] variant of Articulatory Phonology [4] in which gestural overlap is regulated by language-specific constraints on the alignment of these gestures. Each gesture associated with a sound consists of five landmarks to which another sound's gestures can align, shown in Figure 3. Following Delforge's [6] work on devoicing in Andean Spanish, I use such alignment constraints to account for the Malagasy data.

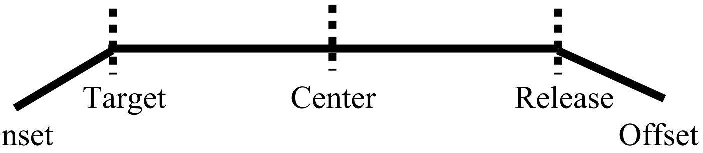


Figure 3. Alignment landmarks for a gesture, adapted from [5]

For Malagasy, the co-articulated devoiced data can be described using two constraints: The first, ALIGN (C_1, CENTER, V, ONSET) assigns a violation to any CV sequence where the onset of the vowel gesture does not coincide with the center of the consonant. In the grammar, this constraint, which favours ease of articulation, competes with a constraint ALIGN (C_1, RELEASE, V, ONSET), which favours perceptibility by aligning the vowel so that it overlaps less with the consonant. In Malagasy, a high ranking for ALIGN (C_1, CENTER, V, ONSET) would result in gestural overlap of CV sequences, including the glottal gesture, which would produce the sort of co-articulation shown in Figures 1 and 2. This is shown in Tableau 1, where underlying /sin/ results in devoicing of /i/, phonetically realised as palatalization of /s/.

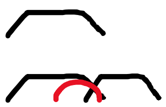
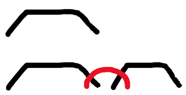
/sin/	Align (C_1, CENTER, V, ONSET)	Align (C_1, RELEASE, V, ONSET)
a. ☞ [s^{j}n] glottal  oral		*
b. [sin] glottal  oral	*!	

Tableau 1. Underlying /sin/ results in devoicing of /i/. In this tableau, candidates consist of two gestural levels for expository purposes, glottal and oral, as well as the corresponding pronunciation in IPA. On each level, consonant gestures are represented with the black angled lines, while the vowel is represented with the red curved line.

The low vowel /a/ as well as stressed vowels do not undergo devoicing. This can be explained by duration: low vowels are inherently longer than high vowels [7], and in Malagasy stressed vowels are longer than unstressed [8]. Even if the onset of the vowel occurs at the center of the preceding consonant, the vowel gesture is long enough that the overlap by the consonant is not complete, leaving a voiced portion of the vowel. In the remainder of the analysis, these Alignment constraints are used similarly to account for vowel deletion that occurs after some sonorants, showing that vowel devoicing and deletion can be uniformly described as one articulatory outcome (overlap), but acoustically, this is realized differently in different segmental environments.

Discussion: Here, I've demonstrated that many so-called devoiced vowels in Malagasy are realized as co-articulated with the preceding consonant; this realization lends support to an account of gestural overlap as an explanation for devoicing, and I show that specific Alignment constraints in an Articulatory Phonology framework neatly account for the acoustic data. This result is theoretically consequential as it indicates that the phonological grammar has access to information about the articulators. In sum, processes like devoicing in Malagasy cannot be described as purely phonetic or phonological, but must take into account both.

References:

[1] Vogel, R. (2022). *Phonology of vowel devoicing: A typological perspective*.

[2] Jun, S. A., Beckman, M., Niimi, S., & Tiede, M. (1997). Electromyographic evidence for a gestural-overlap analysis of vowel devoicing in Korean. *Speech sciences*, *1*, 153-200.

[3] Jannedy, S. (1995). Gestural phasing as an explanation for vowel devoicing in Turkish.

[4] Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *3*, 219-252.

[5] Gafos, A. I. (2002). A grammar of gestural coordination. *Natural language & linguistic theory*, 269-337.

[6] Delforge, A. M. (2008). Gestural alignment constraints and unstressed vowel devoicing in Andean Spanish. In *Proceedings of the 26th west coast conference on formal linguistics* (pp. 147-155). Somerville, MA: Cascadilla Proceedings Project.

[7] Lehiste, I. (1970). Suprasegmentals.

[8] Howe, P. (2019). Central Malagasy. *Journal of the International Phonetic Association*, *51*(1), 103-136.

A Preliminary Study about Disappearing Laryngeal and Supralaryngeal Articulatory Distinction of the Three-way Contrast of Korean Velar Stops

Jiyeon Song[1], Sahyang Kim[2] & Taehong Cho[1]

[1]*Hanyang Institute for Phonetics & Cognitive Sciences of Language, Hanyang University (Korea),*
[2]*Hongik University (Korea)*
jiyeoni00@hanyang.ac.kr, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Classifying stops in many languages often involves two common phonetic features: voicing and aspiration [9]. However, Korean exhibits a typologically rare three-way contrast of word-initial voiceless stops, which are typically categorized as lenis, fortis, and aspirated stops based on their laryngeal characteristics. Previous phonetic studies have described that the lenis stops have a breathy quality and slightly aspirated; the fortis stops are tense, laryngealized, and unaspirated; and the aspirated stops are strongly aspirated [4]. While the three-way contrast was traditionally considered to be primarily reflected in VOT (longest for the aspirated, shortest for the fortis, and intermediate for the lenis stops), recent studies have shown that the VOT distinction between lenis and aspirated stops has been lost, leading to a VOT merger of the two categories, especially in the Seoul dialect of Korean [1,5,7,11]. Furthermore, it has been suggested that females are in the vanguard of this merger [6]. Given the VOT merger underway, a significant question arises regarding whether and to what extent the kinematic properties of the three-way stop contrast may have changed along with VOT, potentially reflecting the effects of the sound change on supralaryngeal articulation. The position of the tongue, for example, might be adjusted based on the articulatory closure duration, which tends to be inversely proportional to VOT when VOT is defined in terms of articulation [3], as pointed out by [10]. In the case of the lenis-aspirated distinction, if the VOT values for these sounds become similar, it might indicate that the differential hold duration-related impacts on articulation would also no longer exist between the two sounds. In other words, the three-way supralaryngeal distinction, which was previously present, would also be lost along with the VOT merger.

Twenty-two Korean speakers in their 20s from Seoul and Gyeonggi-do participated in the experiment as part of a large acoustic and articulatory database of Korean which is in construction at the Hanyang Institute for Phonetics and Cognitive Science of Language [12]. The articulatory data were collected by using the 3D electromagnetic articulography (EMA, AG501, Carstens Electronics). The participants were instructed to produce a randomly ordered sequence of eighteen syllables in the form of /Ca/. Only the velar stops in /ka/, /k^{h}a/, and /k*a/ were selected for examination in this study. The articulatory data from the tongue dorsum sensor have been examined to track the closing and opening movement of tongue dorsum for velar closure. A total of 132 tokens (3 velars x 22 speakers x 2 repetition) are included for the analysis.

In the acoustic data, the results of VOT in the present study showed that the lenis-aspirated contrast is lost in females (/k^h/=/k/>/k*/) while the three-way stop contrast still remains in males (/k^h/>/k/>/k*/). As for the results of articulatory gestures, there was no three-way distinction in closure duration for either females or males: females showed a binary distinction between the lenis and the other two stops (/k/=/k^h/</k*/), while males showed a difference between the lenis and fortis stops only (/k/</k*/). The results for peak velocity of the tongue dorsum opening movement (see Peak Velocity 2 in Fig.2) indicated that in males, there was a three-way contrast between the velar stops (/k^h/</k/</k*/), while in females, there was only a distinction between the aspirated and fortis sounds (/k^h/</k*/). In the case of deceleration duration of the tongue dorsum opening movement, females exhibited a binary contrast, with no differentiation between the aspirated and fortis stops (/k^h/=/k*/</k/), while males showed a binary distinction, with no distinction between the lenis and aspirated stops (/k/=/k^h/</k*/). The findings from opening movement duration showed that for female speakers, there was a difference only between the lenis and fortis stops (/k*/</k/). However, for male speakers, there was a difference between the fortis and lenis/aspirated stops but not between the lenis and aspirated stops (/k*/</k/=/k^h/).

As seen in Figure 1 and 2, the findings from males indicated an inverse association between VOT (laryngeal aspect) and Peak Velocity of the tongue dorsum opening movement (supralaryngeal aspect): the aspirated stop showed the longest VOT and slowest Peak Velocity whereas the fortis stop showed the shortest VOT and the fastest Peak Velocity during the opening movement. Bernoulli's principle, which is a fundamental concept in fluid dynamics, can be used to explain the relationship between VOT and Peak Velocity. According to Bernoulli's principle, when air flows through the wider end of the straw, the velocity of the air decreases, causing the pressure to increase. When air flows through a narrower end of the straw, the velocity of the air increases, causing the pressure to decrease. This pressure difference causes the air to move faster through the narrower end of the straw, and slower through the wider end. Consequently, the fortis stop with the fastest peak velocity of the tongue dorsum during the opening movement can reach a wide opening at the contact area of the tongue dorsum and the palate most rapidly, causing slower air movement and the shortest VOT. On the other hand, the aspirated stop with the slowest peak velocity of the tongue dorsum opening movement can keep a narrow channel at the contact area for the longest duration, causing faster air movement and the longest VOT.

The findings of the current study provide some evidence of disappearing laryngeal and supralaryngeal articulatory distinction of the three-way stop contrast in velars along with the on-going sound change related to the VOT merger. Moreover, the findings also support that the lenis-aspirated distinction is disappearing and that the change is led by females.

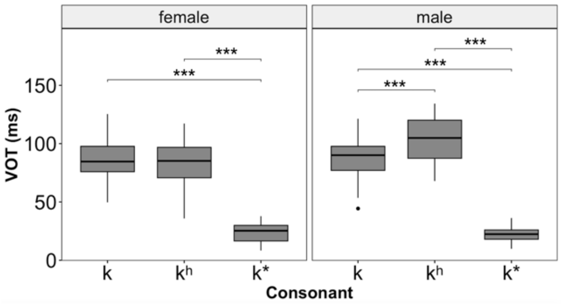


Fig.1 VOT for the three velar stops. Error bars indicate standard errors.

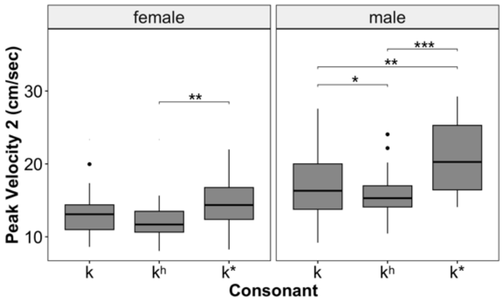


Fig.2 Peak velocity of the tongue dorsum opening movement for the three velar stops. Error bars indicate standard errors.

References

[1] 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.

[2] Cho, T. 1996. Vowel correlates to consonant phonation: an acoustic-perceptual study of Korean obstruents. MA thesis, University of Texas at Arlington.

[3] Cho, T., & Ladefoged, P. 1999. Variation and universals in VOT: evidence from 18 languages. Journal of phonetics, 27(2), 207-229.

[4] Cho, T., Jun, S. A., & Ladefoged, P. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. Journal of phonetics, 30(2), 193-228.

[5] Choi, J., Kim, S., & Cho, T. 2020. An apparent-time study of an ongoing sound changes in Seoul Korean: A prosodic account. Plos one, 15(10).

[6] 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.

[7] Kim, M. R. 2013. Interspeaker variation on VOT merger and shortening in Seoul Korean. In Proceedings of Meetings on Acoustics ICA2013 (Vol. 19, No. 1, p. 060212). Acoustical Society of America.

[8] Ladefoged, P., & Cho, T. 2001. Linking linguistic contrasts to reality: The case of VOT. TravauxDu CercleLinguistiqueDe Copenhague, 31.

[9] Lisker, L., & Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20(3), 384-422.

[10] Maddieson, I. 1997. Phonetic Universals. In 1he handbook of phonetic sciences (J. Laver & W. J. Hardcastle, editors), pp. 619-639. Oxford: Blackwells.

[11] Silva, D. J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. Phonology, 23(2), 287-308.

[12] The Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS). 2022, Dynamics of Speech Production through Articulatory DB Construction.

Pitch Accent in North Kyungsang Korean Spoken Word Recognition

Hyun-ju Kim

The State University of New York, Korea

hjkim@sunykorea.ac.kr

North Kyungsang Korean (NKK) is a pitch accent language where the lexical accent pattern of a word is lexically determined as illustrated in the following minimal triple *káci* ‘kind’, *kací* ‘eggplant’, *kácí* ‘branch’. The role of pitch accent has been understudied for NKK word recognition, while it has been known more for Japanese word recognition (e.g., Cutler and Otake 1999; Katsuda and Steffman 2022), suggesting that pitch accent plays a role in restricting word activation. This study investigates a role of pitch accent in spoken word recognition in NKK, addressing whether pitch-accent information can be used in the lexical access process in NKK. The results revealed that NKK tonal cues facilitate the lexical access with matching segment primes but rather impede it if tone patterns mismatch even with matching segment primes, suggesting that tone mismatch is as detrimental as segment mismatch in NKK word recognition.

Experimental studies were designed to replicate and extend Katsuda and Steffman(2022)’s study to NKK by employing a lexical decision task and priming paradigm but with some modification to make it suitable for NKK lexical activation. In order to test the priming effect of pitch accent in the lexical decision, four prime types (an identity prime, a segment prime, a pitch accent prime, and a control) were included in the study by using 76 bisyllabic word targets and the same number of bisyllabic non-word targets. Two age groups of NKK listeners (12 older; 12 younger) participated in the experiments. In a lexical decision task, listeners judged whether a word presented on the screen is a real word or non-sense word after listening to initial fragments of primes. Response time (in ms.) for correct responses to word target trials was measured, and z-scored response times were analyzed by running linear mixed effects models in R (Baayen et al. 2008) to examine the main effects of age group, prime type and tone type. The dependent variable was z-scored response time.

Results showed that response times(RT) were faster by previous presentation of the identical syllable with the same tone pattern (an identity prime) (1175 ms. in the older group; 724 ms. in the younger group) than either by the identical syllable with a different tone pattern (a segment prime) or by the matching tone pattern but with a different segment (a pitch accent prime) (1211 ms. vs. 1221 ms. in the older group; 788ms. vs. 764 ms.) in both groups, as shown Table 1, although older NKK listeners’ responses were slower than younger NKK responses in general. As presented in Table 2, there was significant interaction between prime type and tone type of the target words (*p*=.02) where the negative estimate indicates faster RT for pitch accent primes in LH type target words, although the main effect of prime type did not reach statistical significance. The interaction between prime type and age group and tone type was also significant (*p*=.0008), indicating that young NKK listeners were slower in LH type target words by pitch accent primes than older NKK listeners. In other words, pitch accent priming effects were stronger for older NKK listeners in LH type words than younger NKK listeners.

This study revealed that lexical decision was facilitated by segment primes only when matching in pitch accent but impeded by segment primes mismatching in pitch accent, implying NKK listeners exploit prosodic information for their lexical access. Furthermore, the significant interaction between prime type and tone type indicates that NKK listeners actively tap into tonal cues as well as to lexical tone types when it comes to lexical decision. These findings suggest that pitch accent information constrains lexical activation in the process of spoken-word recognition by NKK listeners.

Table 1. Mean Response Time by Age Group and Prime Type

Group	Prime_Type	RT (ms.)	S.D.
Old	Control	1259.41	387.63
	Identity	1175.72	299.27
	PitchAccent	1221.08	361.20
	Segment	1211.07	307.43
Young	Control	756.58	281.72
	Identity	724.47	297.68
	PitchAccent	764.41	300.56
	Segment	788.06	405.68

Table 2. Results of the Linear Mixed Effects Model for RT

Fixed effects	Estimate	Std. Error	df	*t* value	*P*
(Intercept)	-2.20E-01	1.50E-01	3.01E+02	-1.463	0.144528
Prime_TypeIdentity	-1.39E-01	1.68E-01	1.08E+03	-0.826	0.409053
Prime_TypePitchAccent	4.99E-02	1.68E-01	1.08E+03	0.298	0.765957
Prime_TypeSegment	-1.58E-01	1.70E-01	1.08E+03	-0.931	0.352116
GroupExp_Y	8.00E-02	1.66E-01	1.08E+03	0.481	0.630598
ToneL	1.74E-01	2.12E-01	2.98E+02	0.823	0.410932
Prime_TypeIdentity:GroupExp_Y	-1.77E-01	2.37E-01	1.08E+03	-0.748	0.454439
Prime_TypePitchAccent:GroupExp_Y	-3.29E-01	2.35E-01	1.08E+03	-1.4	0.161742
Prime_TypeSegment:GroupExp_Y	-8.89E-02	2.38E-01	1.08E+03	-0.374	0.708681
Prime_TypeIdentity:ToneL	-3.65E-01	2.37E-01	1.08E+03	-1.541	0.123708
Prime_TypePitchAccent:ToneL	-5.36E-01	2.36E-01	1.08E+03	-2.269	0.023487*
Prime_TypeSegment:ToneL	-8.32E-02	2.41E-01	1.08E+03	-0.346	0.729677
GroupExp_Y:ToneL	-3.95E-01	2.35E-01	1.08E+03	-1.683	0.09264.
Prime_TypeIdentity:GroupExp_Y:ToneL	6.32E-01	3.35E-01	1.08E+03	1.888	0.059285.
Prime_TypePitchAccent:GroupExp_Y:ToneL	1.11E+00	3.32E-01	1.08E+03	3.342	0.000861***
Prime_TypeSegment:GroupExp_Y:ToneL	5.46E-01	3.37E-01	1.08E+03	1.621	0.105203

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

References

[1] Cutler, A. & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese, *The Journal of Acoustic Society of America* 105(3), 1877-1888.

[2] Katsuda, H. & Steffman, J. (2022). Asymmetrical roles of segment and pitch accent in Japanese spoken word recognition, *JASA Express Letters* 2, 065201

[3] Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items, *Journal of Memory and Language* 59(4), 390-412.

Effect of listeners' linguistic experience on generalization of adaptation

Dae-yong Lee[1] and Melissa Baese-Berk[2]
[1]Hanyang Institute for Phonetics and Cognitive Sciences of Language, Hanyang University
[2]Department of Linguistics, University of Oregon
daeyonglee@hanyang.ac.kr, mbaesebe@uoregon.edu

Listeners often have difficulty understanding non-native speech. While understanding non-native speech may initially be challenging for listeners, listeners may become better at understanding non-native speech after short training sessions (i.e., adaptation; [1]) and better understand a novel non-native speaker from the same language background (i.e., generalization; [2]). Most studies on generalization of adaptation focus on how speaker characteristics affect generalization of adaptation [1,3]. Studies that examine the effect of speaker characteristics on generalization of adaptation tend to control for listeners' linguistic experience by recruiting listeners who do not have extended linguistic experience with target languages [2,3].

However, listeners' linguistic experience with non-native speakers is likely to affect generalization of adaptation. Specifically, short training sessions in the lab facilitate adaptation to non-native speech [1,2] and generalization to a novel speaker [1,2]. Further, sleep between training sessions facilitates generalization of adaptation to a novel speaker [4]. These results suggest that extended linguistic experience with non-native speakers may affect generalization of adaptation. However, it is less understood how listeners' linguistic experience with non-native speakers affects generalization of adaptation.

