TAI 2023

The Second International Conference on Tone and Intonation, 18-20 Nov 2023, Singapore

Proceedings of the Conference

November 18 – 20, 2023

Order copies of this and other COLIPS proceedings from:

Chinese and Oriental Languages Information processing Society (COLIPS)
1 Fisonopolis Way
21-01 Connexis
Singapore 138632
Tel: +65-64082757
colips@asianlp.sg

ISBN

Welcome Message from the Conference Chair

It is with immense pleasure that I welcome you to the vibrant city of Singapore for the Second International Conference on Tone and Intonation (TAI 2023). This event marks a significant gathering of minds in the realms of linguistics and speech research, under the intriguing theme "East Meets West: Languages and Approaches".

I am deeply honored to introduce our esteemed organizers, the Pattern Recognition and Machine Intelligence Association (PREMIA) and the Chinese and Oriental Languages Information Processing Society (COLIPS). Their dedication and hard work have been the driving force behind the success of this conference.

A heartfelt thank you to our Steering Committee, whose wisdom and guidance have been invaluable. Their contributions have significantly shaped the direction and quality of this conference.

Our gratitude extends to our sponsors, the International Speech Communication Association (ISCA), the International Phonetic Association (IPA), and ISCA SProSIG. Their generous support has been crucial in facilitating this global exchange of knowledge. Special thanks to ISCA for providing grants that have enabled students and young researchers to join us here.

I would like to express my deepest appreciation to our keynote speakers - Jennifer S. Cole from Northwestern University, James Kirby from Ludwig-Maximilians-Universität München, Katie Franich from Harvard University, and TAN Ying Ying from Nanyang Technological University. Their groundbreaking research and insights into tone, intonation, and linguistic dynamics are a cornerstone of this conference.

The local organizing committee and our diligent student volunteers have worked tirelessly to ensure a smooth and enriching conference experience. Their commitment and hard work do not go unnoticed, and we are all grateful for their contributions.

Additionally, a special acknowledgment to the Singapore Tourism Board for their local support, helping make our stay in this beautiful city not only productive but also enjoyable.

Our Program Chairs, Dr Yanfeng Lu and Prof Ying Chen, Finance Chair Dr Yan Wu, Organization Chairs Dr Lei Wang and Dr Min Yuan, and Publication Chair Dr Ridong Jiang, deserve special mention for their impeccable planning and execution.

As we embark on this journey of academic and cultural exchange, I encourage everyone to engage in lively discussions, forge new collaborations, and explore the myriad of perspectives presented.

Once again, welcome to Singapore and TAI 2023. Let us make this a memorable and productive conference.

Minghui Dong, Institute for Infocomm Research, A*STAR, Singapore
Conference Chair, TAI 2023

Message from the Program Chairs

Welcome to Singapore for the Second International Conference on Tone and Intonation (TAI 2023). As we gather under this year's theme, "East Meets West: Languages and Approaches," we are set to explore the rich and diverse world of linguistic studies in an international context.

We are pleased to share that this year's conference has achieved a remarkable global reach. We received 86 submissions from 19 countries/regions across Asia, Europe, and America. Following a rigorous review process, 59 papers have been accepted, reflecting the high quality of research in our field. Of these, 39 will be presented in 8 oral sessions and 20 in 2 poster sessions, allowing a diverse range of ideas and research findings to be showcased.

Our program covers a wide range of intriguing topics such as Pitch and Pitch Accent, Prosodic Analysis and Modeling, Tonal Variation and Change, Bilingual and L2 Prosody, Neurolinguistics and Tone Perception, Computational Approaches to Tone and Intonation, and Sociolinguistic Aspects of Tone and Intonation. The presentation sessions promise to provide stimulating and insightful discussions.

We are also honored to have keynote speeches from Prof. Jennifer S. Cole, Prof. James Kirby, Prof. Katie Franich, and Prof TAN Ying Ying. Their presentations are sure to enrich our understanding and provoke thought within the overarching theme of the conference.

Our deepest thanks go to our program committee members for paper review and session chairs for organizing and guiding these sessions. Their expertise and dedication have been fundamental in creating a dynamic and comprehensive program.

We encourage all attendees to immerse themselves in the sessions, engage with the research presentations, and connect with colleagues from around the globe. Let us make TAI 2023 a landmark event for advancing our collective knowledge and fostering international collaboration in the study of tone and intonation.

Looking forward to an enriching and memorable conference experience.

Yanfeng Lu, Institute for Infocomm Research, A*STAR, Singapore
Ying Chen, Nanjing University of Science and Technology, China
Program Chairs, TAI 2023

Organizing Committee

Conference Chair

- Minghui Dong, Insititute for Infocomm Research, A*Star, Singapore

Program Chairs

- Yanfeng Lu, Insititute for Infocomm Research, A*Star, Singapore
- Ying Chen, Nanjing University of Science and Technology, China

Finance Chair

- Yan Wu, Insititute for Infocomm Research, A*Star, Singapore

Organization Chairs

- Lei Wang, Huawei International, Singapore
- Min Yuan, National University of Singapore

Publication Chair

- Ridong Jiang, Institute for Infocomm Research, A*Star, Singapore

Program Comittee

Bistra Andreeva	Saarland University
Francesco Cangemi	University of cologne
Xiaocong Chen	The Hong Kong Polytechnic University
Yiya Chen	Leiden University
Yu-Fu Chien	Fudan University
Poh Shin Chiew	University Malaya
Adam Chong	Queen Mary University of London
Elisabeth Delais-Roussarie	Université de Nantes
Emily Elfner	York University
Hui Feng	Tianjin University
Yan Feng	Nanjing University of Science and Technology
Sonia Frota	University of Lisbon
Jian Gong	Jiangsu University of Science and Technology
Amalesh Gope	Tezpur University
Stella Gryllia	Radboud University
Wentao Gu	Nanjing Normal University; McGill University
Jia Guo	College of Foreign Languages, Nankai University
Luke Horo	Living Tongues Institute for Endangered Languages
Jose Hualde	University of Illinois at Urbana-Champaign
Yishan Huang	University of Sydney
Gwendolyn Hyslop	The University of Sydney
Shinichiro Ishihara	Lund University
Xiaoli Ji	Nanjing Normal University
Pingping Jia	Free University of Berlin
Sujinat Jitwiriyanont	Chulalongkorn University
Sun-Ah Jun	UCLA, Department of Linguistics
Constantijn Kaland	University of Cologne
Tsukada Kimiko	Macquarie University; The University of Melbourne
James Kirby	Institute of Phonetics and Speech Processing
Bjoern Koehnlein	The Ohio State University
Haruo Kubozono	National Institute for Japanese Language and Linguistics
Albert Lee	The Education University of Hong Kong
Seunghun Lee	International Christian University
Lishan Li	Beijing Normal University
Shanpeng Li	Nanjing University of Science and Technology
Ya Li	Zhejiang Ocean University
Bijun Ling	Tongji University
Boquan Liu	Shanghai Jiao Tong University
Katalin Mády	Hungarian Research Centre for Linguistics
Laura McPherson	Dartmouth College
Farah Mehdawi	Hashemite University
Peggy Mok	The Chinese University of Hong Kong
Claire Nance	Lancaster University
Oliver Niebuhr	University of Southern Denmark
Neetesh Kumar Ojha	Indian Institute of Technology Guwahati
Eric Pelzl	The Pennsylvania State University
Bert Remijsen	University of Edinburgh
Tomas Riad	Stockholm University

Krisangi Saikia	Indian Institute of Technology, Guwahati
Heiko Seeliger	University of Cologne
Alif Silpachai	Radboud University
Ping Tang	Nanjing University of Science and Technology
Hiroto Uchihara	Tokyo University of Foreign Studies
Maria del Mar Vanrell Bosch	Universitat de les Illes Balears
Yi Xu	University College London
Mengzhu Yan	Huazhong university of science and technology
Bei Yang	Sun Yat-sen University
Yang Yang	Guangdong University of Foreign Studies
Margaret Zellers	Kiel University
Hongming Zhang	Macau University of Science and Technology

Table of Contents

Cross-language perception of the Japanese singleton/geminate contrast: Case of Vietnamese speakers with and without Japanese language experience

Kimiko Tsukada[1, 2], Đích Mục Đào[3] & Trang Le Thi Huyen[4]
[1]*Macquarie University,* [2]*The University of Melbourne,* [3]*University of Social Sciences and Humanities, Vietnam National University - Ho Chi Minh City,* [4]*Waseda University*
Kimiko.tsukada@gmail.com, dich.daovns@gmail.com, huyentrangle295@gmail.com

Japanese, which is a popular foreign language in both East and West, uses durational variation contrastively for both vowels and consonants. For example, *yoka* 'leisure' contrasts with *yooka* 'eight days' on the one hand and with *yokka* 'four days' on the other hand. It is widely acknowledged that length contrast is difficult for non-native speakers from diverse L1 (first language) backgrounds including Vietnamese, which is the target language of this study. Unlike Japanese, consonant length (i.e., short/singleton vs long/geminate) is not contrastive in Vietnamese. Thus, we were interested in how different experience with Japanese may affect Vietnamese speakers' perception of difficult Japanese contrasts.