Thus, the current study examines whether extended linguistic experience with non-native English speakers affects generalization of adaptation to a novel speaker. Specifically, the current study asks whether listeners' experience with non-native English speakers facilitates generalization and whether different types of experiences have different effects on generalization. With regards to the effect of linguistic experience with non-native English speakers on generalization, it is possible that linguistic experience facilitates generalization to a novel non-native speaker. That is, if short training sessions in the lab and sleep between training sessions help generalization, extended experience with non-native speakers may facilitate generalization. On the other hand, linguistic experience may disrupt generalization. Specifically, listeners who have a lifetime of experience with non-native speakers may have a less malleable representation of non-native speech than listeners who do not have linguistic experience with non-native speakers. With regards to the effect of types of linguistic experience with non-native English speakers, two outcomes are possible. It is possible that linguistic experience with multiple non-native accents is more helpful for generalization than experience with a single non-native accent. Previous studies suggest that exposure to multiple non-native English speakers helps listeners learn the common characteristics of non-native speech and facilitates generalization to a novel speaker [5]. Similarly, it is possible that extended linguistic experience with multiple non-native accents facilitates generalization. Another possibility is that exposure to a single non-native accent and multiple non-native accents have similar effects on generalization of adaptation. That is, extended experience with a single non-native accent may provide enough variability to learn the characteristics of non-native speakers.

75 native English speakers between 18 and 40 years old participated in this study. Participants completed a language experience questionnaire to determine the participants' linguistic experience. Based on the participants' linguistic experience, participants were assigned to one of three linguistic experience conditions: 1) Multiple-accent Exposure, 2) Single-accent Exposure, and 3) No Exposure conditions. Participants were assigned to Multiple-accent Exposure condition if participants had frequent interaction with family members that were non-native English speakers and with non-native English speakers in elementary and high school. Participants were assigned to the Single-accent Exposure condition if participants interacted frequently with family members that were non-native

English speakers (i.e., Spanish learners of English) and did not frequently interact with non-native English speakers other than Spanish learners of English in elementary and high school. Participants were assigned to the No Exposure condition if participants did not have family members that were non-native English speakers, had limited or no interaction with non-native English speakers in elementary and high school, and did not have frequent interaction with non-native English speakers over the past year. Participants in the three conditions were asked to listen to English sentences read by Korean learners of English and transcribe what they heard (i.e., intelligibility task). The task consisted of a training session and a post-test. Participants' performance (i.e., percent correct) in the training session and post-test was scored to measure adaptation to non-native English speakers and generalization to a novel non-native English speaker, respectively.

Figure 1 shows participants' performance in the training session of the intelligibility task. As shown in Figure 1, participants in the No Exposure, Single-accent Exposure, and Multiple-accent Exposure conditions show improvements in intelligibility across the training session, suggesting that participants adapted to the non-native English speakers in the training session. Figure 2 shows participants' performance in the post-test of the intelligibility task. As shown in Figure 2, participants in the Single-accent Exposure (box in the middle) and Multiple-accent Exposure (box on the right) conditions as a group demonstrate lower intelligibility scores than participants in the No Exposure condition (box on the left). Further, participants in the Single-accent Exposure and Multiple-accent Exposure conditions demonstrate similar intelligibility scores. These results suggest that extended experience with non-native English speakers may disrupt generalization to a novel non-native English speaker and extended linguistic experience may be harmful for generalization regardless of the type of linguistic experience. Specifically, listeners' representation of non-native speakers may be less malleable for listeners who have extended linguistic experience with non-native English speakers than listeners who have limited experience with non-native English speakers.

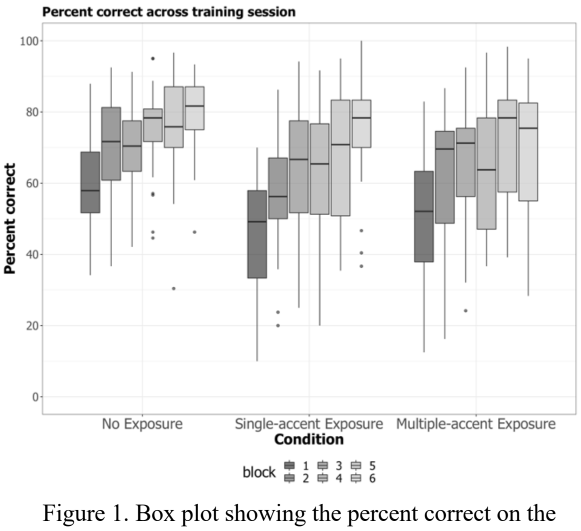


Figure 1. Box plot showing the percent correct on the training session of the intelligibility task as a function of condition and block.

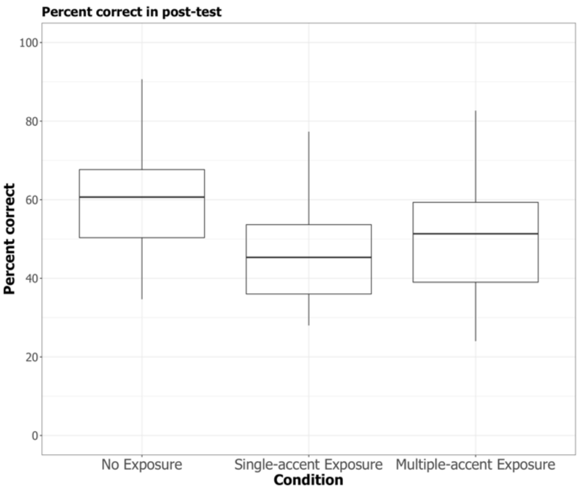


Figure 2. Box plot showing the percent correct on the post-test of the intelligibility task as a function of condition.

References

[1] Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729.

[2] Sidaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America, 125*(5), 3306–3316.

[3] Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174–EL180.

[4] Xie, X., Earle, F. S., & Myers, E. B. (2018). Sleep facilitates generalisation of accent adaptation to a new talker. *Language, Cognition and Neuroscience*, *33*(2), 196-210.

[5] Laturnus, R. (2020). Comparative Acoustic Analyses of L2 English: The Search for Systematic Variation. *Phonetica*, *77*(6), 441-479.

Is Southern Min Tone Circle a Real Thing?

Yishan Huang

University of Sydney (Australia)
yishan.huang@sydney.edu.au

Southern Min, a major Sinitic dialect of the Sino-Tibetan language family, has been best known for its intricate tone sandhi phenomenon. One tone's citation form is conventionally assumed to the sandhi form of another tone, rendering a circular tone change among those unstopped tones that are associated with sonorant-ending syllables. This phenomenon has attracted extensive theoretical discussions to explore and explain why Southern Min tones change in a circular way. The discussions are from various perspectives including Generative Phonology [1-2], Autosegmental Phonology [3-4], Optimality Theory [5-9], Psycholinguistic Theory [10-12], and Lexicalised Phrasal Phonology [13]. However, it turns out to be challenging to use both generative and optimality theories to capture and interpret the naturalness of the process, environment, and mechanism that motivates tones to change in a circular fashion in Southern Min.

This study asserts that the nature of Southern Min tones is morphological, and the relation between sandhi and citation tones is morpho-phonemic. This assertation is proposed based on a systematic examination of multidimensional tonal realisations (F0, duration, vowel quality, voice quality, and obstruent coda) in three linguistic contexts (citation, phrase-initial, and phrase-final) from 21 native speakers in Zhangzhou Southern Min [14]. The phonetically-statistically-grounded results (Fig.1) show that, (1) The relations of phrase-initial (sandhi) tones are entirely unrelated to those of their corresponding citation forms both phonologically and phonetically; it is thus appropriate to consider the sandhi tones and citation tones are in a morphophonemic relation, belonging to two independent systems. (2) The realisations of most phrase-final tones are categorically related to those of their corresponding citation forms with a certain degree of phonetically predictable variations; it is thus appropriate to consider the phrase-final tones and citation tones are in an allophonic relation, belonging to the same system. (3) Tonal contrast neutralisation occurs across citation, phrase-initial (sandhi), and phrase-final contexts, the direction of tonal alternation is essentially indeterminate. Determining which forms are underlying and which forms are derived is difficult. It is thus appropriate to consider that tones at the citation and sandhi contexts are not in a derivational relation, but rather belonging to two independent systems.

Incorporating the three important factors, this study asserts that tonal realisations in Zhangzhou are morphologically motivated. Each lexical tone functions as a single morpheme with alternating allomorphs (tonemes) that are both abstractly stored in the mental grammar of native speakers but phonetically distant on the surface (Fig. 2). This is analogous to the nature of plural morpheme in English that is sometimes pronounced as [s] (as in cats [kæts]), sometimes as [z] (as in dogs [dɒgz]), and sometimes as [-əz] (as in faces [feisəz]) [15], which have the same meaning but occur in different environments and in complementary distribution. It also reflects a close interface between different linguistic levels (phonetics, phonology, and morpho-syntax) to realise tone as an important language phenomenon in Asian languages. This study substantially stretches and advanced our knowledge of the nature of tone sandhi in Southern Min, shedding an important light on how a sophisticated examination of phonetic detail contributes to uncover and establish the cognitive pattern in natural languages.

Keywords: Tone sandhi, acoustics; morphological nature; linguistic interface, Southern Min

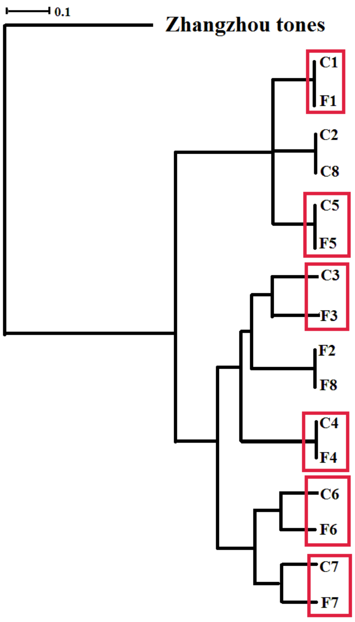


Fig. 1 Mapping of tonal realisations across different contexts (C=citation; F=Phrase-final; I=Phrase-initial).

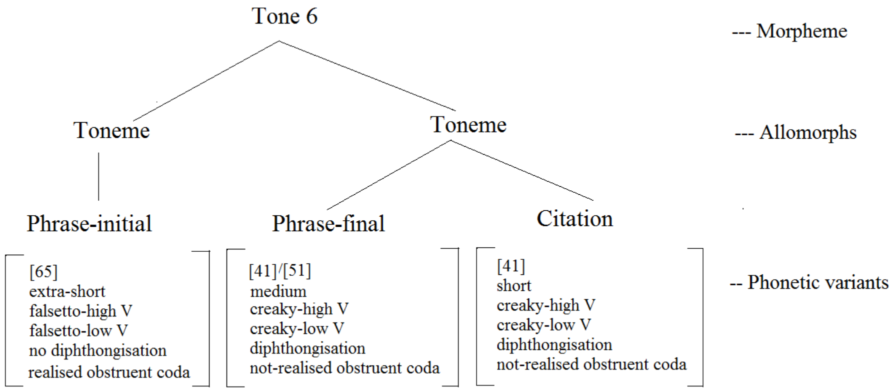


Fig. 2 Illustration of morphologically-conditioned tonal realisations in Zhangzhou Southern Min.

References

[1] Wang, W. S.-Y. (1967). Phonological features of tone. *International Journal of American Linguistics*, 33, 93-105.

[2] Cheng, R. L. (1968). Tone sandhi in Taiwanese. *Linguistics*, 6(41), 19-42.

[3] Yip, M. (1980). *The tonal phonology of Chinese*. Doctoral dissertation: Massachusetts Institute of Technology.

[4] Shih, C.-L. (1986). *The prosodic domain of tone sandhi in Chinese*. Doctoral thesis: University of California at San Diego.

[5] Yip, M. (2002). *Tone.* Cambridge, England: Cambridge University Press.

[6] Mortensen, D. (2002). Semper infidelis: Theoretical dimensions of tone sandhi chains in Jingpho and A-Hmao. *Unpublished manuscript*, University of California, Berkeley.

[7] Hsieh, F.-f. (2005). Tonal chain-shifts as anti-neutralization-induced tone sandhi. *University of Pennsylvania Working Papers in Linguistics*, 11(1), 99-112.

[8] Barrie, M. (2006). Tone circle and contrast preservation. *Linguistic Inquiry*, 37, 131-141.

[9] Thomas, G. (2008). An analysis of Xiamen tone circle. *Proceedings of the 27th West Coast Conference on Formal* Linguistics, Cascadilla Proceedings Project, 422-430,

[10] Hsieh, H.-I. (1976). On the unreality of some phonological rule. *Lingua*, 38(1), 1-19.

[11] Wang, S. (1992). An experimental study on the productivity of Taiwanese tone sandhi. *In Proceedings of the third International Symposium on Langages and Linguistics,* Bankok, Thailand, 116-129.

[12] Chen, S.-w., Myers, J., & Tsay, J. (2010). Testing the allomorph selection hypothesis in Taiwanese tone sandhi. *In Proceedings of the 22nd North American Conference on Chinese Linguistics & The 8th International Conference on Chinese Linguistics,* 283-300.

[13] Tsay, J., & Myers, J. (1996). Taiwanese tone sandhi as allomorph selection. *In Proceedings of the Twenty-Second Annual Meeting of the Berkeley Linguistic Society*, 395-405.

[14] Huang, Y. (2018). *Tones in Zhangzhou: Pitch and beyond*. Doctoral Thesis: The Australian National University.

[15] Haspelmath, M. (2002). *Understanding morphology*. London, England: Arnold.

The effects of ultrasound biofeedback on vowel acoustics

Ching-Hung Lai[1,2] & Chenhao Chiu[2]

[1]National Cheng Kung University (Taiwan), [2]National Taiwan University (Taiwan)
i54051092@gs.ncku.edu.tw, chenhaochiu@ntu.edu.tw

Ultrasound biofeedback has been applied in second language (L2) pronunciation teaching for decades, with many studies showed positive impacts on L2 pronunciation [1, 2, 3]. However, several studies provided mixed results from different training consonants [3, 4] and vowels [5, 6]. Insofar, the effects of ultrasound biofeedback on vowel acoustics remains unexplored. The current study aims to tackle this issue and to investigate how vowel dimensions influence the training effects.

Twenty-eight native speakers of Taiwan Mandarin (12 males, mean: 23.3 year old) who have not learnt the training vowels prior to the experiment were recruited. The training vowels were Cantonese /ɐ/ and Japanese /ɯ/, designated for vowel height and vowel frontness trainings, respectively. Participants received a 20-minute ultrasound biofeedback training session for each training vowel. During the session, participants would see real-time ultrasound tongue images with a superposed target contour. When hearing a syllable, they were asked to mimic the sound and, in the meantime, move their tongue toward the target. There were 20 blocks with 15 syllables each, yield 300 trials for one training session. Pre-training and post-training tests were administered to assess the training effects. The ultrasound tongue images and acoustic signals were recorded simultaneously, and a 3D-printed transducer stabilizer [7] was applied throughout the experiment.

To quantify the training effect on the acoustics performances. Mahalanobis distances (MD) between the target sound and individual tokens in F1 × F2 vowel space were calculated, serving as an index of accuracy. Larger MD differences between pre- and post-training suggest stronger improvement in *accuracy* after training. Meanwhile, the area of ellipse (AE) enclosing a 95% CI from the tokens in F1 × F2 vowel space delineates production variability. Larger AE differences between pre- and post-training indicate stronger improvement in *precision* after training. One-tail *t* test was applied to see if there is a positive training effect and paired-*t* test was employed to see if the training effects from different vowel dimensions are significantly different. Besides, to get the number of participants who improved, if the difference reached 10% of the pre-training test, it would be seen as an individual improvement.

Several findings were obtained from the results. While our articulatory results suggest that learning the difference in the vowel height dimension is easier than that in the vowel frontness dimension, the acoustic results showed the absence of accuracy improvements between pre- and post-training (vowel height: 5.28, p = .481; vowel frontness: -5.08, p = .115) despite a significant difference between the two (p < .05). Second, the trends of individual improvements from the vowel height training (Table 1) were similar between articulatory and acoustic data, where most people improved only in accuracy not in precision. On the other hand, the trends of individual improvements from the vowel frontness training (Table 2) revealed that participants improved either both accuracy and precision or none in terms of articulation while they mostly improved only in precision in terms of acoustics. Collectively, our results provide supportive evidence that different vowel dimensions may induce different biofeedback training outcomes. These results could further offer suggestions for customized training in both L2 pronunciation pedagogy and language therapy.

Table 1. The numbers of participants and ratios (in parentheses) showing accuracy and precision improvements during the vowel height training

Acoustics	Improved *accuracy*	No improved *accuracy*	Total
Improved *precision*	4 (14.3%)	6 (21.4%)	10 (35.7%)
No improved *precision*	14 (50.0%)	4 (14.3%)	18 (64.3%)
Total	18 (64.3%)	10 (35.7%)	28 (100.0%)

Table 2. The numbers of participants and ratios (in parentheses) showing accuracy and precision improvements during the vowel frontness training

Acoustics	Improved *accuracy*	No improved *accuracy*	Total
Improved *precision*	2 (7.1%)	15 (53.6%)	17 (60.7%)
No improved *precision*	4 (14.3%)	7 (25.0%)	11 (39.3%)
Total	6 (21.4%)	22 (78.6%)	28 (100.0%)

References

[1] Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I., & Zampini, M. (2008). Ultrasound imag- ing applications in second language acquisition. *Phonology and second language acquisition*, 36(6), 309–322.

[2] Pillot-Loiseau, C., Antolík, T. K., & Kamiyama, T. (2013). Contribution of ultrasound visualisation to improving the production of the French /y/-/u/ contrast by four Japanese learners. In *PPLC13: Phonetics, phonology, languages in contac: varieties, multilingualism, second language learning.*

[3] Tsui, H. M.-L. (2012). *Ultrasound speech training for Japanese adults learning English as a second language*. Ph.D. thesis, University of British Columbia Vancouver, BC, Canada.

[4] Tateishi, M. (2013). *Effects of the Use of Ultrasound in Production Training on the Perception of English /ɹ/ and /l/ by Native Japanese Speakers*. Master's thesis, Graduate Studies.

[5] d'Apolito, I. S., Sisinni, B., Grimaldi, M., & Fivela, B. G. (2017). Perceptual and ultrasound articulatory training effects on English L2 vowels production by Italian learners. *International Journal of Cognitive and Language Sciences*, *11*(8), 2174– 2181.

[6] Li, J. J., Ayala, S., Harel, D., Shiller, D. M., & McAllister, T. (2019). Individual predictors of response to biofeedback training for second-language production. *The Journal of the Acoustical Society of America*, 146(6), 4625-4643.

[7] Derrick, D., Carignan, C., Chen, W.-r., Shujau, M., & Best, C. T. (2018). Three-dimensional printable ultrasound transducer stabilization system. *The Journal of the Acoustical Society of America*, *144*(5), EL392–EL398.

Cross-linguistic Influences among L1, L2, and L3 Monophthongs by Cantonese Speakers in the Multilingual Context

Chen Hsueh Chu[1], Tian Jing Xuan[2]

[1]The Education University of Hong Kong University (Hong Kong), [2]The Education University of Hong Kong University (Hong Kong)
hsuehchu@eduhk.hk, s1126315@s.eduhk.hk

Third language (L3) phonological acquisition is a complex process (e.g., Chen & Han, 2019; Kellerman, 1983). Language learners in Hong Kong (HK) and Guangdong province (GD) in Mainland China usually acquire three languages Cantonese, as their first language (L1), Putonghua (Mandarin, hereafter) as their second language (L2), the official language, and English as their L3, the foreign language that needs to be learned. However, the language backgrounds of Cantonese speakers in these two areas are different according to the historical issue. For HK Cantonese speakers, English is the L2, and Mandarin is the L3.

In the revised Motor theory (Liberman & Mattingly, 1985), listeners interpret the speaker's intended movements to produce specific phonetic features, such as tongue backing. These intended movements can be seen as abstract control units which would control the production of the phonetic feature. Empirical evidence supports this theory and identifies cross-linguistic influences (CLIs). When the target phonetic features are distinct acoustically or most similar acoustically to sounds belonging to different phonemic categories in the listener's L1, they are easy to learn (Pickett, 1999, p. 213). In L3 acquisition, the acquisition process is complex because of the interactions among the three languages that learners learn and use. L3 acquisition models claim that features of all the languages that learners learned could be transferred to the target language (e.g., Westergaard et al., 2017). Vowel systems of Cantonese, Mandarin, and English are different. According to previous studies (Zee, 1991; Lee & Zee, 2003; Roach, 2009), there are 6 monophthongs in Mandarin, 11 in Cantonese, and 12 in English. Previous studies reported that L3 could be affected by several factors, such as L1 status (e.g., Ringbom, 1987), L2 status (e.g., Wrembel, 2010), and proficiency levels of the L2 and L3 of the learners (e.g., Chen & Han, 2019). The different language situations in HK and GD may cause different CLIs. HK speakers' self-reflection confirmed both progressive and regressive CLIs (Chen & Han, 2019); however, only progressive CLIs were identified by GD speakers (Chen & Tian, 2021). These CLIs were mainly summarized from learners' self-reflections on L2 and L3 learning experiences. Few studies used acoustic data to support this. This study aims to use acoustic data to investigate the possible CLIs on monophthongs by Cantonese speakers.

Participants of this study include Cantonese speakers (CSs) and native speakers (NSs) of English and Mandarin. CS participants are 40 university students whose L1 is Cantonese. Twenty of them are from GD, mainland China, whose L2 is Mandarin and L3 is English. The other 20 are from HK, with English as L2 and Mandarin as L3. Participants from both areas were categorized into high (H) and low (L) groups based on their accuracy rates of a diagnostic test, including a Mandarin reading-aloud task and an English reading-aloud task. All CS participants performed English and Mandarin words-reading aloud tasks with both real and pseudo words. For the English task, words have the [h] sound as the initial, English monophthongs as the vowel, and [d] sound as the final (e.g., had [hæd]). For the Mandarin task, all words are open syllables with the [h] sound as the initial, followed by Mandarin monophthongs (e.g., '壺' [hu˧˥]). Six NSs of English and Mandarin (3 for each) produced the English and Mandarin tasks, respectively. Three of the CS participants produced a reading-aloud task of their L1 Cantonese, with the [h] sound as the initial, followed by Cantonese monophthongs (e.g., '靴' [œ˥]). The NSs' data from the three languages were used as references.

Three L2 and L3 sounds of CSs were measured in this study. The first sound is [u], which exists in Cantonese, Mandarin, and English, and the [u] sound in these three languages belongs to the same phonetic category. The first (F1), second (F2), and third (F3) formants of English and Mandarin [u] produced by CSs were compared with NSs of Mandarin and English's production to identify the possible CLIs. The second and third sounds are English [æ] and Mandarin [ɤ], unique in CSs' L2 or L3. The phonetic features of [æ] and [ɤ] are acoustically distinctive. [æ] is an open front vowel that only exists in L2/L3 English, and [ɤ] is a close back vowel that only exists in L2/L3 Mandarin. CSs' F1, F2, and F3 values of English [æ] were compared with Cantonese and Mandarin NSs' [a]. CSs' production in Mandarin [ɤ] was compared with Cantonese NSs' [œ] and English NSs' [ɜ].

For the [u] sound, which exits in Cantonese, English, and Mandarin, GD participants (both H and L) produced larger F1 values than that of English NSs. However, GD participants produced similar F1 to those of NSs' Cantonese [u] and Mandarin [u]. Acoustic results discovered a combination of L1 (Cantonese) and L2 (Mandarin) progressive CLIs on the vowel in the foreign language English (L3) by GD CSs (both H and L). HK H participants produced larger F2 values on Mandarin [u] than that of NSs but no significant differences with that of Cantonese NSs. Progressive CLIs from L1 to L3 (Mandarin) were identified.

The F1, F2, and F3 values of English [æ] produced by HK and GD CSs received no CLIs from other languages that they learned. The F2 values of Mandarin [ɤ] produced by CSs from the four groups had no statistical differences from that of NSs' English [ɜ]. Their F2 values had statistical differences from that of NSs' Cantonese [œ], except for the HK L group. Acoustic results revealed that CLIs from English to Mandarin (HK H: L2 to L3; GD H & L: L3 to L2) were identified. For the HK L group, a combination of Cantonese and English progressive CLI was identified.

Acoustic results revealed four CLI patterns (both progressively and regressively): 1) a combination of L1 and L2 to L3; 2) L1 to L3; 3) L2 to L3; and 4) L3 to L2. The CLIs for H and L GD participants whose L1 (Cantonese) and L2 (Mandarin) belong to the same language family are consistent. There is a combination of L1 and L2 influences on L3 (English) and a regressive CLI from L3 to L2. For HK speakers, L group participants encountered a combination of L1 and L2 (English) influences on L3 (Mandarin). HK H group participants' L3 received CLIs either from L1 or L2. This study shed light on the monophthongs acquisition in the multilingual context and used acoustic data to support the CLIs.

References

[1] Chen, H. C., & Han, Q. W. (2019). L3 phonology: Contributions of L1 and L2 to L3 pronunciation learning by Hong Kong speakers. *International Journal of Multilingualism*, *16*(4), 492–512.

[2] Kellerman, E. (1983). Now you see it, now you don't. In S. Gass & L. Selinker (eds.), *Language transfer in language learning* (pp. 112–134). New York, U.S.: Newbury House.

[3] Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

[4] Pickett, J. M. (1999). *The acoustics of speech communication : fundamentals, speech perception theory, and technology*. Boston: Allyn and Bacon.

[5] Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *The International Journal of Bilingualism : Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *21*(6), 666–682.

[6] Zee, E. (1991). Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association*, *21*(1), 46-48.

[7] Lee, W. S., & Zee, E. (2003). Standard chinese (beijing). *Journal of the International Phonetic Association*, *33*(1), 109-112.

[8] Roach, P. (2009). *English phonetics and phonology : a practical course* (4th ed.). Cambridge ;: Cambridge University Press.

[9] Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon, U.K: Multilingual Matters Ltd.

[10] Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, *7*(1), 75–90.

[11] Chen, H. C., & Tian, J. X. (2021). The roles of Cantonese speakers' L1 and L2 phonological features in L3 pronunciation acquisition. *International Journal of Multilingualism*, 1–17.

The Effects of VOT on Lexical Access by L1 and L2 Listeners: An Eye-Tracking Study

Eunkyung Sung[1], Sunhee Lee[2] & Sehoon Jung[3]

[1,2]*Cyber Hankuk University of Foreign Studies,*
[3] *Kyungsung University*
eks@cufs.ac.kr, lishanxi@cufs.ac.kr, sehjung0427@gmail.com

The current study examines the effects of within-category differences in voice onset time (VOT) on the dynamics of lexical activation. Specifically, this study compares sensitivity to the VOT cue between native English and Korean listeners in detecting the voicing contrast in English word-initial stop consonants. We used the eye-tracking method to monitor listeners' cognitive processes more closely when dealing with aural input during the picture identification task. The stimuli were modified natural speech tokens varying along six steps of VOT continuum for /b/-/p/, /d/-/t/, and /g/-/k/, respectively. The interval between two steps was 11–19ms for all three pairs. The participants were given aural input in the form of instructions (e.g., *look at the* ____) and asked to pick an image they just heard between the two options (i.e., *palm-bomb*, *pole-bowl*, *tart-dart*, *toe-dough*, *card-guard*, *coat-goat*) on the screen while or after they listened to the input. Figure 1 manifests the screenshot of the experiment.

The results of the experiment showed that both listener groups utilized the VOT cue in recognizing images. Figure 2 displays the results of the keyboard-click responses for each group. The response rates of voiceless stops rose as the VOT values increased for both English and Korean listeners. This means both groups utilized VOT as a major acoustic cue to identify voicing of stop categories.

Figure 3 shows the participants' proportional look on the two images — namely the target and competitor images — starting from the time they heard the target word (i.e., 560ms) and onwards. Comparing the two groups' time-course of fixations to the images, both groups showed similar profiles in that their looks to the target images including voiceless stops (i.e., /p/, /t/, /k/) increased as the VOT level increased. However, there were discrepancies between the two listener groups at least in two respects. First of all, the proportional differences between the 6 levels of VOT were much less for the Korean listeners than for the English listeners. Next, the critical point at which the Korean listeners clearly recognize the target image was found to have occurred at a later point. At the VOT level of 5 or 6, the increase of looks to target images was shown around 780 milliseconds for the Korean listeners, compared to the English listeners (about 720 milliseconds). This means that the English listeners were relatively faster than the Korean listeners in processing the target sounds. In addition, the English listeners demonstrated relatively greater and more stable focus on the targets as the VOT level increased. Therefore, despite some qualitative similarities between the two groups, the analysis of their time course data clearly revealed differences in terms of timing of image recognition and the level of certainty between L1 and L2 listeners.

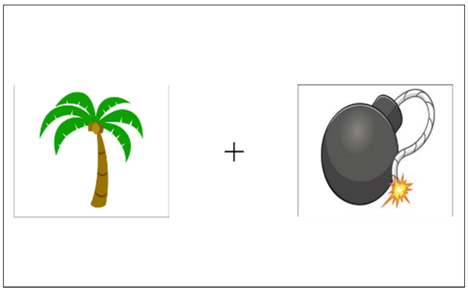


Figure 1. A screenshot of the experiment (*palm* vs. *bomb*)

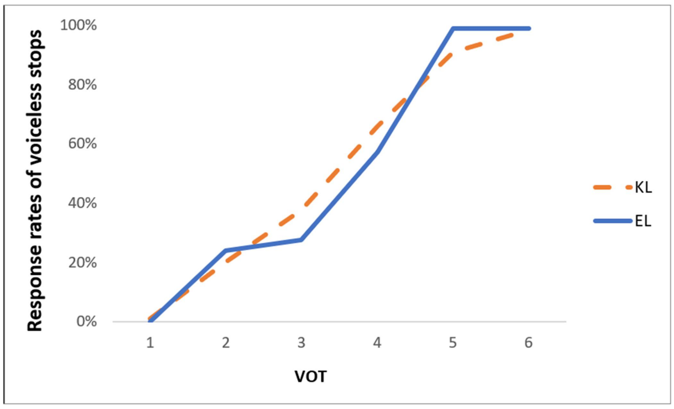


Figure 2. Response rates of voiceless stop (/p/, /t/, or /k/) as a function of VOT by English listeners (EL, solid lines) and Korean listeners (KL, dotted lines)

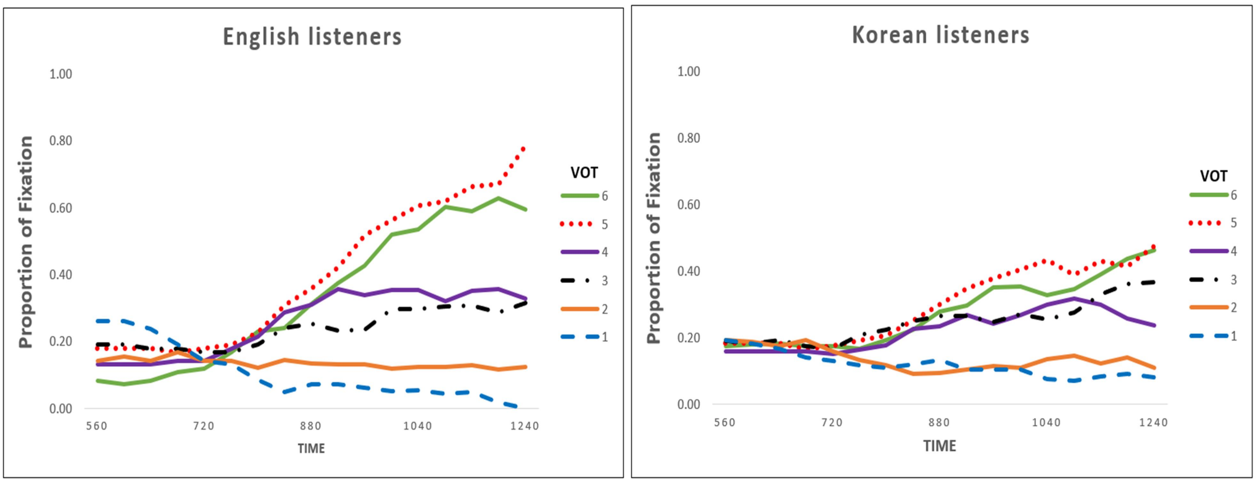


Figure 3. Mean proportion fixation to the target pictures as a function of VOT

References

[1] Kim S., Mitterer, H., and Cho, T. (2018). A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing. *PLoS ONE* 13(8), e0202912.

[2] Kong, E. J. and Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics* 59, 40-57.

[3] Nakai, S. and Scobbie, J. M. (2016). The VOT Category Boundary in Word-initial Stops: Counter-Evidence Against Rate Normalization in English Spontaneous Speech. *Laboratory Phonology* 7(1), 13.

[4] McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86(2), B33-B42.

[5] McMurray, B., Clayards, M. A., Tanenhaus, M. K., and Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review* 15(6), 1064-1071.

[6] Reinisch, E and Mitterer, H. (2022). Phonetics and eye-tracking. In Knight, R. A. and Setter, J. (Eds.), *The Cambridge handbook of phonetics* (pp. 457-479). Cambridge University Press.

SSANOVA as a Method of Examining Nasality in Korean *Aegyo*

Drew Crosby[1]

[1]*University of South Carolina (USA)*
dmcrosby@email.sc.edu

Aegyo, the Korean baby-talk register, is highly performative and has several purported sociophonetic dimensions: rising-falling intonation (LHL%), nasality, and obstruent fortition [6, 8]. Previous anthropological, discursive, and linguistic investigations [5, 7, 10] identify it as a gendered practice associated with “modern and trendy young women in Korean mainstream culture” [9, p. 42], often used for requesting favors, maintaining social harmony, and as a form of politeness to those higher in the social hierarchy [7, 8, 10]. Of its sociophonetic correlates, it has a particularly strong association with nasality. Nasality in *aegyo* has been described as occurring intonation-phrase (IP)-finally in open-syllables [8]. In fact, nasality can be indicated in Korean orthographically by the addition of the Korean symbol for [ŋ], as in (1).

(1) Standard Script | *Aegyo*-style Script
자기야 잘 **자**. | 자기야 잘 **장**~~~ ***^^*** (emoticon: smile with blushing)
/tɕakija tɕal **tɕa**/ | /tɕakija tɕal **tɕaŋ**/
‘Honey good night’ [10, p.13]

Our previous investigation of nasality in *aegyo* (measured over the entire IP-final vowel) showed it to be associated with the age of speakers such that speakers born after 1979 show an increase in nasality when performing *aegyo* whereas those born before 1979 do not. However, contrary to the purported gendered nature of *aegyo*, we failed to find a gender effect [1].

The present study therefore seeks to further examine the gendered nature of nasality in *aegyo* by employing smoothing splines analyses of variance (SSANOVA) to model changes in nasalance (a ratio of intensity of noise from the nasal tract to total intensity from both the oral and nasal tract) across IP final-vowels in performances of *aegyo.* SSANOVA is a method of calculating a best-fit model for curves in a data set using Gaussian process regression [4]. In linguistics it has been used to model tongue shapes [2] and vowel formant trajectories (e.g., [3]). SSANOVA plots are generated with 95% Bayesian confidence intervals and curves that do not overlap are statistically different at that point [3].

The data for the current study consists of interviews with thirteen romantic couples from the central dialect regions of South Korea (Seoul Capital Area, *Chungcheongdo, Gangwondo*) born between 1980 and 1999. Couples were asked to perform three dialogues and three communicative tasks in non-*aegyo* and *aegyo* modes, and to read aloud ten *aegyo*-ful text messages (i.e., orthographic *aegyo*) to their partner. Nasalance measures were calculated at 11 equally spaced time points across the IP-final vowel via intensity measures obtained from earbuds placed under the nostril and in the corner of the mouth [11]. The nasalance measures were then submitted to an SSANOVA model with nasalance as the dependent variable and *aegyo* condition as the independent variable. This modeling is visualized in Figure 1. The plot shows that orthographic *aegyo* (in blue) is far more nasalized than either of the other *aegyo* conditions. Notably, the *aegyo* curve (in green) is higher than the non-*aegyo* curve (in red) and they only overlap at the very beginning of the vowel, supporting our previous results that showed an association between *aegyo* and nasalization. Figure 2 shows separate SSANOVA models for nasalance by gender. The SSANOVA curve for women shows clear separation between the non-*aegyo* and *aegyo* curves throughout the IP-final vowel, whereas the men’s curve has significant overlap between the two curves implying that there is no difference in nasalance between men’s *aegyo* and non-*aegyo* speech. These results corroborate prior assertions that nasality is a feature of *aegyo* and suggest that (at least nasality in) *aegyo* is a gendered practice. This paper also offers a new means of modeling nasality, especially in cases where measures of single points or means can obscure important time-associated details.