Vietnam is currently ranked within the top 6 countries/regions of the world in terms of the number of learners (169,582) of Japanese (The Japan Foundation). Within Japan, Vietnam (31,643 or 14.4%) is the second largest country of origin of non-native learners of Japanese after China (67,027 or 30.5%) as of 2022 (Agency for Cultural Affairs, Government of Japan). As such, improving our understanding of how to facilitate pronunciation pedagogy for Vietnamese learners of Japanese is significant.

We examined the perception of Japanese singleton/geminate contrast by three groups of Vietnamese speakers and a control group of 10 Japanese speakers. Two of the three Vietnamese groups consisted of learners of Japanese with one group participating in the study in Ho Chi Minh City, Vietnam and the other in Tokyo, Japan. The former consisted of 17 learners who were at the N3 level of the Japanese Language Proficiency Test (according to which, the easiest level is N5 and the most difficult level is N1). The latter consisted of 13 learners who had all passed N1 and were considered highly advanced learners of Japanese. The third Vietnamese group consisted of 12 participants who were recruited in Ho Chi Minh City, Vietnam and were inexperienced in Japanese.

Table 1 shows 12 Japanese word pairs audio-recorded by six (3 males, 3 females) L1 Japanese speakers. The speech materials (/(C)V**C(C)**V/ tokens) were arranged in 200 unique triads. Excluding the first 8 trials for practice, the triads contained 96 singleton or 96 geminate tokens intervocalically (underlined and bolded). Only tokens with stops were considered in this study. As voiced geminates are limited in Japanese ([2-4]), only voiceless stops (/t, k/) were used. On average, the closure durations were 96 ms and 262 ms for singletons and geminates, respectively. The geminate-to-singleton ratios were 2.7 for alveolars (/t/-/t:/) and 2.8 for velars (/k/-/k:/), respectively.

The participants responded to 200 trials via a two-alternative forced-choice AXB discrimination task. The presentation of the stimuli and the collection of perception data were controlled by the PRAAT program ([1]). The participants were given two ('A', 'B') response choices on the computer screen. They were asked to select the option 'A' if they thought that the first two tokens in the AXB sequence were the same (e.g., '$yoka_2$'-'$yoka_1$'-'$yokka_3$') and to select the option 'B' if they thought that the last two tokens were the same (e.g., '$soto_3$'-'$sotto_1$'-'$sotto_2$'; where the subscripts indicate different speakers). No feedback was provided during the experimental sessions. The participants could take a break after every 50 trials if they wished. The participants were required to respond to each trial, and they were told to guess if uncertain.

The overall mean discrimination accuracy was 67%, 80%, 91% and 99% for the non-learner group, the learner group in Vietnam, the learner group in Japan and the Japanese group, respectively (Figure 1). All between-group differences were statistically significant [$t(11.3 – 26.6) = -5.9 – 8.7$, $p < .05$]. The advanced learners more closely resembled the Japanese speakers than did the participants in Vietnam not only in their overall accuracy but also in details when factors such as the length category and place of articulation of the target token (X in the AXB sequence) were taken into consideration. A clear difference between the two learner groups in Japan and Vietnam demonstrates learnability of Japanese consonant length beyond early childhood in an immersion setting. At the same time, a non-negligible difference between the L1 Japanese and the highly advanced learner groups suggests genuine difficulty of Japanese consonant length, which has a pedagogical implication.

Proceedings of the Second International Conference on Tone and Intonation, pages 1–2
November 18–20, 2023.

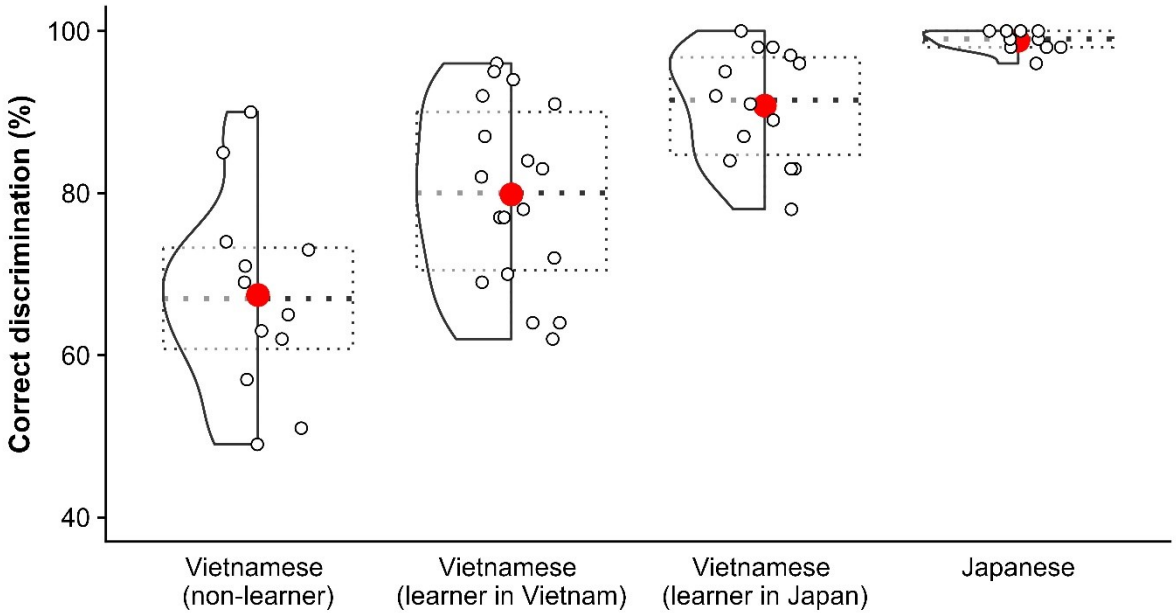


Figure 1: *The distributions of discrimination accuracy (%) by four groups of participants. The horizontal line and the red circle in each box indicate the median and mean, respectively. The bottom and top of the box indicate the first and third quartiles.*

Table 1: *Twelve pairs of Japanese words used with target sounds underlined and bolded.*

	Singleton		**Geminate**	
/t/	He**t**a	'unskilled'	He**tt**a	'decreased'
	Ka**t**o	'transient'	Ka**tt**o	'cut'
	Ma**t**e	'wait'	Ma**tt**e	'waiting'
	O**t**o	'sound'	O**tt**o	'husband'
	Sa**t**e	'well, then'	Sa**tt**e	'leaving'
	Wa**t**a	'cotton'	Wa**tt**a	'broke'
/k/	A**k**e	'open'	A**kk**e	'appalled'
	Ha**k**a	'grave'	Ha**kk**a	'mint'
	I**k**a	'below'	I**kk**a	'lesson one'
	Ka**k**o	'past'	Ka**kk**o	'parenthesis'
	Sa**k**a	'slope'	Sa**kk**a	'author'
	Shi**k**e	'rough sea'	Shi**kk**e	'humidity'

References

[1] Boersma, P., & Weenink, D. 2016. *Praat: doing phonetics by computer.* Computer program.

[2] Hussein, Q., & Shinohara, S. 2019. Partial devoicing of voiced geminate stops in Tokyo Japanese. *Journal of the Acoustical Society of America* 145, 149-163.

[3] Kawahara, S. 2015. The phonetics of sokuon, or geminate obstruents. In Kubozono, H. (Ed.), *Handbook of Japanese Phonetics and Phonology.* Walter de Gruyter, 43-78.

[4] Sano, S. 2019. The distribution of singleton/geminate consonants in spoken Japanese and its relation to preceding/following vowels. *Proceedings of the 19th International Congress of Phonetic Sciences* (Melbourne, Australia), 1833-1837.