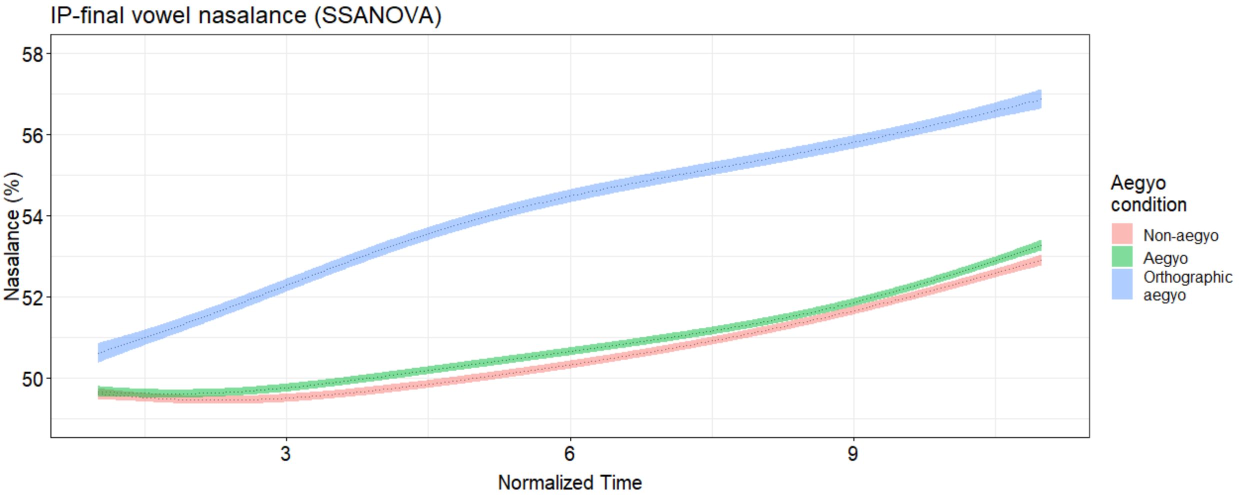


Fig 1. SSANOVA plot of IP-final nasalance by *aegyo* condition

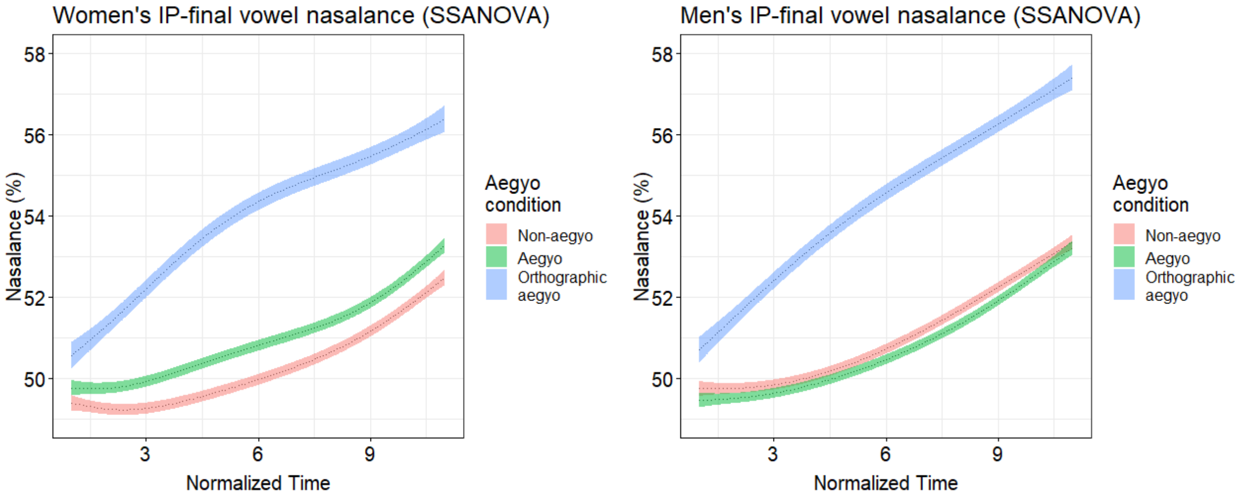


Fig 2. SSANOVA plots of IP-final nasalance gender and *aegyo* condition

References

[1] Crosby, D. & Dalola, A. (2022, October 13-15). Cute nasalization**:** the effect of age on nasalance in Korean *aegyo* [Conference presentation]. NWAV 50, Stanford University, United States.

[2] Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America*, *120*(1), 407-415.

[3] Freeman, V. (2021). Vague eggs and tags: Prevelar merger in Seattle. *Language Variation and change*, 33(1). 57-80.

[4] Gu, C. (2002). Smoothing spline ANOVA models. Springer Series in Statistics. New York: Springer

[5] Han, A. J. (2016). The aesthetics of cuteness in Korean pop music. University of Sussex: PhD diss.

[6] Jang, H. (2021). How cute do I sound to you?: Gender and age effects in the use and evaluation of Korean baby-talk register, Aegyo. Language Sciences, 83.

[7] Manietta, J. B. (2016). Transnational masculinities: The distributive performativity of gender in Korean boy bands. Boulder: University of Colorado: MA thesis.

[8] Moon, K. (2013). Authenticating the fake: Linguistic resources of aegyo and its media assessments. Stanford University: MA thesis.

[9] Moon, K. (2017). Phrase Final Position as a Site of Social Meaning: phonetic variation among young Seoul women. Stanford University: PhD diss.

[10] Puzar, A. & Hong, Y. (2018). Korean Cuties: Understanding Performed Winsomeness (Aegyo) in South Korea. *The Asia Pacific Journal of Anthropology*, 19(4), 333-349.

[11] Stewart, J. & Kohlberger, M. (2017). Earbuds: A Method for Analyzing Nasality in the Field. *Language Documentation & Conservation*, 11, 49-80.

Evaluating forced alignment for under-resourced languages: A test on Squliq Atayal

Chi-Wei Wang[1], Bo-Wei Chen[1], Bo-Xuan Huang[1], Ching-Hung Lai[2] & Chenhao Chiu[1 3]

[1]Graduate Institute of Linguistics, National Taiwan University (Taiwan), [2]School of Medicine, National Cheng Kung University (Taiwan), [3]Neurobiology and Cognitive Science Center, National Taiwan University (Taiwan)

r09142007@ntu.edu.tw, r09142001@ntu.edu.tw, r09142003@ntu.edu.tw, i54051092@gs.ncku.edu.tw, chenhaochiu@ntu.edu.tw

Trainable forced alignment offers feasible solutions to document under-resourced languages. This study aims to assess the performances of a Montreal Forced Aligner (MFA) [1] trained model using a small scale of phonetically transcribed field data in Squliq Atayal, an endangered Austronesian language spoken in Taiwan. Regarding the training dataset, the preliminary corpus consists of a 20-minute recording as an excerpt from a series of fieldwork sessions. All elicited utterances produced by one female Squliq Atayal native speaker were manually labeled at both word and phone levels using Praat [2]. The pronunciation dictionary was generated by combining the word and phone tiers in each manually annotated transcription, ensuring that all words that occurred in the corpus were appended. Once the aligned TextGrid files were retrieved, evaluations were implemented by comparing MFA outputs with manual annotations based on (1) the *accuracy* measurements [3, 4], by measuring the agreements (AG) of interval boundaries at different thresholds, the overlap rates (OvR), and the midpoint displacements (MpD) of each segment; in addition to (2) the *acoustic* measurements [5, 6], by fitting the pitch trajectories through words and the formant trajectories through the most common vowels /a, i, u/ constructed by 30 data points using 7 (F0 + 2 formants * 3 vowels) generalized additive mixture models (GAMMs) [7].

The *accuracy* results suggested that the mean AG of consonants slightly outperformed that of vowels (Table 1) while the mean OvR of consonants was lower, along with the mean MpD being significantly larger when compared to those of vowels (Table 2). Here, the discrepancy might be accounted for by few extreme misalignments. In this case, AG would be more suitable for evaluating overall performances, while OvR and MpD were considered more robust when evaluating the effects of segment types on alignment accuracy. On the other hand, for *acoustic* trajectories (both pitch and formants), no statistical significance was found between the MFA and manual annotations, except for the F2 trajectories of [u], as illustrated in Figures 1 and 2, which positively supported the reliability of the current MFA model. Overall, the current results revealed that MFA outcomes were highly consistent with manual annotations when little but comprehensively labeled data were provided. Moreover, our results also suggested that different evaluation methods may come along with diverged results and should be implemented based on the objectives of the research.

Keywords: forced alignment, phonetic fieldwork, language documentation, Squliq Atayal

Threshold	Vowel	Consonant	Overall
10 ms	36.32%	36.67%	36.49%
20 ms	63.37%	68.66%	66.02%
30 ms	76.26%	83.33%	79.80%
40 ms	79.09%	85.73%	82.41%
50 ms	80.35%	86.40%	83.37%

Table 1. The mean AGs at both boundaries for vowels and consonants, along with the overall AGs, at different thresholds.

Type	OvR	MpD
Vowel	54.67%	33.68 ms
Consonant	41.19%	211.23 ms
Overall	47.37%	129.76 ms
[a]	45.04%	50.62 ms
[i]	59.82%	25.44 ms
[u]	40.73%	39.70 ms

Table 2. The mean OvRs and MpDs of vowels, consonants, overall performances, and the most common vowels [a], [i], [u].

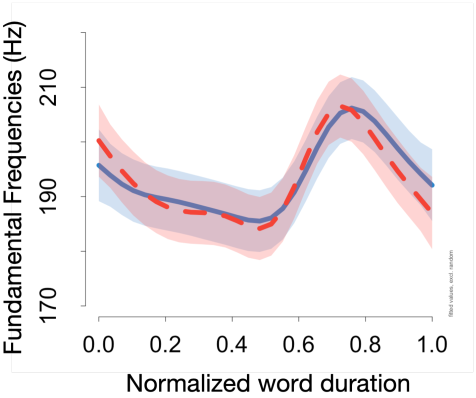


Figure 1. F0 trajectories over normalized word duration fitted by GAMM (MFA = red dashed line; manual = blue solid line).

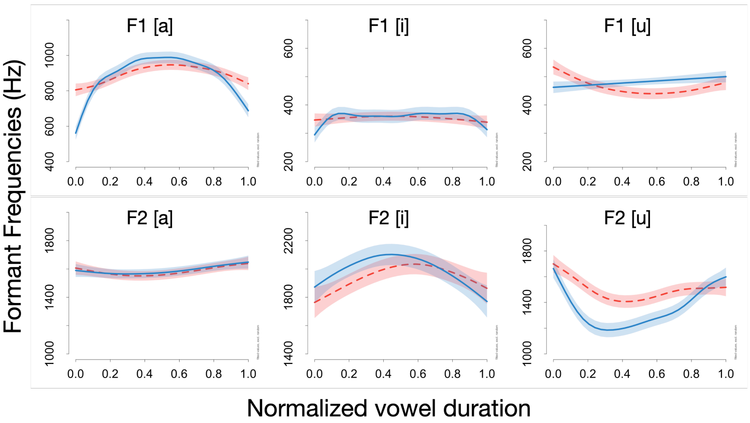


Figure 2. Formant trajectories over the normalized [a, i, u] fitted by GAMMs (MFA = red dashed line; manual = blue solid line).

References

[1] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. In *Interspeech*, 498–502.

[2] Boersma, P. and Weenink, D. (2021). Praat: Doing phonetics by computer.

[3] Jones, C., Li, W., Almeida, A., and German, A. (2019). Evaluating cross-linguistic forced alignment of conversational data in North Australian Kriol, an under-resourced language. *Language Documentation and Conservation*, *13*, 281–299.

[4] Gonzalez, S., Travis, C., Grama, J., Barth, D., and Ananthanarayan, S. (2018). Recursive forced alignment: A test on a minority language. In *Proceedings of the 17th Australasian International Conference on Speech Science and Technology*, 145–148.

[5] Babinski, S., Dockum, R., Craft, J. H., Fergus, A., Goldenberg, D., and Bowern, C. (2019). A robin hood approach to forced alignment: English-trained algorithms and their use on Australian languages. In *Proceedings of the Linguistic Society of America*, 1–12.

[6] Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America, 142*(4), 2469–2482.

[7] Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86-116.

Denasalization and the Phonological Representation of Voiced Stops in Lushootseed

Ted Kye

University of Washington, Seattle (USA)
tkye29@uw.edu

Denasalization is a sound change where nasal stops have historically changed to voiced oral stops (i.e., */n/>/d/ and */m/>/b/). There are several languages in the Pacific Northwest that lack contrastive nasals, among them Lushootseed. The historical literature claims that in Lushootseed nasals became voiced stops (Hockett 1955, Thompson & Thompson 1972). However, there hasn't been an acoustic phonetic study of these sounds. In this study, I perform an acoustic phonetic analysis to look for residual nasality in voiced stops. High quality archival recordings were examined. The research question is whether some nasality was present for the voiced stops /b/ and /d/ in Lushootseed. Prenasalization of voiced stops occurs in prosodic domain-initial position for some languages (Gudschinsky et al. 1970, Iverson & Salmon 1996, Piñeros 2003). Therefore, nasality may be observed in similar positions in Lushootseed. I compare voiced stops /b d/ in prosodic domain-initial position with /b d/ in intervocalic and non-prominent (unstressed) initial positions. I predict that partial (or weak) nasality can be observed for the voiced stops /b/ and /d/ in domain-initial position (realized as prenasalized stops [ᵐb] and [ⁿd] respectively) but not intervocalic or non-prominent initial position.

I conducted a qualitative and quantitative analysis of voiced stops by listening for audible nasal murmur as well as inspecting spectrograms and waveforms. Frequency components of the voiced stops /b/ and /d/ were compared with nasal stops /m/ and /n/ in the word /mimuʔan/ 'small, little', which is the only word in Lushootseed that retained its nasals (the word occurred only once in these recordings). Nasals have the property of introducing extra pole zero pairs in the transfer function below and above frequencies of the first formant (Chen 1996, Fujimura & Lindqvist 1964). The zeros are located near the peak of F1, which introduces peaks below F1 (near 250Hz) and above F1 (near 1000Hz). In this study, the amplitudes of the peaks were calculated from an LPC. The amplitude of the peak corresponding to the extra peak above F1 (near 1000Hz) was labelled P1, whereas the amplitude of F1 was labelled A1. These measures were compared between the voiced bilabial stop [b], prenasalized bilabial [ᵐb], and the bilabial nasal [m] in the word /mimuʔan/ 'small'. The percentage of voiced stops that were prenasalized was also calculated.

Results: There were 68 voiced bilabial stops /b/ and 50 voiced alveolar stops /d/ that were produced by the speaker. 19 of the 68 voiced bilabial stops /b/ and 13 of the 50 voiced alveolar stops /d/ occurred word-initially (domain-initially). For the voiced bilabial stops /b/ in word-initial position, 63.2% (12/19) were realized with partial (weak) nasality. For the voiced alveolar stops /d/, 62% (8/13) were realized with partial (weak) nasality. Of the 5 word-initial voiced alveolar stops /d/ that were not partially nasalized, 2 were devoiced (as in [d̥]). Partial (weak) nasality occurred to voiced stops /b d/ only in prosodic domain-initial position (i.e., word-initial position or initially under focus). Figure 1 and 2 shows the partial nasality of /b d/ in prosodic domain-initial position. None of the voiced stops were realized with partial nasality in intervocalic position or initially in unstressed environments. Amplitudes A1 and P1 for the bilabial stops [b], [ᵐb] and [m] are summarized in Table 1. As Table 1 summarizes, A1 for [ᵐb] was greater than [b]. Moreover, A1 and P1 for the prenasalized variant [ᵐb] was considerably lower than [m]. This suggests that prenasalized variants were produced with a weaker nasal airflow than fully nasalized stops.

Discussion: These results indicate that there is partial (weak) nasality for voiced stops in Lushootseed. However, partial nasality is restricted to domain-initial position exclusively. There are two explanations for the partial nasality of voiced stops in these recordings. One explanation is that the nasal series was partially preserved in domain-initial position for voiced stop reflexes. Under this view, voiced stops pattern with sonorants in Lushootseed, where voiced stops lack an underlying [voice] feature (Rice 1993). On the other hand, the partial nasality could be interpreted as voice *enhancement* in prosodically strong environments (Wetzels & Nevins 2018). The low-frequency amplitude of the voiced stop is increased by preventing a buildup of intraoral pressure during a portion of the closure interval (Stevens et al. 1986:439). This enhances the audibility of the low-frequency noise that is characteristic of voiced stops. The view that partial nasality is due to voice enhancement makes the most sense because partial nasality is observed exclusively in prosodic domain-initial position. Moreover, the partial nasality is observed in only a portion of the stops in this position (it is not observed initially in unstressed syllables). For this reason, voiced stops should underlyingly be represented with a [voice] feature, contrary to Rice (1993).

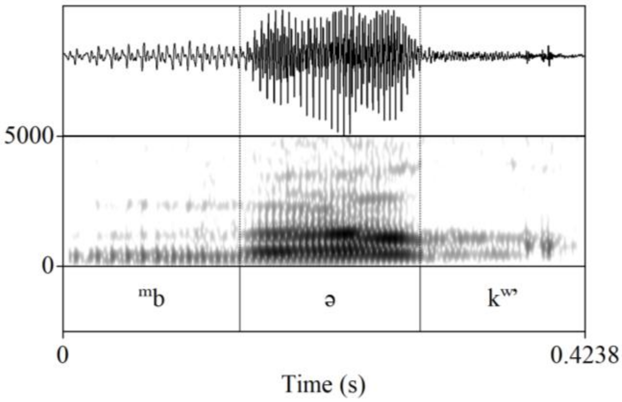


Fig.1 Prenasalized bilabial stop [ᵐb] in prosodic domain-initial position, from the word /bəkʷ'/ 'all'.

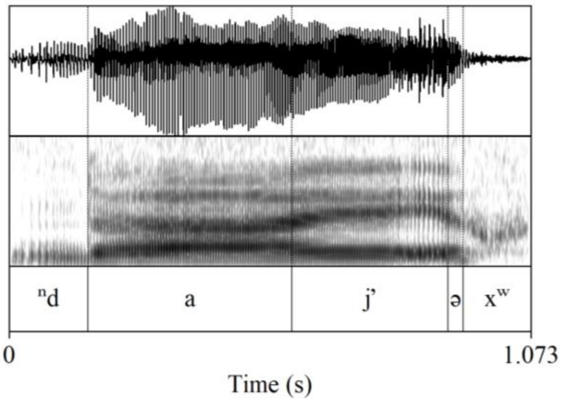


Fig.2 Prenasalized alveolar stop [ⁿd] in prosodic domain-initial position, from the word /daj'əxʷ/ 'just now'.

Table 1. Amplitudes A1 and P1 for voiced bilabial stop [b], prenasalized [ᵐb], and nasal stop [m] at the beginning of the word /mimuʔan/ 'small'

Segment	A1 (in dB)	P1 (in dB)
[b]	23.0 dB	N/A
[ᵐb]	33.9 dB	11.4 dB
[m]	53.2 dB	25.2 dB

References

[1] Campbell, Lyle (1997). *American Indian Languages: The Historical Linguistics of Native America.* Oxford: Oxford University Press.

[2] Chen, M. Y. (1996). Acoustic correlates of nasality in speech. (Doctoral dissertation, Massachusetts Institute of Technology).

[3] Fujimura, O., & Lindqvist-Gauffin, J. (1964). The sinewave response of the vocal tract. In *Speech Transmission Laboratory Quarterly Progress and Status Report, 2*(3), 17-37.

[4] Gudschinsky, Sarah C., Harold Popovich, & Frances Popovich (1970). Native reaction and phonetic similarity in Maxakali phonology. In *Language, 46,* 77-88.

[5] Hockett, Charles F. (1955). A manual of phonology. In *International Journal of American Linguistics, 21*(4).

[6] Iverson, Gregory K., & Joseph C. Salmons (1996). Mixtec prenasalization as hypervoicing. In *International Journal of Americal Linguistics, 65*, 163-175.

[7] Piñeros, Carlos-Eduardo (2003). Accounting for the instability of Palenquero voiced stops. In *Lingua, 133,* 1185-1222.

[8] Rice, Keren D. (1993). A reexamination of the feature [sonorant]: The status of 'sonorant obstruents'. In *Language, 69*(2), 308-344.

[9] Stevens, Kenneth N., Samuel J. Keyser, & Haruko Kawasaki (1986). Toward a phonetic and phonological theory of redundant features. In J.S. Perkell & D.H. Klatt (eds.), *Invariance and variability in speech processes,* pp. 426-463. New Jersey: Lawrence Erlbaum Associates.

[10] Thompson, L. C., & Thompson, M. T. (1972). Language universals, nasals, and the Northwest Coast. In *Studies in Linguistics in Honor of George L. Trager, Mouton, The Hague,* 441-56.

[11] Wetzels, W. Leo, & Andrew Nevins (2018). Prenasalized and postoralized consonants: The diverse functions of enhancement. In *Language, 94*(4), 834-866.

Phonetic development of English stress production and perception by EFL and ESL Korean speakers

Hijo Kang (Chosun University) and Hyun-ju Kim (SUNY Korea)
hijo.kang@gmail.com, hjkim@sunykorea.ac.kr

L2 learners of English are known to experience significant challenges in acquiring native-like perception and production [1,2]. L2 learners may employ different strategies to perceive and produce English stress from native speakers as previous studies reported that Korean learners used pitch more than intensity and duration to realize English stress unlike native speakers [3]. This study examines the development patterns of English stress by Korean learners to investigate how their strategies for English stress production and perception are different depending on their proficiency level and learning environment.

Phonetic experiments were designed to compare the weights of the three suprasegmental cues used to perceive and produce English stress: intensity, duration, and pitch, among three groups of L2 learners: mid/high level EFL and ESL. In the production experiment, we recorded the three groups of 36 Korean speakers and one group of 6 English native speakers, each of whom produced 33 English words twice (without and with stress marking on the words). One stressed vowel and one unstressed vowel were segmented from each word and measured to obtain the values of three acoustic cues. In the perception experiment, two bisyllabic nonce words were modulated into five steps of each acoustic cues. The participants listened to each token and asked to choose between initial and final stresses. Unlike the previous studies, the values in the production experiment and the choices in the perception experiment were subject to primary component analysis to reveal which acoustic cue explains the participants' subconscious production and perception of English vowels. Production results revealed that Korean speakers make most use of duration unconsciously and of pitch consciously, while native speakers rely most on intensity irrespective of condition (see Table 1). ESL speakers were more like native speakers in that the weight of pitch was much lighter than the other two Korean groups (see Figure 1). However, all the Korean speakers tended to emphasize pitch when they intended to give stress. In the perception experiment, pitch and intensity best accounted for the native and ESL speakers' choices for initial stress, which was followed by duration. In contrast, pitch alone was the most important acoustic cue that was responsible for the EFL speakers' choice for initial stress (see Table 2). For final stress, duration was the most reliable acoustic cue for Korean speakers, while intensity was for native speakers.

All the results suggest that Korean learners of English mainly adopt pitch to realize and recognize English stress. EFL speakers did not differ depending on their levels, but ESL speakers showed they had learned to make more use of intensity, as native speakers do. The results indicate that learning environment should significantly affect the perception and the phonetic realization of English stress. Also, the results indicate that perception and production do not go in parallel in L2 learning.

Table 1. Primary Cues for Production

	Without Stress Marking	With Stress Marking
Mid Level EFL	Duration > Intensity > Pitch	Pitch > Duration > Intensity
High Level EFL	Duration > Intensity > Pitch	Pitch > Intensity > Duration
ESL	Intensity = Duration > Pitch	Pitch > Duration > Intensity
Native	Intensity > Duration > Pitch	Intensity > Duration > Pitch

Table 2. Primary Cues for Perception

	Initial Response	Final Response
Mid Level EFL	Pitch > Intensity > Duration	Duration > Pitch > Intensity
High Level EFL	Pitch > Intensity > Duration	Duration > Pitch > Intensity
ESL	Pitch = Intensity > Duration	Duration > Pitch > Intensity
Native	Pitch = Intensity > Duration	Intensity > Duration > Pitch

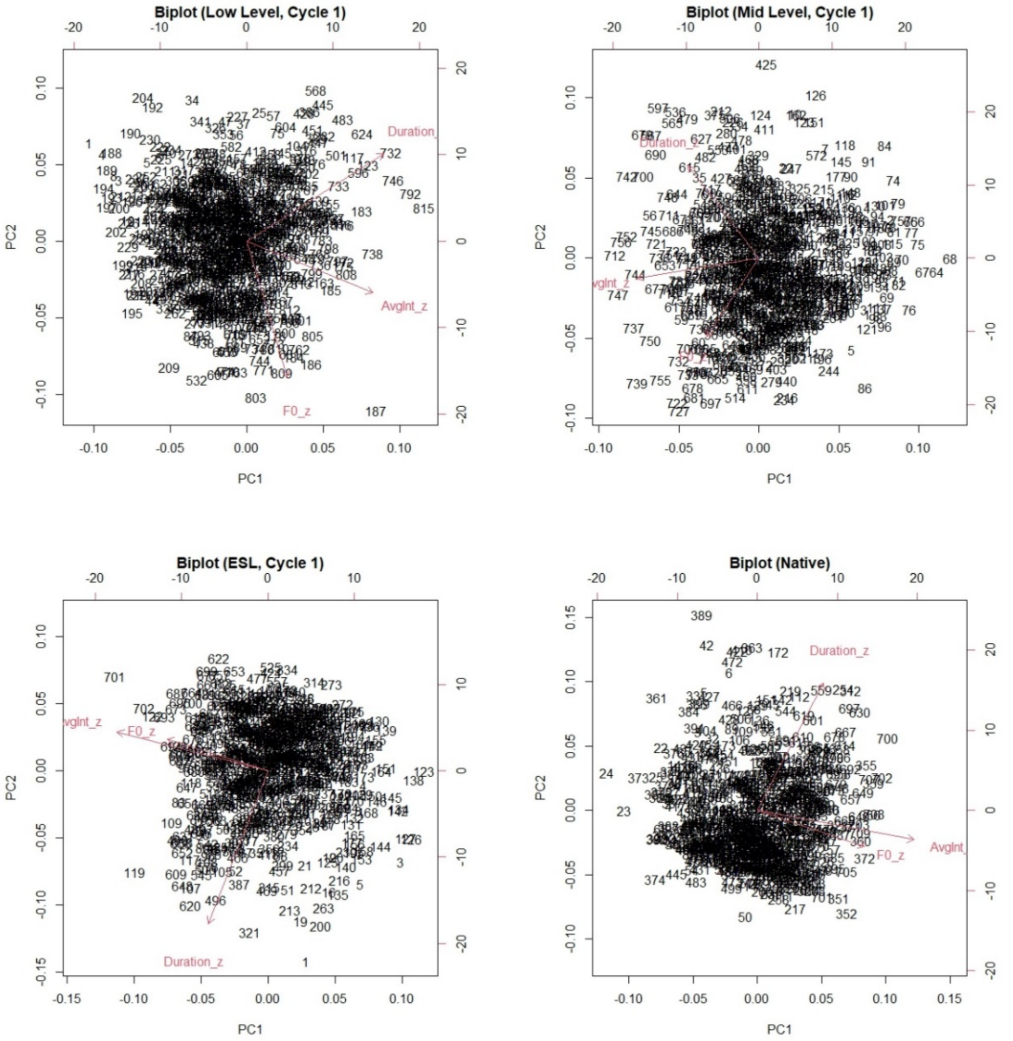


Fig.1 Biplots of PCA for low level, mid level, ESL, and native groups

References

[1] Archibald, John. 1993. *Language Learnability and L2 Phonology: The Acquisition of Metrical Parameters*. Dordrecht: Kluwer.

[2] Zhang Y., Nissen, S. L. & Francis, A. L. 2008. “Acoustic characteristics of English lexical stress produced by native Mandarin speakers,” *The Journal of the Acoustical Society of America* 123, 4498–4513.

[3] Lee, B., Susan G. G., and Tetsuo H. 2006. “Acoustic Analysis of the Production of Unstressed English Vowels by Early and Late Korean and Japanese Bilinguals,” *Studies in Second Language Acquisition* 28, 487-513.

Within-category cue sensitivity in native language perception and its relation to non-native phonological contrast learning

Jieun Lee[1] & Hanyong Park[2]

[1]*University of Kansas (USA),* [2]*University of Wisconsin-Milwaukee (USA)*
jieunlee@ku.edu, park27@uwm.edu

Previous research has challenged assumption of categorical perception and provided evidence of gradient encoding of speech categories [1]. Previous studies with alternative measures of speech perception, such as 4IAX and the Visual Analogue Scaling tasks, have consistently shown listeners' ability to discriminate speech stimuli within the same category and a wide range of individual differences in their sensitivity to subtle within-category acoustic differences [2, 3, 4]. The current study attempted to quantify individual variability in within-category acoustic cue sensitivity in the perception of the native language (L1) category and its relation to second language (L2) phonological contrast learning (Experiment 1). We also investigated the effectiveness of Cue-Attention Switching Training (CAST). CAST is designed to downweight and upweight the irrelevant and the relevant acoustic cues of a target L2 phonological contrast, respectively, in native category perception. We examined whether such training would decrease the possible L2 learning gap due to the listeners' different sensitivities to the acoustic cues (Experiment 2).

Experiment 1: 24 Korean adult learners of English participated in learning three English vowel contrasts /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/ -/u/, which are distinguished primarily by spectral properties and secondarily by duration. The AXB oddity task was adopted to quantify participants' sensitivity to within-category differences induced by spectral and duration cue changes. A set of stimuli was constructed from the Korean /i/ vowel, with five different steps of spectral and duration properties. A and B stimuli in AXB pairs had either a one-step spectral or a one-step duration difference from the X stimulus, while the other step remained the same as the X stimulus. Participants were asked to judge whether A or B was more distinct from X and pick the "most odd one" out. We hypothesized that participants who dominantly selected stimuli with different spectral steps (i.e., High Sensitive (HS) group) would master the distinctions between English vowels of target contrasts better than those who did not show such high sensitivity to spectral steps (i.e., Low Sensitive (LS) group). In a pretest-posttest design, both groups received a five-day computer-based auditory L2 training with a set of stimuli comprising minimal pairs from the endpoints of a five-step spectral and duration continuum of either /i/-/ɪ/, /ɛ/-/æ/, or /ʊ/ -/u/: hid/heed, head/had, and hood/who'd. Overall, the HS group demonstrated an initial advantage in L2 contrast learning. This group showed higher scores in tests taken after each training day and more nativelike and increased use of spectral cues in the post-test compared to the LS group (Figs. 1 & 2). The individual-level analysis showed that despite individual variability in developmental trajectories in spectral and duration cue weightings over time, most HS group participants presented a desired trajectory pattern of enhancing the reliance on spectral dimension after the training. There were also positive correlations between participants' AXB oddity task results and their spectral cue weights for the target /i/-/ɪ/ and /ʊ/-/u/ contrasts, indicating that the higher selections of spectrally different stimuli are associated with the higher use of spectral dimension after the training (Fig. 3).

Experiment 2: A new group of 25 participants (LS2 group) took part in the same procedure as in Experiment 1 with short, simple two-alternative forced choice CAST before each training session. CAST stimuli were a subset of the AXB oddity task stimuli. We included stimuli varying along all five duration steps but only at the two most extreme spectral steps (steps 1 & 5). Participants identified each stimulus as either "Type 1 Korean /i/" or "Type 2 Korean /i/" and received feedback. The spectral properties of stimuli always determined the correct answer: "Type 1 Korean /i/" for stimuli with spectral step one and "Type 2 Korean /i/" for stimuli with spectral step five. The primary purpose of CAST was to downweight duration (i.e., make it an uninformative dimension for participants' decisions) and reallocate participants' attention to spectral cues. We observed some benefits of CAST, especially for the /i/-/ɪ/ contrast (Fig. 4). The

LS2 group identified the target English vowel contrasts better than the LS group in Experiment 1. Notably, the LS2 group exhibited more native-like use of the spectral cue than the LS group.

Our study demonstrated that L2 learners' individual differences in sensitivity to sub-phonemic acoustic details in L1 could explain variability in the learning of novel nonnative phonological contrasts. This suggests the transfer of L1 cue sensitivity to L2 cue utilization: how successfully L2 learners progress to become more nativelike listeners can be predicted in terms of their sensitivity to the L2 informative acoustic cue in L1 speech perception. Furthermore, we have shown that the addition of CAST can be helpful for L2 learners to overcome initial disadvantages due to such individual differences in L1 perception.

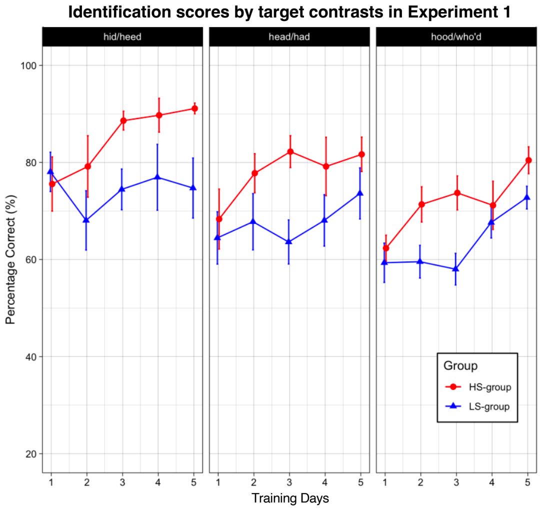


Fig.1 Identification test scores (% correct) after each training day by target contrasts for the HS-group (red) and the LS-group (blue) in Experiment 1.

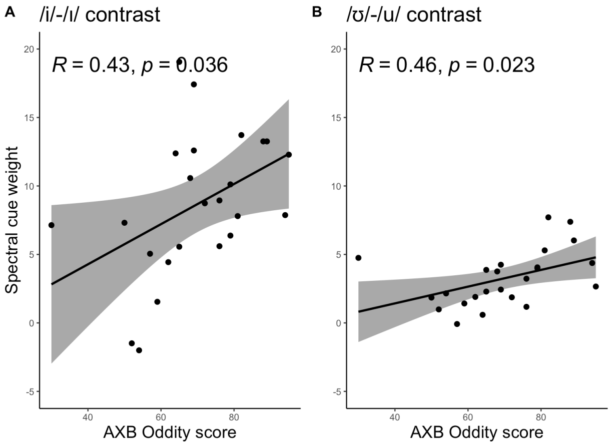


Fig.3 Correlations between individual participants' AXB oddity task scores and their beta-coefficients of spectral dimension for the /i/-/ɪ/ (left) and /ʊ/-/u/ (right) contrasts from the logistic regression analysis fitted to each participants' response data of the post-test.

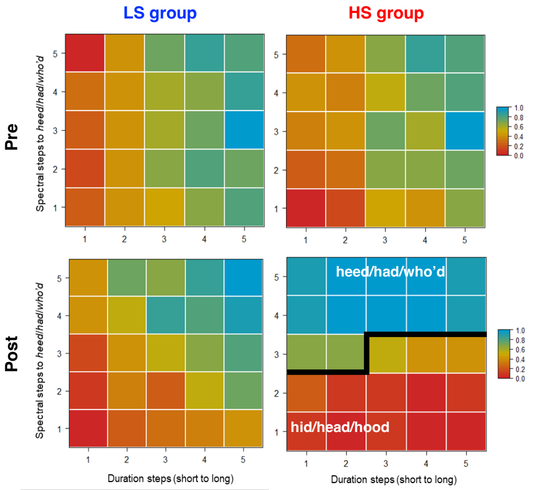


Fig. 2 Heat plots of overall responses of the pretest and the post-test in Experiment 1 to each combination of spectral and duration dimensions averaged across three vowel contrasts. The darkest red cells elicited 100% '*hid/head/hood*' responses, while the darkest blue cells elicited 100% '*heed/had/who'd*.'

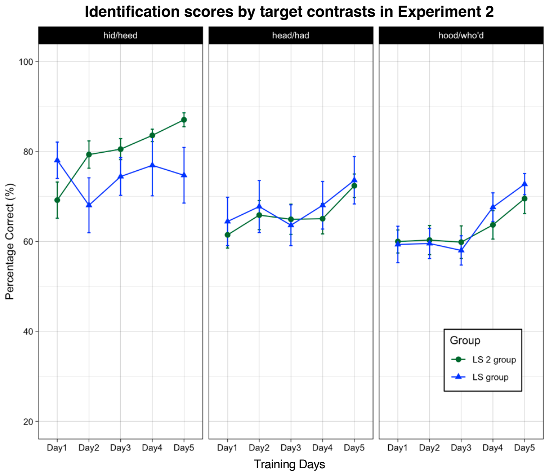


Fig. 4 Identification test scores (% correct) after each training day by target contrasts for the LS 2 group (green) and the LS-group (blue).

References

[1] McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, *152*(6), 3819–3842.

[2] Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America*, *152*(6), 3728–3745.

[3] Kong, E. J. & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

[4] Pisoni, D. B. & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*(2), 328–333.

Cross-linguistic Influences on Speech Prosody by Cantonese Multilingual Speakers

TIAN JINGXUAN[1]

[1]*The Education University of Hong Kong (Hong Kong)*
xuanxuan1028123@msn.com

Multilingual context is common in modern society, and previous studies reported that multilingual language acquisition is a complex process and different from first language (L1) or second language (L2) acquisition (e.g., Chen & Han, 2019). In the Chinese context, there are many multilingual language learners whose L1 is usually their dialect. Putonghua (Mandarin, hereafter) is the dominant language in mainland China and L2 for these multilinguals. For example, Cantonese is language learners' L1 in Guangdong province (GD). Mandarin is their L2, and English is their third language (L3) which is the foreign language that Mainland students are required to learn at school. These three languages are also three common languages used in Hong Kong (HK) (Wang & Kirkpatrick, 2015); however, language learners' background in HK is different from that of learners in GD. English is their L2 owing to the historical issue. Mandarin is their L3 and is more frequently used after the handover and the implementation of the 'biliteracy and trilingualism' policy (Wang & Kirkpatrick, 2013).

Cantonese, Mandarin, and English have different prosodic features. For fundamental frequency (F0), Han et al. (2022) revealed that the F0 of Mandarin speakers is larger than that of English speakers. In Mandarin, there are words with neutral tones, which are considered as unstressed syllables. However, the syllables with neutral tones are not toneless but have a target F0 and are independent of the surrounding tones. The syllables with a neutral tone have a static and mid F0 (Chen & Xu, 2006). The pitch range of a syllable with a neutral tone and its surroundings is large in Chinese. Chinese is a syllable-timed language, while English is stress-timed. Cantonese is a typical syllable-timed language, while Beijing Mandarin is less syllable-timed compared with Cantonese (Mok, 2009). Owing to the different prosodic features of the three languages, Cross-linguistic influences (CLIs) could influence the acquisition of speech prosody in the multilingual context.