OCP and downstep in Japanese

Manami Hirayama[1], Hyun Kyung Hwang[2] & Takaomi Kato[3]
[1]*Seikei University,* [2]*University of Tsukuba,* [2]*Sophia University*
hirayama@fh.seikei.ac, hwang.kyung.gu@u.tsukuba.ac.jp, t-kato@sophia.ac.jp

The Obligatory Contour Principle (OCP), originally proposed in tonal phenomena (e.g., [1, 2], et seq.), has been argued to account for other phenomena, where identical elements are avoided to occur adjacently within a domain. The OCP has been said to be at work in Japanese segmental phonology as well, in cases such as the avoidance of multiple voiced obstruents in *rendaku* (e.g., [3]) and obstruent devoicing in loanword phonology ([4]). In Japanese prosody, OCP was proposed to be a potential blocking factor in downstep (e.g., [5]) to account for the fact that the process was blocked in a sequence of two adjective phrases with the same ending *-i* independently modifying a head noun ([A-*i* [A-*i* N]]) in one study [5] while a combination of two adjacent adjectives with different endings in the same structure ([A-*i* [A-*na* N]]) did not block downstep in another study [6]. In an acceptance judgment experiment [7, n. 16], the hypothesis was supported that a phrase with adjectives with the same endings adjacently would be less favoured than a phrase with adjectives with different endings.

This paper investigates whether OCP plays a role in downstep in Japanese through a production experiment. Downstep in Japanese is a process where an accented word triggers the pitch register of the following phrase to be lowered within the domain of Major Phrase (e.g., [8, 9, 10]). In Figure 1(a), the first accented phrase *shirói* (acute accent mark indicates lexical accent) triggers the next phrase *nagái* to be downstepped, while in 1(b), the second phrase *nagái* is not downstepped since the first phrase *amai* is unaccented. This study uses four phrase types as in [7], with all combinations of two adjectives with endings *-i* and *-na* modifying the head noun (i.e., [A-*i* [A-*i* N]], [A-*i* [A-*na* N]], [A-*na* [A-*i* N]], and [A-*na* [A-*na* N]]). If the OCP is in effect in downstep inhibition, a phrase with the same ending in a sequence ([A-*i* [A-*i* N]], [A-*na* [A-*na* N]]) might disfavour downstep to occur while downstep would not be blocked in the other two phrases ([A-*i* [A-*na* N]], [A-*na* [A-*i* N]]).

Two items were prepared for each of the four phrase types, for both accented and unaccented triggers (e.g., [*aói* [*hirói ie*]] 'blue large house' for the former, [*kurai* [*hirói ie*]] 'dark large house' for the latter), totalling 16 test phrases. Put in a carrier phrase *ane-wa _ to itta* '(my) sister said _', and together with 16 distractor sentences, the sentences were near-randomized, and 8 versions were prepared. 10 speakers (Tokyo accent system) (age 20 to 22) participated. Excluding 17 sentences due to mispronunciation, 1,263 sentences remained for the analysis. Annotation of phrase boundaries were manually made with Praat [12] and pitch peak of each phrase was automatically measured by using *ProsodyPro* [13].

Figure 2 shows boxplots with the mean peak f0s in the target phrase in the four phrase types. On average, the peak f0 of the target phrase with the accented trigger (red) is lower than that with the unaccented trigger (blue), suggesting that downstep occurred in all four types. A statistical analysis was conducted using a linear mixed-effects model with R [14] and *lmerTest* package with the dependent variable being the peak f0 value and predictors being accent of trigger (accented vs. unaccented), phrase type (*ii* vs. *ina* vs. *nai* vs. *nana*), and the interaction of these two. Speaker and item were included as random effects in the model. A Wald chi-square test, using the *Anova* function of the *car* package [15], determined that the factors accent and type are (marginally) significant, while the interaction is not (Table 1). Pairwise comparisons revealed that no pairs (e.g., *ii* vs. *ina*) were significantly different. These suggest that overall downstep occurred in the same way in all four types and OCP is not in effect.

However, if item was removed from the model, the interaction of the accent and type is marginally significant in a Wald chi-square test (Table 2), indicating that the effect of accent difference is different in different phrase types. Pairwise comparisons indicate that all pairs except *ii* vs. *ina* and *nai* vs. *nana* are significantly different, indicating that *ii* is individually different from *nai* and *ina*. Also, in the results showing the estimates, all except the interaction of accU:type*ina* are (marginally) significant: the peak f0 in *ii* is different from that in *ina*, *nai* or *nana* individually in the accented trigger phrases as well. If this result is more representative of the population, the degree of downstep is different among the phrase types: *ii* has the smallest degree, *nana* the largest, and *ina* and *nai* situating in between. OCP might partially explain the smallest degree of pitch lowering in [A-*i* [A-*i* N]], though given the different statistical results depending on the treatment of item, it is not conclusive at this point whether this is truly representing the behaviour of the phrase types in general.

Proceedings of the Second International Conference on Tone and Intonation, pages 3–4
November 18–20, 2023.

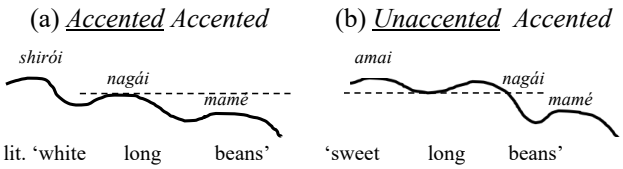


Figure 1: *Schematic pitch curves in downstep* (adopted from [5, 11]).

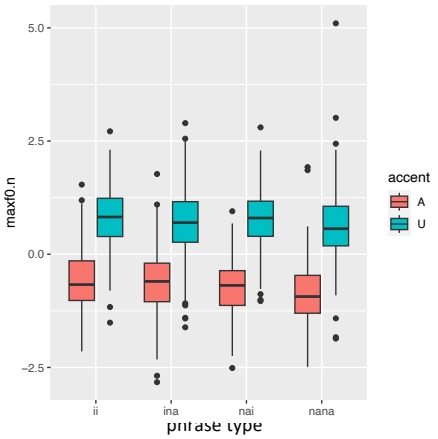


Figure 2: *Boxplots of peak f0s (y-axis) in the four phrase types (x-axis).*

Table 1: *Chi-square test in the linear mixed-effects model.*

predictor	χ^2	*df*	*p*-value
Accent	96.1104	1	< 0.001 ***
Type	7.7724	3	0.050
Accent*type	2.4325	3	0.487

Table 2: *Chi-square test without item.*

predictor	χ^2	*df*	*p*-value
Accent	113.2320	1	< 0.001 ***
Type	22.2844	3	< 0.001 ***
Accent*type	7.0271	3	0.071

References

[1] Leben, W. 1973. *Suprasegmental Phonology*. Doctoral dissertation, MIT.
[2] Goldsmith, J. 1976. *Autosegmental Phonology*. Doctoral dissertation, MIT.
[3] Kawahara, S. & Sano, S. 2014. Identity Avoidance and Rendaku. *Proceedings of the Annual Meeting on Phonology 2013*, DOI: https://doi.org/10.3765/amp.v1i1.23
[4] Kawahara, S. & Sano, S. 2016. /p/-driven geminate devoicing in Japanese: Corpus and experimental evidence. *Journal of Japanese Linguistics* 32, 57–77.
[5] Hwang, H. K. & Hirayama, M. 2021. Downstep in Japanese revisited: Morphology matters. *NINJAL Research Papers* 21, 15–23.
[6] Kubozono, H. 1991. Modeling syntactic effects on downstep in Japanese. In: *Papers in Laboratory Phonology II*. Cambridge University Press, 368–387.
[7] Hirayama, M., Hwang, H. & K., Kato, T. 2022. Lexical category and downstep in Japanese. *Languages* 7, 25.
[8] Kubozono, H. 1989. Syntactic and rhythmic effects on downstep in Japanese. *Phonology* 6, 39–67.
[9] Pierrehumbert, J. & Beckman, M. 1988. *Japanese Tone Structure*. MIT Press.
[10] Poser, W. 1984. *The Phonetics and Phonology of Tone and Intonation in Japanese*. Doctoral dissertation, MIT.
[11] Hwang, H. K., Hirayama, M. & Kato, T. 2022. Perceived prominence and downstep in Japanese. *Proceedings of Interspeech 2022*, 908.
[12] Boersma, P. & Weenink, D. 2022. Praat: doing phonetics by computer [Computer program]. http://www.praat.org/
[13] Xu, Y. 2013. ProsodyPro — A tool for large-scale systematic prosody analysis. In: *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France, 7–10.
[14] R Core Team. 2022. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
[15] Fox, J. & Weisberg, S. 2019. *An R Companion to Applied Regression*, 3rd edition. Sage.

The perception of German wh-phrase-final intonation: a contour clustering evaluation

Heiko Seeliger[1], Anne Lützeler[1] & Constantijn Kaland[1]
[1]*University of Cologne*
<heiko.seeliger, anne.luetzeler, ckaland>@uni-koeln.de

This study is a follow-up to [1], which presented the results of a cluster-based analysis (cf. [2]) of a large experimental corpus of German *wh*-phrases (read speech, elicited for [3]), focusing particularly on utterance-final pitch. In [1], special attention was paid to two clusters, termed here F and L, which consisted of rises to medium-high plateaus. The difference between the two was that F featured many late falls, while the contours in L remained level.

These two clusters were selected for validation in a perception study. These contours are interesting because both types are hard to account for in the standard GToBI inventory [4] – late, boundary tone-related falls are explicitly not part of the model, while medium-high, non-rising plateaus are, but only with the function of a continuation rise. The medium-high plateaus in this data set intuitively do not function as continuation rises, however. If these clusters are found to be distinct in perception, we cannot exclude the possibility that they might form two different intonational categories in the standard GToBI system, although more work is needed to confirm this (cf. [5] on a medium-high plateau that is formally similar to but functionally distinct from a continuation rise).

In order to keep the number of stimuli manageable, we selected 10 F0 contours from each cluster. We picked the 10 most prototypical contours from each cluster, i.e. the contour that deviated from their cluster mean the least. Stimuli were resynthesized in Praat [6] from interpolated, smoothed Pitch objects. The resulting stimuli contained only F0, harmonics and (static) formants, but no segmental information. Duration of the original stimuli was left as-is.

21 listeners participated in the experiment. Stimuli were presented using OpenSesame [7], in randomized order. Each stimulus from one cluster was paired with each stimulus from the other cluster. Two stimuli from each cluster were paired with each member of their own cluster (one of them also having one cross-cluster comparison). While unintended, these stimuli served as fillers and controls. Participants could listen to each stimulus pair as often as required. They were then asked to rate how similar the two contours sounded on a five-point scale. We translated the ratings to a numeric scale (1 = 'very dissimilar'; 5 = 'very similar').

An overview of the results split up by across- vs. within-cluster comparisons is shown in Fig. 1. Fig. 2 and 3 correlate mean ratings and acoustic properties of contour pairs – difference in mean F0 and in duration. A striking finding is that participants were sensitive to the differences in speaker gender despite the low-pass filtering of the stimuli. In Fig. 2, the between-gender pairs are the L-L comparisons and those with an F0 difference of greater than 6 ST. Note that an F0 difference between contours of less than 6 ST was a necessary, but not sufficient requirement for high perceptual similarity: The between-gender L-L comparisons were rated as dissimilar despite differing by less than 6 ST. Ultimately, we cannot say for this cluster whether its members were truly dissimilar, since all comparisons here were between-gender. For the F-F comparisons, we can also see a gender effect: between-gender pairs are in the bottom right of Fig. 1, while within-gender (and, as it happens, within-speaker) pairs are in the top left. Fig. 3 shows that larger duration differences led to lower similarity ratings, perhaps best visible from the outlier among the very similar F-F pairs.

Despite these overall correlations between general acoustic properties and perceived similarity, the clusters were distinct in perception: The results of a mixed linear model predicting ratings from CLUSTER (same / different) and GENDER (same / different) (random intercepts for participant, stimulus pair; random per-participant slopes for CLUSTER, GENDER) indicate that ratings were more similar for stimulus pairs from the same cluster ($b = 0.3$, $t = 3.7$, $p < 0.001$) and from the same gender ($b = 0.7$, $t = 8.8$, $p < 0.001$). Mean ratings per condition are given in Table 1.

Taken together, the results indicate that the cluster analysis successfully separated contours that German listeners can distinguish from another perceptually, even in a very abstract, delexicalized format, albeit only if speaker gender is controlled for. A further step, which must be left for future work, would be to investigate to which extent the late falls are also perceptually distinct from e.g. L+H* L-% and the level plateaus perceptually distinct from canonical continuation rises, which would allow phonological conclusions about the inventory of German boundary tones.

Proceedings of the Second International Conference on Tone and Intonation, pages 5–6
November 18–20, 2023.

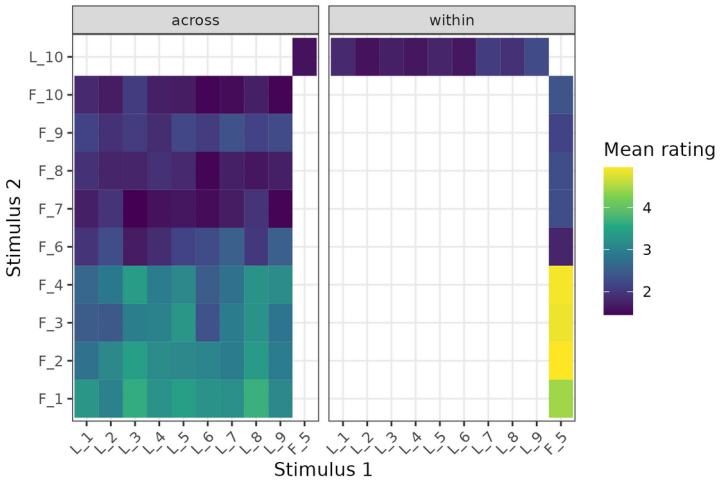


Figure 1: Mean similarity ratings for every stimulus combination

Table 1: Mean ratings (SDs in brackets) for cluster-gender combinations

	Cluster: Same	**Cluster: Different**
Gender: Same	4.8 (0.5)	3 (1)
Gender: Different	1.9 (0.9)	1.9 (0.9)

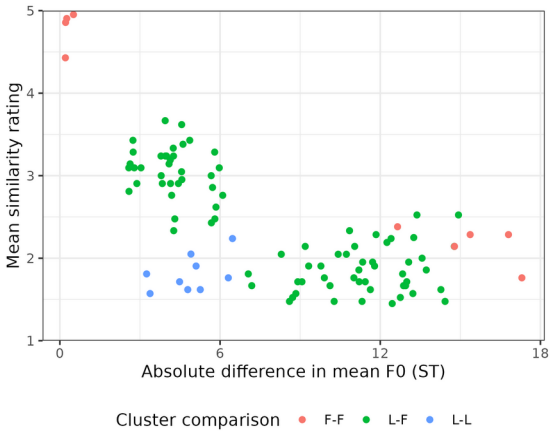


Figure 2: Correlation between absolute difference in mean F0 in ST and similarity ratings

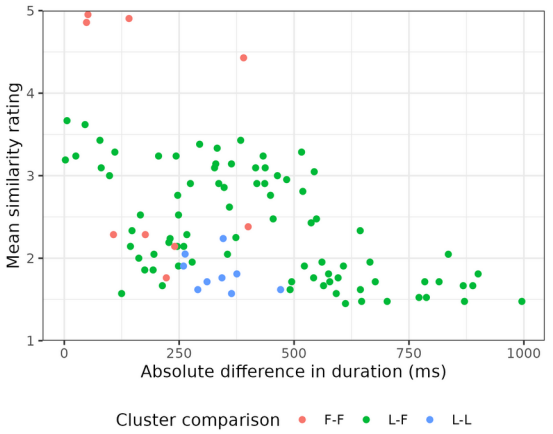


Figure 3: Correlation between absolute difference in contour duration and similarity ratings

References

[1] Seeliger, H., & Kaland, C. 2022. Boundary tones in German *wh*-questions and *wh*-exclamatives – a cluster-based approach. *Proceedings of Speech Prosody 2022* (Lisbon, Portugal), 27–31.

[2] Kaland, C. 2021. Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association*, 1–30.

[3] Seeliger, H., & Repp, S. In prep. Facets of prosodic prominence marking in non-assertive speech acts: cumulativity, non-locality, and constructional defaults.

[4] Grice, M., Baumann, S., & Benzmüller, R. 2005. German Intonation in Autosegmental-Metrical Phonology. *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun (ed.), 55–83.

[5] Niebuhr, O. 2013. Resistance is futile – the intonation between continuation rise and calling contour in German. *Proceedings of INTERSPEECH 2013* (Lyon, France), 225–229.

[6] Boersma, P., & Weenink, D. 2022. *Praat: doing phonetics by computer*. Computer program. Version 6.2.09.

[7] Mathôt, S., Schreij, D., & Theeuwes, J. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mutual predictability of the accent patterns of Tokyo and Osaka Japanese

Hiroto Noguchi
Sophia University, Tokyo Medical and Dental University
noguchih425@gmail.com

Many studies have been conducted on accent patterns in Japanese dialects in linguistics, including their differences [1]–[3]. They have reported historical correspondence in the accent patterns of short native words among dialects in Japanese, including those of Tokyo and Osaka Japanese. In natural language processing, data-driven research using machine learning has recently been performed on the accent patterns for words not registered in accent dictionaries in Tokyo Japanese [4]–[6]. Accent correspondence has been studied in linguistics, and machine learning has been used to predict the accent patterns of Tokyo Japanese. However, there has been no research connecting the two. Therefore, this study investigated the extent to which one accent contributed to the prediction of the other using machine learning focusing on Tokyo and Osaka Japanese.