However, few previous studies focused on CLIs patterns of L3 prosody. Zhu et al. (2019) investigated L3 prosody produced by Cantonese-English-German trilinguals. L3 rhythm receives effects from L1 transfer and L2 interlanguage transfer. However, no Cantonese data of the Cantonese-English-German participants were collected as a reference. Han et al. (2022) examined influences from L1 and L2 to L3 prosodic features (mainly on pitch span, pitch level, and duration ratio). For GD multilinguals, progressive transfer from L2 large pitch span to L3 was identified; however, regressive transfer from L3 large pitch span to L2 was also identified when Hong Kong Cantonese speakers' L3 is more proficient than L2. Han et al. (2022) also discovered that GD Cantonese speakers' L2 neutral tone duration has an impact on their L3 English function word duration. However, Han et al. (2022) only investigated the prosody features at an initial level and measured the duration ratio on the phrasal level. Duration metrics (e.g., nPVI) should be calculated. This study aims to calculate duration metrics and pitch span on the phrasal level and investigate the possible CLIs on speech prosody by Cantonese speakers.

Participants of this study were 30 university students from HK and GD (15 from each area) who have Cantonese as their L1. The 15 HK participants have English as L2 and Mandarin as L3. However, for the 15 GD participants, Mandarin is their L2, and English is their L3. Ten out of the 15 HK participants had relatively higher proficiency in L2 (HK L2 H) English. The other 5 HK participants had relatively higher proficiency in L3 Mandarin (HK L3 H). For GD participants, all 15 have higher proficiency in their L2, Mandarin. None of them have a higher L3 proficiency level than their L2. Chen and Tian (2021) also reported this limitation in their study because Mandarin is the official and dominant language. It would be hard to recruit participants whose English, as a foreign language, has a higher proficiency level than the dominant language. However, 5 GD participants had relatively higher proficiency in L3 English, who received 18 to 19 (out of 20) in the college entrance English exam (GD L3 H), compared with the other 10 participants who received 10 to 15 (GD L2 H).

All 30 participants performed Cantonese, Mandarin, and English passage reading-aloud tasks. Cantonese and Mandarin versions of ‘The North Wind and the Sun’ were selected. The English passage was retrieved from Chen and Chung (2008), in which different sentence types were included (e.g., Wh-questions & Yes-No questions). 2 native speakers (NSs) of English and Mandarin were also recruited and performed the English and Mandarin passage reading, which was used as the reference. Duration and pitch were measured using Praat (Boersma & Weenink, 2021). The syllable duration was normalized by calculating the nPVI for syllables (nPVI_S). The pitch span for phrasal level utterances is calculated.

HK participants (both groups) produced relatively smaller L1 nPVI_S compared with that of GD participants. The small L1 nPVI_S was transferred to HK participants’ L3 Mandarin, which also had smaller nPVI_S compared with that of GD participants and Mandarin NSs. Progressive CLIs from L1 to L3 were identified. For the HK L2 H group, the CLIs from L2 to L3 and L1 could not be identified. However, 2 out of the 10 HK L2 H participants, who had larger L2 English nPVI_S compared with the other 8 HK L2 H participants, produced relatively larger L1 Cantonese nPVI_S. Regressive CLIs (from L2 to L1) were also identified. However, the L2 regressive CLIs were limited and only occurred when the participants’ L2 English nPVI_S was close to that of NSs. HK participants’ pitch span mainly received CLIs from L1 Cantonese. Their L1 pitch span (both groups) was relatively large, which was transferred to HK participants’ L2 English and L3 Mandarin. HK participants’ pitch span of English and Mandarin was 2 or 3 times larger than those of NSs.

GD participants’ L1 and L2 nPVI_S values were not statistically different. For GD participants, the nPVI_S values of L3, English, were also close to those of their L1 and L2. Strong L1 and L2 CLIs to L3 could be identified. For pitch span, GD participants’ L2, Mandarin, received CLIs from their L1. GD participants (both groups) produced a larger pitch span (almost two times larger) than that of NSs. A progressive CLI pattern, from L1 Cantonese to L2 Mandarin, could be identified. This study provided valuable contributions to speech prosody acquisition in the multilingual context.

References

[1] Chen, H. C., & Han, Q. W. (2019). L3 phonology: Contributions of L1 and L2 to L3 pronunciation learning by Hong Kong speakers. *International Journal of Multilingualism, 16*(4), 492–512.

[2] Wang, L., & Kirkpatrick, A. (2015). Trilingual education in Hong Kong primary schools: An overview. *Multilingual Education*, *5*(3), 1–26.

[3] Wang, L., & Kirkpatrick, A. (2013). Trilingual education in Hong Kong primary schools: A case study. *International Journal of Bilingual Education and Bilingualism*, *16*(1), 100–116.

[4] Han, Q. W., Tian, J. X., & Chen, H. C. (2022). L3 Prosody: Cross-linguistic influence of prosodic features in Mandarin and English by Cantonese multilinguals. *L3 Acquisition After the Initial State.* Amsterdam, The Netherlands : John Benjamins Publishing Company.

[5] Chen, Y., & Xu, Y. (2006). Production of weak elements in speech–evidence from f_0 patterns of neutral tone in Standard Chinese. *Phonetic*a, *63*, 47-75.

[6] Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics,* 2, 148-154.

[7] Zhu, Y., Chen, A., Sudhoff, S., Mok, P., Calhoun, S., Escudero, P., & Warren, P. (2019). Third language prosody: Evidence from Cantonese-English-German trilinguals. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, 2019*. (pp. 3735-3739). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

[8] Chen, H. C., & Tian, J. X. (2021). The roles of Cantonese speakers’ L1 and L2 phonological features in L3 pronunciation acquisition. *International Journal of Multilingualism,* 1–17.

[9] Chen, H. C., & Chung, R. F. (2008). Interlanguage analysis of phonetic timing patterns by Taiwanese learners. *Concentric: Studies in Linguistics, 34,* 79-108.

[10] Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer* [Computer program]. Version 6.2.

Crosslinguistic Influence on the Gestural Dynamics of Focus Prosody in Native Mandarin Learners of L2 English

Yichen Wang[1], Benjamin Kramer[2], Noah Macey[2], Michael Stern[2], Yuyang Liu[2], Jason Shaw[2]

[1]Michigan State University (USA), [2]Yale University (USA)
wangy176@msu.edu, ben.kramer@yale.edu, noah.macey@yale.edu,
michael.stern@yale.edu, yuyang.liu.yl2472@yale.edu, Jason.shaw@yale.edu

1. Introduction. Studies on L2 prosody have found evidence for the influence of L1 on L2 prosody production across a range of language pairs [1–5]. L2 speakers often exhibit prosodic patterns—accentual peaks, stress patterns and tonal contours—that can be attributed to L1 influence (*Transfer Hypothesis*; [6]). However, little has been done to capture such patterns formally. Using the parameters of Articulatory Phonology (AP; [7]), we explore how L2 prosody relates to L1 and to the target language. We investigate this issue in the kinematics of L2 focus production, considering highly proficient L2 learners of English from Mandarin L1. In both of these languages, as in others, syllables under informational focus tend to be longer in duration than unfocused syllables [8–13]. In AP, increased duration can be explained by a π-gesture [14]. The temporal scope of the π-gesture may vary across languages, possibly in ways that could be obscured by acoustic measures of syllable duration. We therefore collected kinematic data using Electromagnetic Articulography and analyzed sub-intervals of the syllable based on articulatory landmarks. Focused and unfocused syllables produced by L1 Mandarin and L1 English speakers established baseline patterns of π-gesture alignment in each language, against which we assessed L2 focus prosody.

2. Methods. Participants were 12 native speakers of Mandarin Chinese who were also L2 speakers of English and 12 native speakers of American English. Materials included eight word-initial CV sequences, consisting of /b/ or /m/ followed by /ɑ/ or /i/ in each language. All Mandarin target syllables carried a falling tone (T4). Each was produced in a *focused* condition, in which it was located in an informationally prominent position in the sentence, and a *non-focused* one. Carrier sentences were preceded by a context presented auditorily and orthographically. Nine sensors attached to the articulators and head were tracked using the NDI Wave Speech Research System at a sampling rate of 100 Hz. Acoustic data were recorded concurrently. Sensors were attached to the tongue tip, tongue blade, tongue dorsum, lower incisor, upper lip, and lower lip. Reference sensors were attached on the nasion or bridge of the nose, as well as the right and left mastoids. After computationally correcting for head movements, target C and V gestures were parsed from sensor trajectories in MVIEW [15]. We used the lips to parse labial consonants and the tongue body for vowels. Gestural landmarks were used to defined four key intervals: *CV lag* (consonant onset to vowel onset), *consonant closing interval* (C_{CLOS}; consonant onset to consonant target), *consonant opening interval* (C_{OPEN}; consonant release to consonant offset), and *vowel opening interval* (V_{OPEN}; vowel onset to point of minimum vowel velocity). A total of 3612 tokens entered into the analysis.

3. Results. Table 1 presents means and standard deviations of the target intervals for L1 English, L1 Mandarin, and L2 English, organized from the beginning of the syllable (left) to the end of the syllable (right). Most of the intervals were longer in syllables produced in the *focused* condition than in syllables produced in the *non-focused* condition. We fit linear mixed effects regression (LMER) models to the four sub-intervals (Table 2). Models included random intercepts for item and subject, a control fixed factor for vowel, and an experimental fixed factor for focus. Results indicated a significant effect of focus on all intervals in L1 English, whereby each interval was longer in the *focused* condition than in the *non-focused* condition. The effect increased in magnitude over time: smallest for CV lag, progressively larger for C_{CLOS}, C_{OPEN}, and V_{OPEN}. For Mandarin, the model indicated a significant lengthening effect of focus only on the final interval V_{OPEN}, and this effect was smaller than the effect in English (9.25 in Mandarin, 27.53 in English). For L2 English, the effect of focus was significant for the last two intervals

C$_{\text{OPEN}}$ and V$_{\text{OPEN}}$. Notably, the sizes of the effects in L2 English were intermediate between the sizes of the effects in L1 English and Mandarin.

Language	Condition	CV lag	C$_{\text{CLOS}}$	C$_{\text{OPEN}}$	V$_{\text{OPEN}}$
L1 English	non-focused	36.02 (13.17)	74.39 (8.44)	86.89 (14.06)	146.90 (13.45)
	focused	41.02 (17.82)	79.73 (10.35)	102.38 (14.21)	174.08 (19.07)
L1 Mandarin	non-focused	38.18 (13.25)	81.38 (7.92)	81.58 (6.24)	135.57 (16.09)
	focused	36.40 (10.91)	83.11 (7.67)	82.93 (10.55)	145.92 (10.96)
L2 English	non-focused	43.34 (13.43)	83.01 (5.37)	96.10 (12.69)	142.25 (15.77)
	focused	46.55 (17.57)	84.12 (10.54)	103.71 (11.90)	159.33 (15.46)

Table 1. Means (standard deviations) of key interval durations (ms).

Focus estimate (ms) & significance	CV lag	C$_{\text{CLOS}}$	C$_{\text{OPEN}}$	V$_{\text{OPEN}}$
L1 English	5.52***	5.56***	16.87***	27.53***
L1 Mandarin	0.51✗	1.39✗	2.34✗	9.25***
L2 English	4.32✗	0.81✗	7.10**	15.66***

Table 2. Focus estimates and ANOVA results of LMER models.
Note: ✗ $p \leq 1$, . $p \leq .1$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

4. Discussion. The results indicate that temporal slowdown occurred in both English and Mandarin under focus. However, the two languages exhibited different patterns in the realization of focus. Since the slowdown started earlier in English than in Mandarin, we posit an earlier onset of the π-gesture. In English, the π-gesture aligns to the start of the syllable, ramping up from the onset of the syllable to maximum slowdown (π-gesture peak) at the vowel opening. In Mandarin, the π-gesture is aligned later in the syllable, possibly coupled to the start of the vowel.

The L2 speakers of English showed an intermediate pattern. The π-gesture seemed to exert its influence later in the syllable than in L1 English but earlier than in L1 Mandarin. One possible interpretation is that L1 speakers of Mandarin maintain a coupling relation between the vowel and the π-gesture even as they acquire English, adding a new coupling relation between the π-gesture and the consonant. In this way, the intermediate pattern could surface because the π-gesture is competitively coupled to both the consonant and the vowel gesture resulting in a "C-center effect" [16]), reflecting influences of both languages. The competition account can also be formalized in Dynamic Field Theory (DFT; [17, 18]). In a DFT planning model, a dynamic neural field governing the timing of the π-gesture within the syllable could stabilize under input from L1 and L2 at an intermediate value (cf. [19, 20]).

References

[1] Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). A study of sentence stress production in mandarin speakers of american english. *The Journal of the Acoustical Society of America*, *109*(4).

[2] Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors.

[3] Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from l1 mandarin and german learners of english. *Studies in Second Language Acquisition*, *36*(2).

[4] Hosseini, A. (2013). L1 interference in l2 prosody: Contrastive focus in japanese and persian. *Journal of Logic Language and Information*, *11*.

[5] Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language history questionnaire (lhq3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5).

[6] Clark, H. (1987). The stress accent in adult second language acquisition.

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

[8] Cooper, W., Eady, S., & Mueller, P. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *J. Acoust. Soc. Am.*, *77*(6).

[9] Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonol. Yearb.*, *3*.

[10] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *J. Phon.*, *27*(1).

[11] Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *J. Phon.*, *33*(2).

[12] Roessig, S., & Mu¨cke, D. (2019). Modeling dimensions of prosodic prominence. *Front. Commun.*, *4*(44).

[13] Roessig, S., Winter, B., & Mu¨cke, D. (2022). Tracing the phonetic space of prosodic focus marking [Art. no. 842546]. *Front. Artif. Intell.*, *5*.

[14] Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2).

[15] Tiede, M. (2005). Mview: Software for visualization and analysis of concurrently recorded movement data. *New Haven, CT: Haskins Laboratories*.

[16] Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*(2-4).

[17] Scho¨ner, G., & Spencer, J. P. (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press.

[18] Erlhagen, W., & Scho¨ner, G. (2002). Dynamic field theory of movement preparation. *Psychological review*, *109*(3).

[19] Stern, M. C., & Shaw, J. A. (2022). Neural inhibition during speech planning contributes to contrastive hyperarticulation. *arXiv preprint arXiv:2209.12278*.

[20] Stern, M. C., Chaturvedi, M., & Shaw, J. A. (2022). A dynamic neural field model of phonetic trace effects in speech errors. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Tonal Coarticulation in Taiwan Mandarin and Taiwan Southern Min

Po-Hsuan Huang[1], Chenhao Chiu[2]
[1, 2]*National Taiwan University*
r09142003@ntu.edu.tw, chenhaochiu@ntu.edu.tw

Background Tonal coarticulation has been found to have an asymmetric distribution, with carry-over effects being stronger than anticipatory ones, and with the former being assimilatory and the later dissimilatory, as reported in a number of tonal languages, such as Beijing Mandarin, Cantonese, Thai, and Vietnamese [1]. Some inconsistent results, however, have been found in languages like Taiwan Southern Min and Malaysian Hokkien, where the two languages were found to have more symmetric distributions [1,2]. This raises the question as to whether tonal coarticulation is a universal constraint applied to all tone languages, or can be voluntarily controlled. Several scholars support the first stance, contending that tone variation under coarticulation is likely caused by articulatory constraints (e.g., [3]). Some findings, on the other hand, point to the possibility that such variation may be manipulatable (e.g., [4]). A comparison between Taiwan Mandarin and Taiwan Southern Min is therefore a good testing ground for the two theories. On one hand, Taiwan Mandarin, as Beijing Mandarin, only has four lexical tones, and therefore is likely to have strong tonal coarticulation; on the other hand, Taiwan Southern Min has a much larger tone inventory, where the same amount of tonal coarticulation would lead to higher risk of perceptual confusion, which presumably would be avoided if it is indeed manipulatable. Otherwise, tonal coarticulation is likely not entirely voluntary.

This study thus aims to measure and compare coarticulatory effects in the two languages and to provide insights concerning the nature of tonal coarticulation.

Methods A production experiment was conducted to examine tonal coarticulation in Taiwan Mandarin and Taiwan Southern Min. The experiment quantified both carry-over and anticipatory effects in disyllabic words.

Participants & stimuli This study recruited 25 Taiwanese college students as participants (15 females; 20–27 y.o., mean=22.64), including 14 Taiwan Mandarin monolingual and 11 Taiwan-Mandarin-Taiwan-Southern-Min bilingual. The monolingual speakers produced only Taiwan Mandarin stimuli while the bilingual speakers produced both the Taiwan Southern Min and Taiwan Mandarin stimuli. A disyllabic word was chosen for each of the 16 (4 tones×4 tones, for Taiwan Mandarin) and 25 (5 tones×5 tones, for Taiwan Southern Min, with the two check tones excluded) tone combinations as stimuli. Each word was produced 10 times.

Pitch extraction F0 values were extracted using Praat [5], with the time step set at 0.01s. The F0 values were divided into 11 proportions, and the mean of the F0 values in each proportion was calculated, and then converted to z-transformed semitones for between-subject comparisons.

Analyses To quantify the magnitude of tonal coarticulation, pitch onsets and offsets were obtained, determined by the F0 means of the first and last 11 time points from each tone production. To compare Taiwan Mandarin and Taiwan Southern Min, realized tones after sandhi were converted into a five-level scale. The F0 values were then fitted through linear mixed effect models (LMM). In this study, a positive coefficient is regarded as indicator of assimilatory effects, and vice versa.

Results & discussion The results are shown below in Fig. 1. Significant coarticulatory effects were found for both carry-over and anticipatory effects (all p<.001***), and no differences of magnitudes of coarticulatory effects between Taiwan Mandarin and Taiwan Southern Min were found, be it carry-over or anticipatory (p=.31 and p=.65, respectively). Contra previous findings in Beijing Mandarin, Taiwan Mandarin had more symmetric distribution, with both carry-over and anticipatory effects being assimilatory. No differences were found between the Mandarin results of monolingual and bilingual speakers (p=.92 and p=.19, carry-over and anticipatory, respectively). The results in our study seem to support the view of [3] and others, which suggests that tonal coarticulation is likely to a large extent involuntary. As argued previously, for a language such as Taiwan Southern Min, with its large tone inventory, to have the same amount of tone variation as Taiwan Mandarin would mean that perception in continuous speech can be more challenging, as the odds of confusing a coarticulated tone with another lexical tone are bound to be higher. The fact that the magnitudes of coarticulatory effects in the two languages are not different suggests highly that such variation is likely not entirely voluntary.

It is, however, interesting to note that discrepancies indeed were found between Beijing Mandarin and Taiwan Mandarin, with the later having typologically more singular symmetric distribution just as Taiwan Southern Min. One possible explanation is that Taiwan Mandarin, as a more syllable-timed language than Beijing Mandarin, could have more even emphasis on the former and the later syllables, resulting in the more symmetric distribution of tonal coarticulation attested. Further investigation is required to test such theory.

Overall, this study investigated tonal coarticulation in Taiwan Mandarin and Taiwan Southern Min, finding that both languages, despite the large size difference of tone inventory, had similar amount of tone variation under coarticulation. This study is the first to investigate tonal coarticulation in Taiwan Mandarin, and to quantitatively compare coarticulated tones in two languages with different sizes of tone inventory. We hope to provide novel perspectives concerning the nature of coarticulation and tone variation.