This study used an accent dictionary [7] for learning. For the data without accent variants, spaCy [8] and GiNZA [9] split each word into morphemes with the split mode A [9]. Wordfreq [10] provided word frequency. Lexical strata were also provided on the assumption that words were (i) Sino-Japanese if they contained only Chinese characters, (ii) loan words if they contained only *katakana*, and (iii) native words otherwise. The data with accent variants in either Tokyo or Osaka were excluded. Scikit-learn [11] was used for some preprocessing procedures, and Tensorflow [12] was used for preprocessing and machine learning.

Surface-form characters and accent patterns in the dictionary [7] were used as features. The model architecture was defined as a sequential neural network with two dense layers: the first with 128 units and ReLU (rectified linear unit) activation, and the second with several units equal to the number of labels (51 for Tokyo Japanese; 45 for Osaka Japanese). The model was trained using a batch size of 250 and 10 epochs and compiled using the Adam optimizer, the sparse categorical cross-entropy loss function, and the accuracy metric.

The accent patterns of the 10,000 most frequent words were predicted in Tokyo Japanese based only on the surface forms. Next, the accent patterns of the other dialect were added. The surface form and the accent pattern of Osaka Japanese were given for the prediction of Tokyo Japanese, and the surface form and the accent pattern of Tokyo Japanese were given for the prediction of Osaka Japanese. Moreover, DummyClassifier [11] was used to confirm the baseline, always predicting the most frequent class in the training data.

The results are summarized in Table 1. When given the accent pattern of the other dialect, the percentage of correct predictions increased in both dialects but was greater in Tokyo Japanese. The results from the baseline generated by DummyClassifier are summarized in Table 2. For example, the descending condition represents that the model was built with the 80% most frequent words and tested with the 20% least frequent words. The descending-ordered data had a higher prediction rate than the ascending data in both prediction directions (from Tokyo to Osaka and vice versa). Other models were also evaluated in conditions by part-of-speech and morphological structure to examine supplementary factors. Table 3 shows the results for these conditions. The prediction rates improved with more restrictions.

Osaka accent patterns were more difficult to predict, which can be explained by the fact that there were two registers (low-beginning and high-beginning [2]) and more possible class labels in machine learning. Historically, Kansai Japanese, including Osaka Japanese, was the standard language, so Tokyo Japanese has limited deviation patterns from Kansai Japanese, making learning easy to converge. On the other hand, the prediction in the opposite direction was impossible because the standard language was inferred from a randomly derivated peripheral one.

The fact that low-frequency words are predictable from high-frequency words suggests that the accent patterns of low-frequency words are analogous to those of high-frequency words, rather than default accent patterns being more likely to appear in low-frequency words. Concerning Tokyo Japanese, accent prediction rates improved when the accent patterns of the other dialects were given, suggesting that machine learning with Kansai Japanese may help improve the efficiency of describing accent patterns of endangered dialects in other regions in Japan.

Proceedings of the Second International Conference on Tone and Intonation, pages 7–8
November 18–20, 2023.

Table 1: *Accuracy rates of the neural network model.*

Target	Accent info of the other dialect	Accuracy
Tokyo	No	0.5845
	Yes	0.7630
Osaka	No	0.5435
	Yes	0.5580

Table2: *Accuracy rates of DummyClassifier.*

Direction	Order by frequency rates	Accuracy
Tokyo □ Osaka	ascending	0.3755
	descending	0.5100
Osaka □ Tokyo	ascending	0.6235
	descending	0.6700

Table 3: *Accuracy rates for other conditions.*

Target	Noun only	Simplex	Accuracy
Tokyo	Yes	Yes	0.8102
	No	Yes	0.7415
	No	No	0.6978
Osaka	Yes	Yes	0.6017
	No	Yes	0.5977
	No	No	0.5600

References

[1] S. Kawahara, "The phonology of Japanese accent," in *Handbook of Japanese phonetics and phonology*, De Gruyter Mouton, 2015, pp. 445–492.

[2] The Society for Japanese Linguistics, *Nihongogaku daijiten [The encyclopedia of Japanese linguistics]*. Tokyodo Shuppan, 2018.

[3] Kindaichi H., *Kokugo akusento no shitekikenkyu [A historical study of Japanese accent patterns: Principles and methodology]*. Hanawashobo, 1974.

[4] H. Nakajima, M. Nagata, H. Asano, and M. Abe, "Estimating Japanese person name accent from mora sequence using support vector machines," *IEICE(D)*, vol. 88, no. 3, pp. 480–488, 2005.

[5] R. Kuroiwa, N. Minematsu, Y. Den, and K. Hirose, "Accent labeling of a large-scale database by a single labeler and its use in statistical learning of accent sandhi," *Tech. Rep. IEICE*, vol. 106, no. 614, pp. 31–36, 2007.

[6] H. Tachibana and Y. Katayama, "Accent estimation of Japanese words from their surfaces and romanizations for building large vocabulary accent dictionaries," Sep. 2020, doi: 10.1109/ICASSP40776.2020.9054081.

[7] Sugito M., *CD-ROM accent dictionary of spoken Osaka and Tokyo Japanese*. Maruzen, 1995.

[8] M. Honnibal and I. Montani, "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing," *Appear*, vol. 7, no. 1, pp. 411–420, 2017.

[9] "GiNZA NLP Library." Megagon Labs. [Online]. Available: https://github.com/megagonlabs/ginza

[10] R. Speer, "rspeer/wordfreq: v3.0." Zenodo, Sep. 26, 2022. doi: 10.5281/zenodo.7199437.

[11] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.

[12] M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems".

Modelling Fijian focus prosody using PENTAtrainer: A pilot study

Albert Lee[1], Candide Simard[2], Apolonia Tamata[2], Yi Xu[3]
Santitham Prom-on[4] & Jiaying Sun[1]
[1]*The Education University of Hong Kong,* [2]*University of the South Pacific,*
[3]*University College London,* [4]*King Mongkut's University of Technology Thonburi*
albertlee@eduhk.hk, candide.simard@usp.ac.fj, apolonia.tamata@usp.ac.fj,
yi.xu@ucl.ac.uk, santitham.pro@kmutt.ac.th, jsun@eduhk.hk

Studying the prosody of endangered languages is hard due to practical challenges in the field such as background noise and access to willing and capable speakers. Analysis-by-synthesis can help by offering an additional way to verify analyses based on limited speech data. Here we demonstrate the effectiveness of PENTAtrainer [1] as a PRAAT-based, no-code intonation synthesiser through Fijian, an understudied language which still has young native speakers.

Fijian is an Austronesian language spoken by about 400,000 as a first language [2] in Fiji. It has positional stress. The basic word order of Fijian is verb-object-subject (but note alternative accounts such as [2]). In an item-naming task [3], Fijian narrow focus was found to be marked by a general elevation of fundamental frequency (f_0) regardless of focus location. In another production study [4] using more natural target sentences, it was found that (i) **VP focus** (i.e. sentence-initial) had significantly higher on-focus mean f_0 than corresponding **neutral focus**, (ii) **Corrective-Subject focus** had significantly lower post-focus mean f_0 than corresponding **Subject focus**, and (iii) **Corrective-Object** focus had significantly lower pre-focus mean f_0 than corresponding **Object focus**. However, cross-sentence and cross-speaker variations were also observed.

For this study we chose PENTAtrainer to conduct analysis-by-synthesis for several reasons: (i) it is a no-code PRAAT-based synthesiser that would be convenient for the typical field linguist, (ii) it implements Target Approximation [5] which incorporates our current understanding of articulatory mechanisms (*contra* a purely mathematical model without physiological realism), and (iii) its functional annotation is blind to surface acoustic data (i.e. more objective).

We trained PENTAtrainer with the speech data from [4] (552 semi-spontaneous utterances from 10 native speakers), following procedures in [1]. In this pilot study, three communicative functions were hypothesised to be relevant, namely **Stress** (**S**tressed vs. **U**nstressed), **Focus** (Pre, On, Post, Neutral, *ala* [6]), and **Demarcation** (Sentence-Left, Sentence-Right, Word-Left, Word-Right, Medial) (see annotation scheme in attached **Appendix 1**). We first trained PENTAtrainer with TextGrid files that contained the function **Demarcation**, and extracted articulatory parameters for synthesis. This step was repeated twice, first with the **Stress** function added, then the **Focus** function. Figure 1 shows the user interface which displays the natural utterance and corresponding resynthesis, extracted pitch targets, and functional annotation.

Table 1 shows that including the **Focus** function in the TextGrid files led to much better synthesis accuracy (from r = .539 to r = .684). This is despite [3], [4] in which fundamental frequency was not found to mark narrow focus systematically compared with other prosodic cues. Compared with similar studies like [7]–[9], the present pilot study has yielded somewhat lower synthesis accuracy.

Table 1: *Summary of mean synthesis accuracy*

Communicative functions modelled	RMSE	*r*
Demarcation	4.634	.530
Demarcation + Stress	4.613	.539
Demarcation + Stress + Focus	4.274	.684

Table 1 suggests that the present functional annotation scheme may not sufficiently capture the variation in the data. In the next step, we will experiment with different functional annotation schemes. Alternatively, it could simply be that fundamental frequency does not play a very big role in marking prosodic focus (*ala* [10], [11], also *cf* [3], [4]). That is, Fijian as a verb-first language that uses word order to encode information structure may use fundamental frequency to mark narrow focus to a lesser extent compared with languages such as Mandarin [12]. A perception experiment is underway to shed light on this.

Proceedings of the Second International Conference on Tone and Intonation, pages 9–10
November 18–20, 2023.

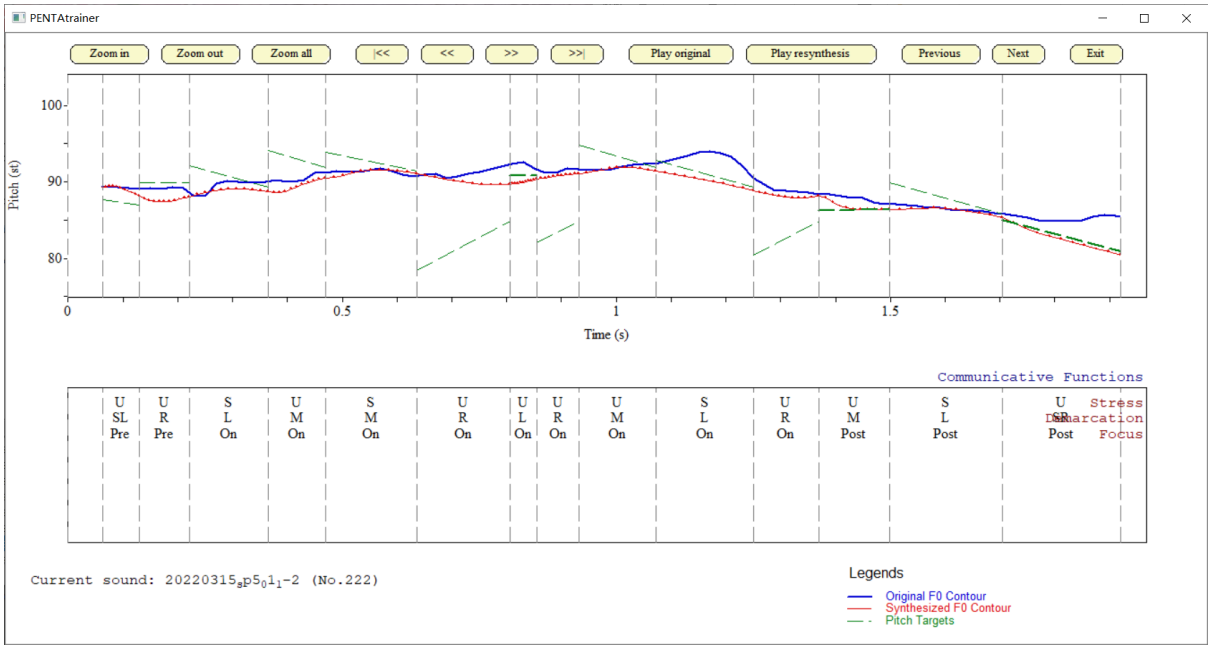


Figure 1: *User interface displaying a natural utterance (blue) and the corresponding resynthesis (red)*

References

[1] Xu, Y., & Prom-on, S. (2014). Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech Communication*, 57, 181–208.

[2] Geraghty, P. A. (2009). Fijian. In K. Brown & S. Ogilvie (Eds.), *Concise encyclopedia of languages of the world* (pp. 412–413). Elsevier.

[3] Mut, T. C., Simard, C., Tamata, A., & Lee, A. (2023). Focus prosody in Fijian: In-situ focus marking. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*.

[4] Lee, A., Simard, C., Tamata, A., Sun, J., & Mut, T. C. (2023). Focus prosody in Fijian: A pilot study. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*.

[5] Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33(4), 319–337.

[6] Xu, Y., Xu, C. X., & Sun, X. (2004). On the temporal domain of focus. *Proceedings of the 2nd International Conference on Speech Prosody (SP2004)*, 81–84.

[7] Lee, A., & Xu, Y. (2015). Modelling Japanese intonation using PENTAtrainer2. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015),* 86, 7–11.

[8] Simard, C., Wegener, C., Lee, A., Chiu, F., & Youngberg, C. (2014). Savosavo word stress: A quantitative analysis. *Proceedings of the 7th International Conference on Speech Prosody (SP2014)*, 512–514.

[9] Liu, F., Xu, Y., Prom-on, S., & Yu, A. C. L. (2013). Morpheme-like prosodic functions: Evidence from acoustic analysis and computational modeling. *Journal of Speech Sciences*, 3(1), 85–140.

[10] 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* (pp. 454–467). Oxford University Press.

[11] Yang, Y. (2023). First language attrition and second language attainment of Mandarin-speaking immigrants in Hong Kong: Evidence from prosodic focus. *Language Acquisition*, 30(2), 201–203.

[12] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27(1), 55–105.

Acknowledgements

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. EdUHK 18600621) awarded to AL.

Cross-generational variability of laryngeal contrasts in Shuangfeng Xiang Chinese

Menghui Shi[1] & Yiya Chen[2]
[1]*Fudan University,* [2]*Leiden University*
shimenghui@fudan.edu.cn, yiya.chen@hum.leidenuniv.nl

Languages differ in their utilization of acoustic and articulatory cues to signal laryngeal contrasts[1]. Most studies thus far have examined non-tone languages with a binary system of laryngeal contrast. A handful of studies on Asian tone languages show that the number of laryngeal contrasts, their phonetic realizations, and the effect of laryngeal contrasts on the *f0* patterns of their following vowels are typically more complicated than that in non-tone languages[2][3][4][5]. We report data which suggest that the phonetic realizations of multiple laryngeal contrasts can also vary across speakers of different generations within the same language. Understanding those complex interactions can help us gain insights into tonogenesis and the evolution from laryngeal contrasts to tonal systems.

Shuangfeng Xiang Chinese (SF hereafter), a Sinitic tone language, features a three-way laryngeal contrast in obstruents (unaspirated, aspirated, and voiced), which interacts with a five-way lexical tonal contrast (h-level, l-rising, h-falling, h-rising, and l-level, T1 to T5 hereafter), as plotted in Figure 1. T1, T3, and T4 only co-occur with voiceless onsets, while T5 occurs exclusively with voiced onsets. T2, however, can co-occur with all three-way onsets. In addition, phonation has been argued to serve as an important cue for signaling the voiced onsets[6]. Given the three-way contrast in VOT, the co-occurrence of lexical tones with consonant onsets, and the potential effect of onset voice quality on the following vowel in SF, we bring in empirical data to examine (i) how the phonetic cues interact to give rise to the laryngeal contrasts; and (ii) how speakers of different generations may utilize and weight the cues differently.

Two repetitions of a total of 20 morphemes were produced by speakers of two generations (22 old speakers with an average age of 58 yr vs. 15 young speakers with 35 yr). These morphemes consist of two items under each laryngeal condition of the five lexical tones. Acoustic and electroglottograph (EGG) signals were recorded simultaneously. Three parameters – voice onset time (VOT) of the onset as well as fundamental frequency (*f0*) and laryngeal contact quotient (CQ) of the following vowel – were analyzed. Principal component analysis (PCA) and multilevel regression models (i.e., GCA, GLMMs, and LMMs) were applied to data analyses.

Results suggest a relatively stable phonological distinction of the three-way laryngeal contrast across generations. On the one hand, the *f0* contours following aspirated onsets were consistently lowered relative to those following unaspirated onsets. On the other hand, the unaspirated onsets were consistently realized with shorter VOT and higher CQ, compared to that after aspirated onsets. The voiced onsets consistently co-occurred with low *f0* contours but their realizations showed varied relationships between VOT and CQ across the two generations of speakers. The old-generation speakers produced predominately negative VOT without significant differences in the following vowel's CQ, while the young-generation speakers produced fewer negative-VOT tokens and decreased CQ. Conjointly, our results show that cues (i.e., *f0*, VOT, and phonation) for laryngeal contrasts are interdependent, and the utilization of these cues can vary across generations. The changing relationship between laryngeal timing and phonatory state suggests a possible intermediate stage from voicing contrast to tonal contrast, lending support to the laryngeally-based model of tonal development[7].

Proceedings of the Second International Conference on Tone and Intonation, pages 11–12
November 18–20, 2023.