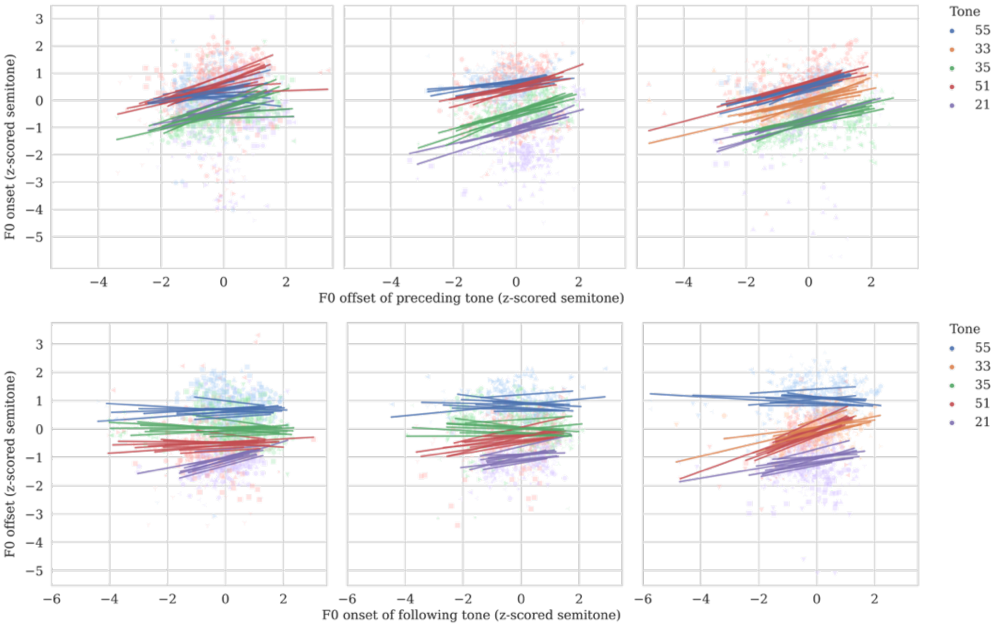


Fig. 1: Fitted LMM models of tone onsets and offsets in carryover (top) and anticipatory (bottom) positions (left: Taiwan Mandarin (monolingual); middle: Taiwan Mandarin (bilingual); right: Taiwan Southern Min)

References

[1] Chang, Y.-C., & Hsieh, F.-F. (2012). Tonal coarticulation in Malaysian Hokkien: A typological anomaly? *The Linguistic Review, 29*, 37–73. doi: 10.1515/tlr-2012-0003

[2] Wang, H. (2002). The prosodic effects on Taiwan Min tones. Language and Linguistics, 3, 839–852.

[3] Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics Monograph Series, 17*, 1–31.

[4] DiCanio, C. T. (2014). Triqui tonal coarticulation and contrast preservation in tonal phonology. In Bennett, R., Dockum, R., Gasser, E., Goldenberg, D., Kasak, R., & Patterson, P. (Eds.), *Proceedings of the Workshop on the Sound Systems of Mexico and Central America*. Yale University.

[5] Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer*. Version 6.2.14 [Computer Software]. Retrieved from http://www.praat.org/

The role of L2 English in the perception of L3 Korean sibilants by L1 speakers of French and Vietnamese

Minkyoung Hong[1] & Jeffrey J. Holliday[2]

Korea University (Korea)
[1]gophang@hanmail.net, [2]holliday@korea.ac.kr

The PAM-L2 model [1], which is based on the contrast assimilation types found in the Perceptual Assimilation Model [2], states that L2 phonological contrasts whose members perceptually assimilate to different L1 categories will be easier to discriminate. While this outcome is frequently observed (e.g., [3]), it has recently been shown that when the L2 is actually an L3, listeners may assimilate L3 sounds to either their L1 or the L2 (e.g., [4]). Given the ubiquity of English education, most "L2" learners of Korean have already, in fact, had substantial experience with English, making Korean their L3. If these listeners can assimilate Korean sounds to either their L1 or L2, exploring any differences in perceptual assimilation to their L1 and L2 can help us better understand L3 discrimination accuracy.

In this ongoing work, we tested the perceptual assimilation and discrimination of L3 Korean sibilant fricatives and affricate contrasts by L1 speakers of French and Vietnamese with respect to both their L1 and to English. Korean contrasts fortis and non-fortis sibilant fricatives /s*/ and /s/, and fortis, lenis, and aspirated alveolopalatal affricatives /ʨ*, ʨ, ʨʰ/. In terms of coronal sibilants, both French and Vietnamese contain an /s/-/z/ contrast, with French also having /ʃ/-/ʒ/. And while Vietnamese has a single affricate /ʨ/, French has none.

Three female speakers of Seoul Korean produced isolated CV monosyllables combining a coronal stop or sibilant /t*, t, t^h, s*, s, ʨ*, ʨ, ʨʰ/ with the vowels /a, i, u/. These recordings were used as stimuli in two perceptual assimilation (PA) tasks and an AXB discrimination task, administered online using Pavlovia. As analysis is still ongoing, only the results from the /a/ stimuli are presented here. In the first PA task, listeners heard the /Ca/ stimuli (n = 16) produced by two of the speakers and were asked which L1 sound (PA-L1) it was most similar to, whereas in the second PA task listeners heard the same stimuli and were asked to choose the most similar L2 (English) sound (PA-L2). In the AXB discrimination task trials (n = 40), listeners heard three /Ca/ stimuli, each one produced by a different speaker, and were asked whether the first or third sound was the same as the second.

The participants reported on here were only those participants whose English ability was assessed as high. These were advanced L3 learners of Korean whose L1 was also French (FK, n=11) or Vietnamese (VK, n=7), and also naïve listeners with no experience with Korean who were either L1 French (FN, n=11) or L1 Vietnamese (VN, n=9).

Turning first to fricatives, in the case of L1 Vietnamese, both the naïve listeners (VN) and L3 learners (VK) assimilated both /s/ and /s*/ to their Vietnamese /s/ category, but both were able to discriminate them reasonably well (71 to 83%). This can be partially explained by their PA-L2 results, in which VN listeners assimilated Korean /s/ more to English /s/, and Korean /s*/ more to English /ʃ/, but the VK listeners assimilated both Korean fricatives to English /s/ at roughly equally rates. Thus, PA-L2 seems to predict VN listeners discrimination accuracy, but that of the VK listeners is unexplained. In the case of L1 French, however, both the L3 learners (FK) and naïve (FN) listeners assimilated both Korean /s/ and /s*/ to a single /s/ category in both PA-L1 and PA-L2, and their discrimination accuracy was quite low, following PAM/PAM-L2.

Turning next to affricates, the VN listeners assimilated Korean /ʨ*/ and /ʨʰ/ to their Vietnamese /ʨ/ category, whereas Korean lenis /ʨ/ was assimilated mostly to Vietnamese /ʨ/ but also sometimes to /tʰ/. Discrimination accuracy was highest (71%) on the Korean /ʨ*-ʨʰ/ contrast, however, which again suggests a role for English: VN listeners assimilated Korean /ʨ*/ to English /ʧ/, but Korean /ʨʰ/ remained uncategorized, being perceived weakly as both English /t/ and /ʧ/. The VK listeners, on the other hand, assimilated all three Korean affricates to their Vietnamese /ʨ/ category and their English /ʧ/ category, but nevertheless exhibited very high discrimination accuracy of all three contrasts. This result is again not in line with the segment-based predictions of PAM/PAM-L2, but it is important to note that there may be a role for perceptual assimilation on a tonal level, as Korean lenis obstruents are produced with lower f0 than fortis or aspirated ones [5].

In the case of L1 French, FN listeners assimilated all three Korean affricates mostly to French /t/, whereas their assimilation to English varied: Korean /tɕ/ to English /t/, Korean /tɕʰ/ to English /ʧ/, and Korean /tɕ*/ to no single category. Discrimination accuracy ranged from 61 to 78%. The clearest difference between FN and FK listeners was that FK listeners assimilated all three Korean affricates not to French /t/, but to French /ʃ/ (/tɕ, tɕʰ/) and /ʒ/ (/tɕ*/). Their PA-L2 results were slightly different, with Korean /tɕʰ/ assimilating to English /ʧ/, Korean /tɕ*/ to English /ʤ/, and Korean /tɕ/ to both English /ʧ/ and /ʤ/. FK listeners' discrimination accuracy was highest for Korean /tɕ*-tɕʰ/, reflecting their two-category assimilation pattern in both their L1 and their L2.

Finally, the effect of L3 experience on perceptual assimilation was most visible in the perception of Korean affricates. In Vietnamese, naïve listeners assimilated them less consistently, while L3 learners perceived them nearly categorically as Vietnamese /tɕ/. Their assimilation to English sounds also differed, with naïve listeners sometimes perceiving them as /t/ whereas L3 learners perceived them only as affricates. In French, naïve listeners perceived Korean affricates mostly as stops in both French and English, whereas L3 learners perceived them as fricatives in French and either fricatives or affricates in English.

To summarize, the results of this study suggest the following. First, accuracy on the discrimination of L3 contrasts cannot be predicted only by perceptual assimilation to L1 categories. Second, considering the case here in which /ʃ/ exists in both French and English but not Vietnamese, the existence of L2 categories can be leveraged by L3 learners to discriminate between members of a difficult L3 contrast. Lastly, as illustrated by the perceptual assimilation of Korean affricates by L1 French listeners, phonological knowledge gained early on in L3 learning can influence how L3 sounds are perceptually assimilated, supporting the main claim of PAM-L2.

References

[1] Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: commonalities and complementarities. In O.-S. Bohn & M. Munro (Eds.), Second-language speech learning: *The role of language experience in speech perception and production* (pp. 13-34). John Benjamins.

[2] 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). York Press.

[3] Holliday, J. J. (2016). Second language experience can hinder the discrimination of nonnative phonological contrasts. *Phonetica*, 73(1), 33-51.

[4] Wrembel, M., Marecka, M., & Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism, 16*(4), 513-533.

[5] Lee, H., Holliday, J. J., & Kong, E. J. (2020). Diachronic change and synchronic variation in the Korean stop laryngeal contrast. *Language and Linguistics Compass*, e12374.

Syllabic perception of vowels: evidence from interlanguage

Li Junkai[1] & Lu Chen[2]

[1]*Tianjin University (China),* [2]*Shaanxi Normal University (China)*
lijunkai@tju.edu.cn, luchen@snnu.edu.cn

Existing models of L2 vowel perception (*e.g.* Flege & Bohn, 2021) predict the accuracy of interlanguage sounds by classifying "identical", "similar" or "different" phonemes. Many predictions assume that identical phonemes are the easiest to learn and similar phonemes are the most difficult to acquire. However, Li (2021) observes that even there exist identical phonemes in the vocalic inventory of the learners' L1 compared to the target language, it is difficult for them to acquire some of those vowels.

In a cognitive point of view, we suggest that the basic unit of vowel perception and recognition may not be segmental sounds, but syllables, *i.e.* words, for Chinese learners of L2, as there is a strict "syllable-morpheme" correspondence in Chinese languages (Chen *et al.*, 2002). The cognitive factor of syllables in vowel perception and recognition is the subject to some previous studies from the perspective of phoneme and allophone: whether there is an "Allophone-to-Phoneme" process followed by association with lexicon in the vowel perception and recognition process? McClelland & Elman (1986), Norris (1994), Luce, Goldinger, Auer, & Vitevitch (2000) support setting up the process, while Marslen-Wilson & Warren (1994), and Klatt (1989) take the opposite view. We argue that phoneme, allophone remains an abstract unit/concept, or a mode of analysis. In the process of vowel perception and recognition, there exists the possibility of studying vowel perception and recognition through higher unit than phoneme and allophone. Some recent studies placed greater importance on syllables (Yasufuku & Doyle, 2021), but the scope of the material and the experimental approach have focused more on drawing scenario-specific conclusions than on theoretical exploration, *i.e.*, on uncovering the theoretical value of syllables in vowel recognition.

Thus, we propose the hypothesis that phonological system is based on the lexical system. Language is a sound-meaning integrated system. As a result, the phonological system is never established independently, but along with the building of the meaning system, in the process of distinguishing or expressing meaning, of constructing morphemes or words. The segment (phone or phoneme)-morpheme-word-phrase-clause-sentence represents a hierarchical system of linguistic units from bottom to top. Syllable plays an important role as a link between segments and morphemes: segments distinguish or express meaning and construct morphemes through syllables, and morphemes are realized as segments through syllable distinctions or expressions of meaning. Therefore, compared to phoneme and allophone alone, the syllable in the cognitive process of meaning recognition may be more salient than segmental sound.

To verify this hypothesis, we first tested the perception of three marked vowels ([y], [ø] and [ɯ]) in L1 Chinese (Mandarin, Cantonese and Xi'an dialect) and in L2 (French, German and Japanese) by Chinese learners in isolated form and in syllabic forms respectively (See Tableau 1). Then we conducted acoustic analysis of the production of the these L2 vowels by L1 Chinese learners. Our results reveal that:

1) the subjects cannot fully perceive the marked vowels of L2, even they exist in their L1;

2) it is difficult for the subjects to perceive the marked vowels in isolated form than in syllabic forms, in L2 and also in L1;

3) we observe more accurate production of the L2 vowels in the syllabic forms that correspond to their native language than in the syllabic forms that do not.

Our conclusion is that vowel perception and recognition are based on syllables. The basic level of cognitive categorization of vowels is syllabic. Phonotactic structure plays a more important role than the equivalence of vocalic inventories both in perception and in production. Cognitively speaking, it is difficult for Chinese learners to manipulate sounds at a level lower than syllabic level.

Tableau 1: Recognition rate of three marked vowels in L1 and L2.

	Context	Sound	Recognition rate	Significance
L1 Mandarin	isolated	[y]	67.27%	]*
	syllabic (word)	[ly] *lü*	96.58%	
L2 French	isolated	[y]	43.93%	]*
	syllabic (word in L1)	[ly] *lu*	82.38%	]*
	syllabic (not word in L1)	[ty] *tu*	59.82%	
L1 Cantonese	isolated	[ø]	45.25%	]*
	syllabic (word)	[ʃøn] *shoen*	84.94%	
L2 German	isolated	[ø:]	35.63%	]*
	syllabic (word in L1)	[ʃø:n] *schön*	71.79%	]*
	syllabic (not word in L1)	[ø:l] *öl*	55.52%	
L1 Xi'an Dialect	isolated	[ɯ]	51.65%	]*
	syllabic (word)	[kɯ] *ku*	90.82%	
L2 Japanese	isolated	[ɯ]	38.26%	]*
	syllabic (word in L1)	[kɯ] *ku*	77.19%	]*
	syllabic (not word in L1)	[nɯ] *nu*	63.38%	

*: p>0.05

References

[1] Flege, J. E., & Bohn, O.-S. (2021) "The revised speech learning model (SLM-r)." *Second language speech learning: Theoretical and empirical progress* 10.9781108886901.002.

[2] Li, J. (2021). *Interphonologie segmentale du français dans la perspective de la théorie de l'optimalité*. PhD dissertation. Sun Yat-sen University (China).

[3] Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, 46, 751–781. https://doi.org/10.1006/jmla.2001.2825

[4] McClelland, J. L., & Elman, J. L. "The TRACE model of speech perception." *Cognitive psychology* 18.1 (1986): 1-86.

[5] Norris, D. "Shortlist: A connectionist model of continuous speech recognition." *Cognition* 52.3 (1994): 189-234.

[6] Luce, P. A., et al. "Phonetic priming, neighborhood activation, and PARSYN." *Perception & psychophysics* 62.3 (2000): 615-625.

[7] Marslen-Wilson, W., & Warren, P. "Levels of perceptual representation and process in lexical access: words, phonemes, and features." *Psychological review* 101.4 (1994): 653.

[8] Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226).

[9] Yasufuku, K., & Doyle, G. "Echoes of L1 Syllable Structure in L2 Phoneme Recognition." *Frontiers in Psychology* (2021): 1282.

Audiovisual enhancement in clear speech production of English laterals

Jonathan Havenhill, Ming Liu, Shuang Zheng & Jonah Lack

The University of Hong Kong
jhavenhill@hku.hk, {mingliu_7, ivyzheng, jlack}@connect.hku.hk

The role of auditory perceptibility in the maintenance and enhancement of phonological contrast is well established [1]. In clear speech styles, speakers are observed to increase the acoustic distance between contrastive phones to improve auditory perceptibility [2]. Non-auditory perceptual cues (notably vision) also influence speech perception [3] and have long been known to improve perceptibility under noisy conditions [4]. As such, clear speech may also involve enhancement of visible articulations, e.g., lip rounding [5]. The finding that blind speakers show less rounding than sighted speakers in clear speech [6] suggests that such modifications are at least partly mediated by visual factors. Yet while increasing the degree of lip protrusion may improve visual perceptibility, doing so simultaneously increases acoustic distance by lowering F2. In many cases it is therefore difficult to determine the extent to which speakers are optimizing their speech for auditory-acoustic and/or visual-articulatory cues.

To address this question, we investigate the production of laterals and other coronal consonants in normal and listener-oriented clear speech. In English, visibly articulated variants of /l/ have been noted to occur in lip syncing [7] and can also be observed in other clear or emphatic speech styles. Such variants have not been systematically investigated, however, so their frequency, phonetic properties, and phonological distribution, as well as their communicative function, remain unknown. The goal of this study is to examine how speakers use visible articulatory gestures in producing English laterals during clear speech, to test the hypothesis that some articulatory gestures serve a visuoperceptual rather than auditory enhancement function.

18 adult native English speakers (8 men, 10 women, mean age 28.2) participated in a two-part speech production experiment. Speakers were asked to produce a pseudo-randomly ordered list of 96 words containing /l n d θ/ in syllables with primary stress. Target syllables were balanced for vowel height (/i/ vs. /æ/) and position of the target consonant (onset, coda, monomorphemic intervocalic, pre-boundary intervocalic) positions. In the first block, speakers produced three repetitions of each word in citation form while seated alone in a sound-attenuated booth. Audio was recorded using an Earthworks Ethos condenser microphone, while high-speed (120 fps) video was recorded using a Sony DSC RX10-IV camera. In the second block, speakers repeated the wordlist in a cooperative game with a native English-speaking listener. Audio and video were recorded and simultaneously transmitted to the listener (seated in another room) over Zoom. Ambient recordings from a noisy bar were overlaid at −9dB SNR, such that the speaker's speech was effectively unintelligible. The speaker was instructed to produce three repetitions of each word as clearly as possible, while the listener attempted to guess the word after each repetition. The speaker clearly heard the listener's guesses with the same ambient noise at +3dB SNR.

Each token was visually coded according to its articulatory configuration. Non-visible (NV), dental (D), visible alveolar (V), interdental (ID), and linguolabial (LL) configurations were observed, as shown in Figure 1. Frequency of each variant is provided in Figure 2a. The typical realization of /θ/ in normal speech was interdental (69.2%) or dental (19.9%). For /d/ and /n/, over 98% of tokens were realized with non-visible articulations in normal speech, with only a handful of dental and visible alveolar tokens in the clear speech task. In contrast, /l/ was produced with both dental (11.9%) and interdental (5.4%) articulations, even in normal speech. Multinomial logistic regression analysis indicates significantly higher rates in clear speech of dental and interdental variants for /l/ ($p < 0.001$), dental and visible variants of /n/ ($p < 0.001$) and interdental variants of /θ/ ($p < 0.001$). Linguolabial variants of both /θ/ and /l/ also occur in clear speech, albeit rarely, but were never observed in normal speech. Visibly articulated variants of /l/ occur in all syllable positions, as in Figure 2b, but with some variability by vowel height.