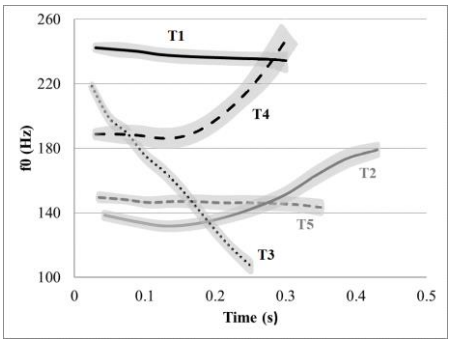


Figure 1: *f0* contours of the five lexical tones in SF (measured at 20 equidistant points). Light gray areas indicate ±SE.

References

[1] Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, *72*, 52–65.

[2] Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, *39*(4), 612–625.

[3] Pittayaporn, P., & Kirby, J. P. (2017). Laryngeal contrasts in the Tai dialect of Cao Bằng. *Journal of the International Phonetic Association*, *47*(1), 65–85.

[4] Shi, M., Chen, Y., & Mous, M. (2020). Tonal split and laryngeal contrast of onset consonant in Lili Wu Chinese. *The Journal of the Acoustical Society of America*, *147*(4), 2901–2916.

[5] Ge, C., Xu, W., Gu, W., & Mok, P. (2023). The change in breathy voice after tone split: A production study of Suzhou Wu Chinese. *Journal of Phonetics*, *98*, 101239.

[6] Zhu, X. 朱晓农, & Zou, X. 邹晓玲. (2017). 清浊同调还是气声分调—在音节学和类型学普适理论中安排湘语白仓话的声调事实 [One tone with onset voicing contrast or two tones with vowel breathy contrast]. *南方语言学* [South Linguistics], *12*, 1–10.

[7] Thurgood, G. (2007). Tonogenesis revisited: Revising the model and the analysis. In J. G. Harris, S. Burusphat, & J. E. Harris (Eds.), *Studies in Tai and Southeast Asian Linguistics* (pp. 263–291). Ek Phim Thai Co.

The Effects of Imitation and Emphasis Levels on the Learning of Post-Focus Compression: A Case Study of Cantonese Speakers on English

Ann Wai Huen To[1,2] & Yi Xu[2]
[1]*The Education University of Hong Kong,* [2]*University College London*
towaihuenann@gmail.com, yi.xu@ucl.ac.uk

Focus is a communicative function to convey emphasis. Prosodic focus is marked differently in different languages by not just the features of the on-focus interval, but also the intervals before and after the focus. In particular, post-focus compression (PFC), which is the lowering of pitch range and amplitude of the whole interval after the focus, is not a universal property for all languages [1]. For example, it is prevalent in languages like English [2, 3] and Mandarin [4], but absent in languages like Cantonese, in which focus is mainly correlated with the increase in on-focus duration and intensity [5]. It is reported that PFC can be acquired naturally in L2 learning by living in the L2 environment for many years [6, 7]. But is not yet clear whether and how PFC can be taught to second-language learners through training [8].

This study aims to further investigate how PFC can be taught to Cantonese learners of English. Two research questions are raised: (i) can the production of PFC by native Cantonese speakers of English be improved by imitation training? (ii) can the production of PFC by native Cantonese speakers of English be affected by different emphasis levels? A production experiment with three sections was conducted on native Cantonese speakers to answer these questions. Section one assessed participants' pre-training focus production; in section two participants imitated recordings of the target sentences read by a native British English speaker; section three tested how much of the benefit of the imitation training was maintained in mini dialogues with three levels of emphasis. The declarative sentence *May saw Nel in the morning* was used as the target sentence. The sentence was first recorded by a native British English speaker with focus on the initial word *May*, the medial word *Nel*, the final word *morning*, and no focus on any of the words.

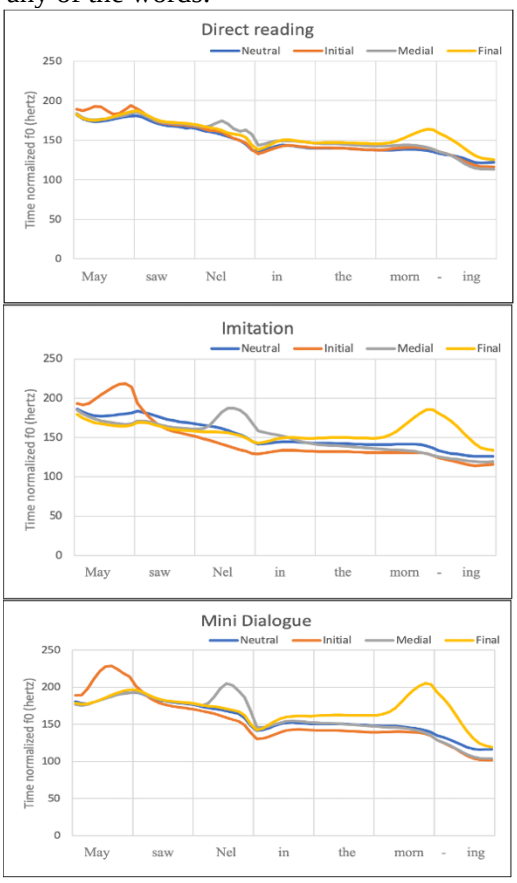


Figure 1: *Time normalized f_0 contour averaged across the 20 native Cantonese speakers for the three sections*

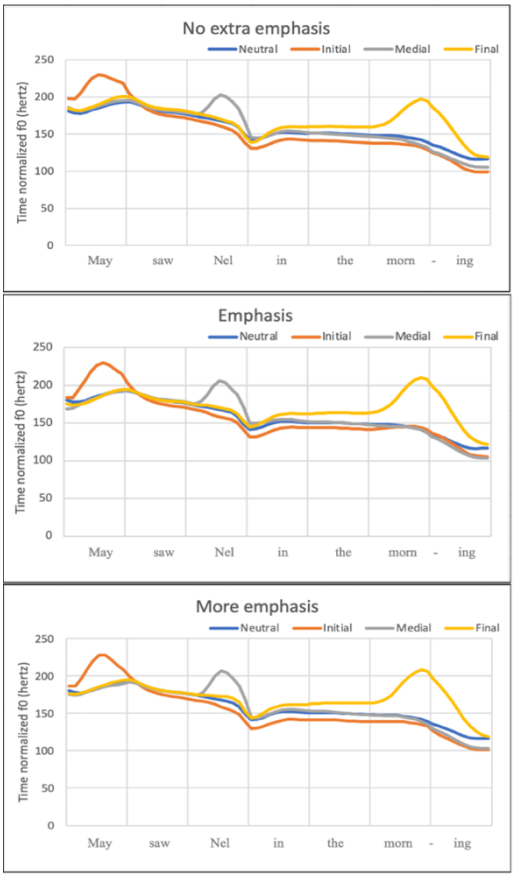


Figure 2: *Time normalized f_0 contour averaged across the 20 native Cantonese speakers for the three emphasis levels*

Proceedings of the Second International Conference on Tone and Intonation, pages 13–14
November 18–20, 2023.

Acoustic analysis was done with the Praat script ProsodyPro [9] and the time-normalized f_0 contours are shown in Figure 1. Results show that there was no evidence of PFC and only very slight on-focus expansion before training. The imitation training results showed a clear increase in on-focus expansion and PFC. For the post-training mini dialogue session, on-focus expansion could be clearly seen, but PFC is less evident compared to the imitation section. For all three emphasis levels, there was clear on-focus expansion and slight amount of PFC (Figure 2). From no extra emphasis (first level of emphasis) to contrastive emphasis (second level of emphasis), there was an increase in on-focus expansion, but PFC was reduced. From contrastive emphasis to extra emphasis (third level of emphasis), on-focus expansion had decreased, but there was slightly more PFC. When comparing the three emphasis levels, on-focus expansion was the most evident at the contrastive emphasis level and the least evident at no extra emphasis level, while PFC was the most evident at extra emphasis level and the least evident at contrastive emphasis level. The fact that the extent of on-focus expansion and PFC did not increase as the level of emphasis increased suggests that the effects of emphasis might be capped at a certain level, which is consistent with the finding of Chen and Gussenhoven [10].

The results of this study show that learners' direct mimicking of native speaker's focus prosody is effective for L2 focus learning. However, it is not yet clear how long the learning effect can be retained after a brief training session. Longitudinal studies are needed in the future to further test whether L2 learners could really associate the communicative function of focus to its prosodic features, how long the effects of imitation training can last, and how many training sessions are needed in order to achieve a long-term learning effect. One way of investigation is by systematically implementing this method of training in L2 pedagogy, such as in listening and oral dialogue practices. This method could also benefit L2 speakers by guiding them to speak and convey ideas with native-like intonation, in which communication effectiveness could be enhanced.