Acoustic analysis indicates that visibly articulated variants of /l/ (particularly ID and LL) do not consistently preserve /l/-like acoustic features. As seen in Figure 3, interdental [l̪] not only lacks the characteristic resonance patterns of alveolar [l] (3a), but also exhibits frication, particularly toward the vowel onset (3b), yielding an audibly [ð]-like sound. This finding suggests that listener-oriented speech does not necessarily prioritize auditory perceptibility.

Rather, speakers may choose to provide listeners with a direct visual cue of a segment's articulatory/gestural properties, e.g., tongue lengthening and narrowing for /l/. However, this strategy is not available for all segments, suggesting that speakers recruit visually exaggerated gestures when such gestures a) also occur in normal speech and b) are visually distinct from similar, contrasting segments. These results call into question the hypothesis that the objects of speech perception are primarily auditory and suggest that speakers may prioritize visual perceptibility, consistent with the inherently multimodal nature of speech perception [8]. Theories of contrast and representation must therefore incorporate both auditory and non-auditory perceptibility.

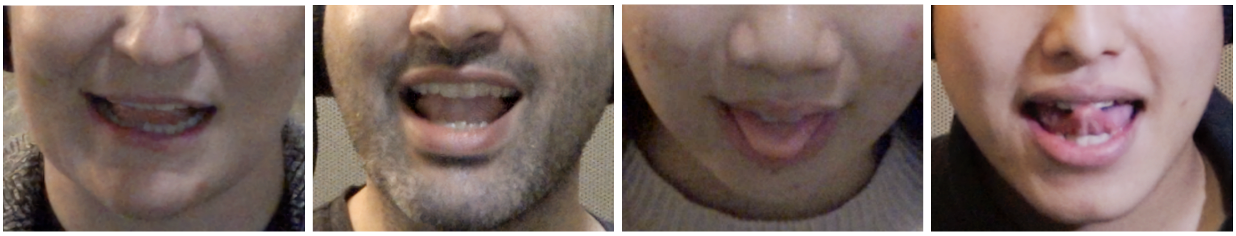


Figure 1: From left to right: dental (D), visible alveolar (V), interdental (ID), and linguolabial (LL) productions of /l/ in clear speech task.

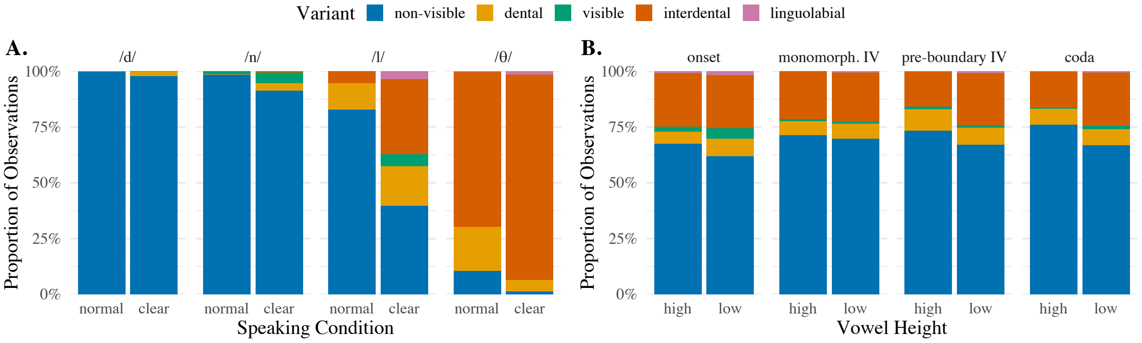


Figure 2: (A) Variants of /d n l θ/ observed in normal and clear speaking tasks. (B) Variants of /l/ by syllable position and vowel quality.

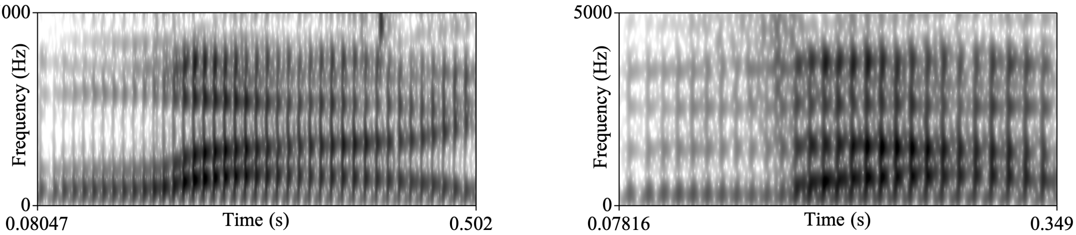
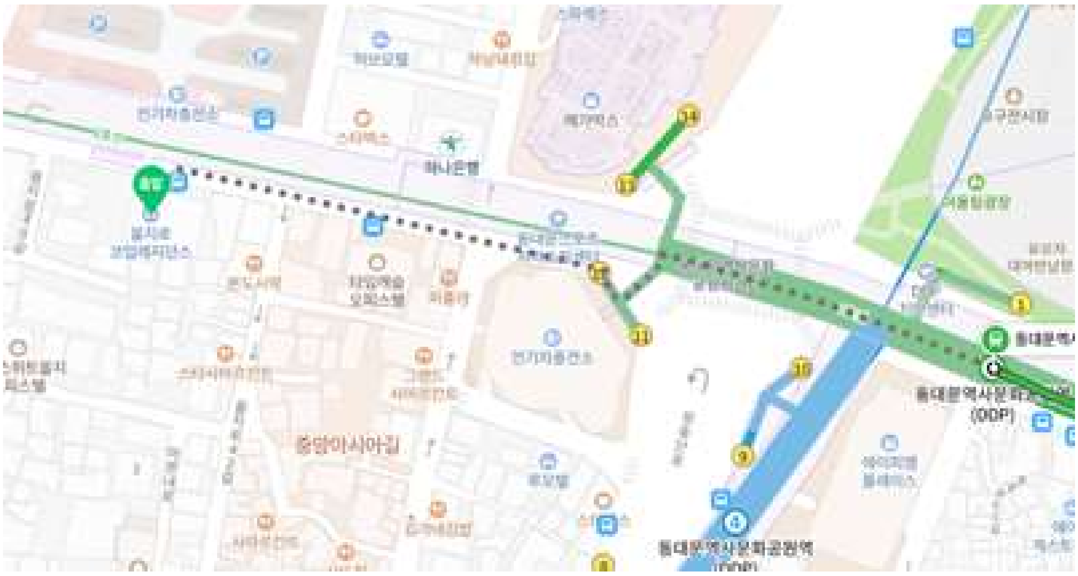


Figure 3: Spectrograms for a) non-visible [læg] (left) and b) interdental [l̪æp] in clear speech.

[1] R. L. Diehl and K. R. Kluender. "On the objects of speech perception". In: *Ecol. Psychol.* 1.2 (1989), pp. 121–144. **[2]** R. Smiljanić and A. R. Bradlow. "Speaking and hearing clearly: Talker and listener factors in speaking style changes". In: *Lang. Linguist. Compass* 3.1 (2009), pp. 236–264. **[3]** J.-P. Gagné, A.-J. Rochette, and M. Charest. "Auditory, visual and audiovisual clear speech". In: *Speech Commun.* 37.3-4 (2002), pp. 213–230. **[4]** W. H. Sumby and I. Pollack. "Visual Contribution to Speech Intelligibility in Noise". In: *JASA* 26.2 (1954), pp. 212–215. **[5]** J. Havenhill. "Constraints on Articulatory Variability: Audiovisual Perception of Lip Rounding". PhD thesis. Georgetown Univ., 2018. **[6]** L. Ménard et al. "Speaking Clearly for the Blind: Acoustic and Articulatory Correlates of Speaking Conditions in Sighted and Congenitally Blind Speakers". In: *PLOS ONE* 11.9 (2016), e0160088. **[7]** M. Liberman. *Apico-labials in English*. Language Log, Accessed: 2023-02-10. 2010. **[8]** L. D. Rosenblum. "Primacy of Multimodal Speech Perception". In: *The Handbook of Speech Perception*. Ed. by D. B. Pisoni and R. E. Remez. Oxford: Blackwell, 2005, pp. 51–78.

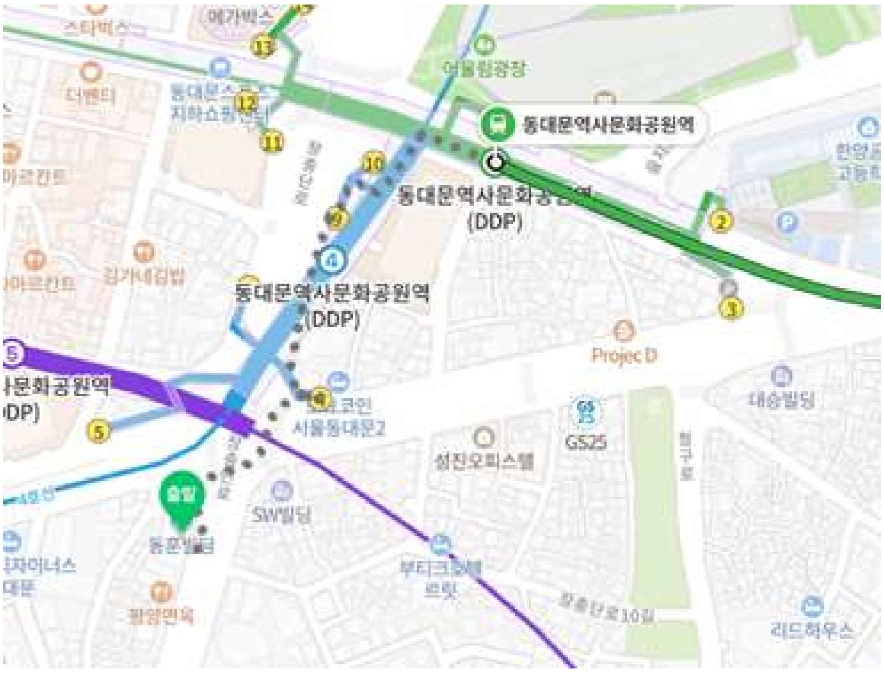
FROM THE HOTELS TO THE CONFERENCE

FROM ULJIRO CO-OP RESIDENCE:



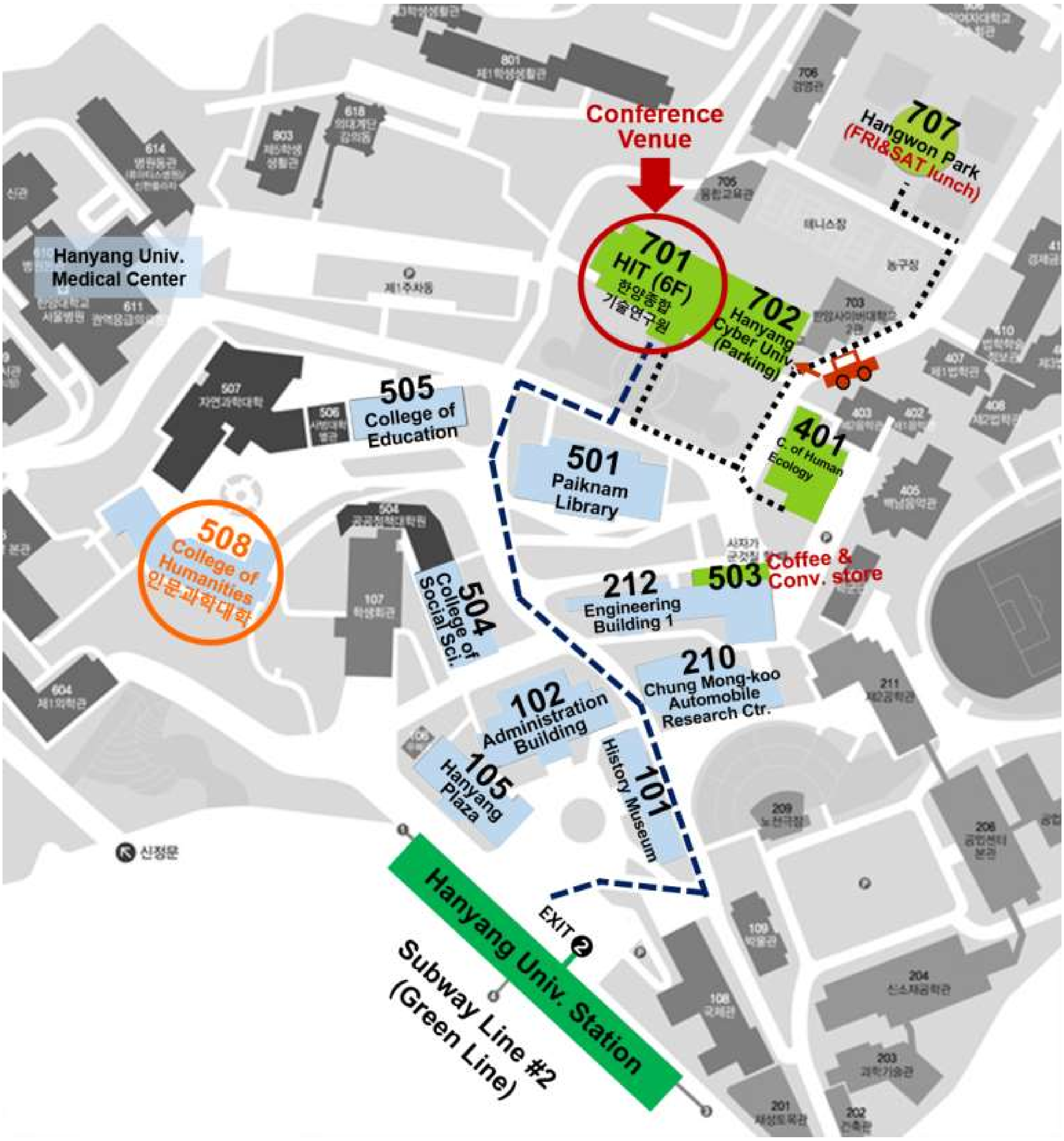
1. Enter Dongdaemun History & Culture Park Station from Entrance 12.
2. Take Subway Line 2 (🟩) towards Sindang, Seongsu and Wangsimni station.
3. Exit after four stops at Hanyang University Station (exit 2).
4. Walk about 10 minutes to the Hanyang University HIT Building.

FROM GUESTHOUSE DONGDAEMUN PREMIUM:



1. Enter Dongdaemun History & Culture Park Station at entrance 4.
2. Take Subway Line 2 (🟩) towards Sindang, Seongsu and Wangsimni station.
3. Exit after four stops at Hanyang University Station (exit 2).
4. Walk about 10 minutes to the Hanyang University HIT Building.

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.
- ➢ 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.

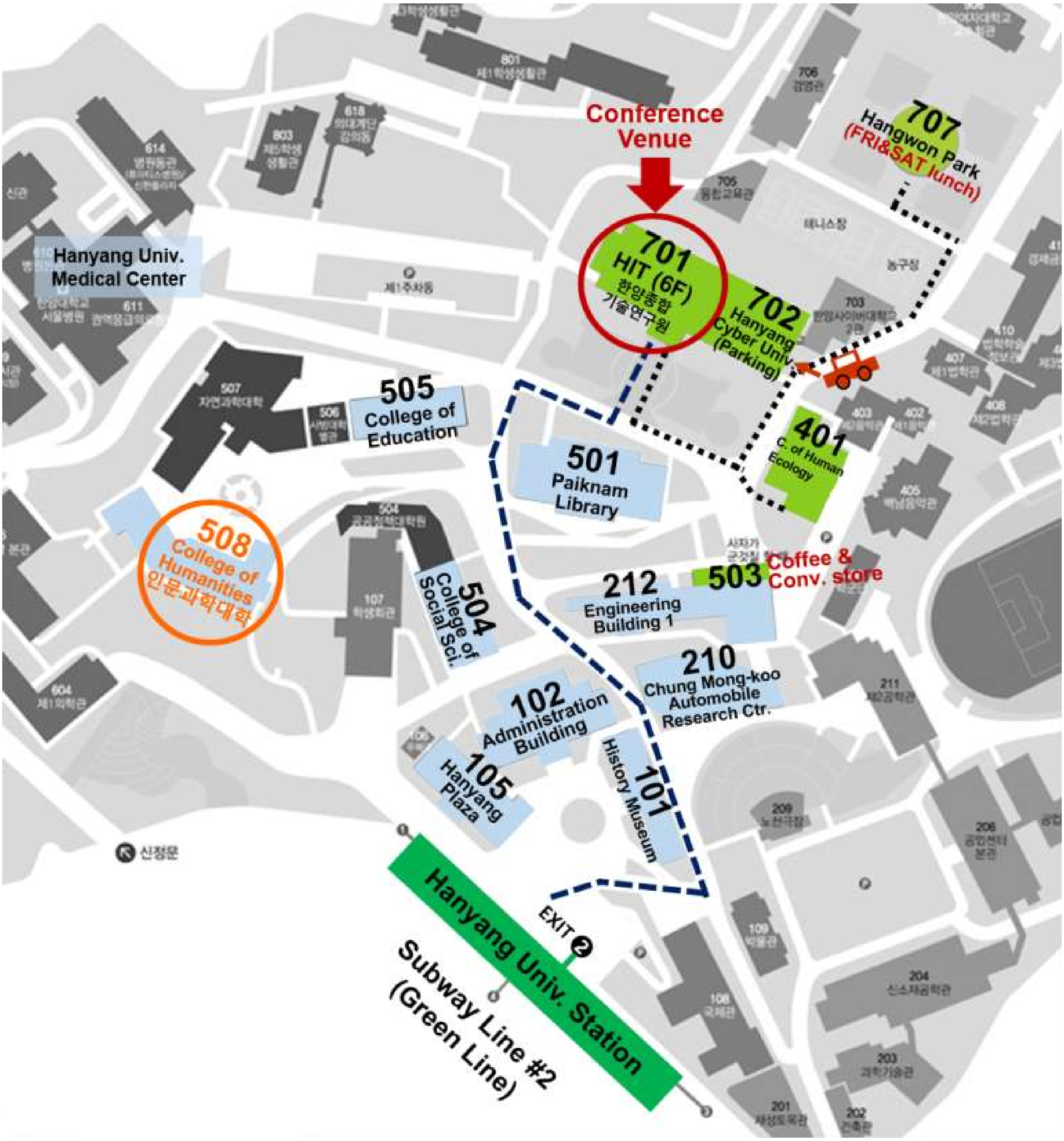
HISPhonCog 2023: Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2023

●

May 25, 2023.
Hanyang University, Seoul, Korea

인쇄일	:	2023년 5월 25일
발행일	:	2023년 5월 25일
발행인	:	조태홍
발행처	:	한양 음성•언어인지과학연구소

Printing Date	:	May 25, 2023
Publishing Date	:	May 25, 2023
Publisher	:	Cho, Taehong
Publishing Place	:	Hanyang Institute for Phonetics and Cognitive Sciences of Language



비매품/무료
95700
9 791198 325907
ISBN 979-11-983259-0-7 (PDF)