References

[1] Xu, Y. 2011. Post-focus compression: Cross-linguistic distribution and historical origin. *Proceedings of The 17th International Congress of Phonetic Sciences* (Hong Kong, China).

[2] Cooper, W. E., Eady, S. J. and Mueller, P. R. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77, 2142-56.

[3] Xu, Y. and Xu, C. X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159-97.

[4] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27, 55-105.

[5] Wu, W. L. and Xu, Y. 2010. Prosodic Focus in Hong Kong Cantonese without Post-focus Compression. *Proceedings of Speech Prosody 2010* (Chicago, US).

[6] Chen, Y. 2015. Post-focus compression in English by Mandarin learners. *The 18th International Congress of Phonetic Sciences* (Glasgow, UK).

[7] Chen, Y., Xu, Y. and Guion-Anderson, S. 2014. Prosodic realization of focus in bilingual production of Southern Min and Mandarin. *Phonetica* 71, 249-70.

[8] Gao, W, Xu, Y. and Mu, F. 2015. An experimental study of teaching prosodic focus to Chinese learners of English [中国英语学习者焦点教学的初步实验研究]. *Foreign Language Teaching and Research* 6, 861-73.

[9] Xu, Y. 2020. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis.

[10] Chen, Y., & Gussenhoven, C. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics* 36(4), 724–46.

Syntax-Prosody mismatches in Teotitlán Zapotec

Hiroto Uchihara[1] & Ambrocio Gutiérrez[2]
[1]*Tokyo University of Foreign Studies; Universidad Nacional Autónoma de México,*
[2]*University of Colorado at Boulder*
hirotouchihara81@gmail.com, Ambrocio.GUTIERREZLORENZO@colorado.edu

Intonational Phrase is a prosodic constituent that is larger than the Prosodic Word or the Prosodic Phrase, which corresponds more or less to the syntactic clause (Nespor & Vogel 1986: Ch. 7), which is a linguistic unit consisting of a subject and a predicate. In Teotitlán Zapotec, an Otomanguean language spoken in the state of Oaxaca, Mexico, the domain of Tone Sandhi and Final Glottalization is the Intonational Phrase. Intonational Phrase is often defined as the domain of intonational contours (cf. Nespor & Vogel 1986: Ch. 7), but intonation contours are not reported to be employed for pragmatic purposes in this language.

Tone Sandhi is a tonal process where a syllable with a low (a) or mid (ā) tone alternates with a falling (â) or a high (á) tone when it follows a syllable with a mid tone (ā), as illustrated in (1). In this example, the syllable *tɷ:* (in boldface) alternates with a falling tone (*tɷ̂:*) after a syllable with the mid tone, *ʒɯ̄:ʒ*. In the following examples, the first line shows the surface forms that are pronounced by the speakers, and the second line the underlying forms. The hyphen (-) connects affxes while the equal sign = connects clitics (which are outside of the phonological word). The intonational phrase boundaries are indicated by ‖ ... ‖ and the syntactic clause boundaries are indicated by [...].

The domain of Tone Sandhi is larger than the prosodic word boundaries indicated by the spaces in the first line as we can observe in (1), but it is not the case that this tone rule is applied when a syllable with a low or a mid tone follows a syllable with a mid tone. For instance, in (2), the syllable *zit* (in boldface) follows a syllable with the mid tone, *pān*, but Tone Sandhi is not applied here. This is because the domain of application of Tone Sandhi is the Intonational Phrase, and that there is an Intonational Phrase boundary between *pān* and *zit.*

The other process, Final Glottalization, is a process where a vowel with a low or high tone in an atonic open syllable at the final position of an Intonational Phrase is glottalized. In (3) the final 1SG enclitic *=a* is glottalized because it is found in the final position of the Intonational Phrase, while in (4) the same clitic is found within the Intonational Phrase and thus Final Glottalization is not applied.

The Intonational Phrase in Zapotec defined this way in many cases correspond to the syntactic clause, but not necessarily. First, there are cases where the Intonational Phrase is larger than the syntactic clause, as in (5). Here, the mid-tone syllable *dān* and the following syllable *gu* (in boldface) constitute independent syntactic clauses (quotation clause and the main clause), but Tone Sandhi is applied and the tone on *gu* alternates from a low tone to the high tone. Thus, the structure here is ‖ […] […] ‖ .

On the other hand, there are cases where the Intonational Phrase is smaller than the syntactic clause. In (6), the initial element in the prepositional phrase, *kun*, follows a syllable with a mid tone, *kī,* but Tone Sandhi is not applied. This is because the phrase beginning with *kun* constitutes an independent Intonational Phrase. Thus, the structure here is [‖ … ‖ ‖ … ‖].

In this talk, we examine the cases of such Syntax-Prosody mismatches in Teotitlán Zapotec at the level of the Intonational Phrase, and situate it in a typological and theoretical perspectives. From the typological perspectives, Himmelman et al. (2018) and Himmelman (2022) conclude that phonetic Intonational Phrase is universal and even non-native speakers can judge its boundaries, while phonological Intonational Phrase, as in the cases discussed so far, is grammaticalized and differ in individual languages. From theoretical perspectives, it has been argued that the illocutionary clauses show more consistent mapping to an Intonational Phrase than standard clauses (Selkirk 2011). Furthermore, Gussenhoven (2004: 292) claims that reporting clauses are either incorporating or cliticizing, while complex reporting clauses cannot be incorporated. The goal of this presentation is to examine if these generalizations also hold in Zapotec, where the Intonational Phrase is not defined with intonational contours as in other better described languages.

Proceedings of the Second International Conference on Tone and Intonation, pages 15–16
November 18–20, 2023.

(1) ‖ [baˈʒu̯ːʒˈ**tæ̯ː**] ‖
ba-ʒu̯ːʒ tæ̯ː
COMPL-fray INT
‘frayed/scratched (something) a lot’

(2) ‖ [ˈzitˈtæ̯ː ˈmĕːdy ˈgūpān,] ‖ ‖ [ˈ**zit**ˈtæ̯ː ˈmĕːdy ˈgūpān] ‖
zit tæ̯ː mĕːdy gu-(ā)p=ān zit tæ̯ː mĕːdy gu-(ā)p=ān
much INT money COMPL-have=3SG.F much INT money COMPL-have=3SG.F
‘He had a lot of money, he had a lot of money! ’

(3) ‖ [ˈnisrú riˈkāːz**áʾ]** ‖
nis=rú ri-kăːz=a
water=more HAB-want:1SG=1SG
‘I want more water.’

(4) ‖ [riˈkāːz**á** ˈgâː ˈkyǽ̯ː] ‖
ri- kăːz =a ø-́ gaː kyǽ̯ː
HAB-want:1SG=1SG POT-trim head.of:1SG
‘I want to get a haircut.’

(5) ‖ [ryubˈlạːz ˈtæ̯ːdān] [**gú**ˈnnį̣ː biʾnˈgĭːʒ nazaˈkǣʾnkī] ‖
r-yublạːz tæ̯ː=dān gu-nnį̣ː biʾn+gĭːʒ nazak-ǣʾn=kī
HAB-be.in.a.hurry INT=3PL.F COMPL-say youngster+toff good-DIM=INV.DEM
‘They are very much in a hurry, said the Little Prince’

(6) [‖ guˈnnậː biʾnˈgĭːʒ nazaˈkǣʾnkī zúːbán ˈkyæːkī ‖ ‖ **kun**ˈnáll ˈnyæ̂ʾn ‖]
gu-nnậː biʾn+gĭːʒ nazak-ǣʾn=kī zŭːb=ān kyæː=kī **kun**=n-áll nyæʾ=(a)n
COMPL-witness:1SG youngster+toff good-DIM=INV.DEM be.sitting=3SG.F head=INV.DEM with=STAT-hang leg=3SG.INF
‘I saw that the Little Prince sitting over up there, (with) his legs (were) hanging’

References

[1] Gussenhoven, Carlos. 2004. The phonology of tone and intonation. Cambridge: Cambridge University Press
[2] Himmelmann, N., Sandler, M., Strunk, J., & Unterladstetter, V. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. Phonology, 35(2), 207-245. doi:10.1017/S0952675718000039.
[3] Himmelmann, Nikolaus P. 2022. Prosodic phrasing and the emergence of phrase structure. Linguistics 60.3: 715-743
[4] Nespor, Marina and Irene Vogel. 1986 [2007]. Prosodic Phonology. Dordrecht: Foris.
[5] Selkirk, Elizabeth. 2011. Syntax-phonology interface. In John Goldsmith ed., the Handbook of Phonological Theory, 2nd ed, 435-484. Wiley-Blackwell.

Benefits of Targeted Memory Reactivation in Perceptual Learning of Non-native Tones are Associated with Slow-oscillation Phase and Delta-theta Power

Xiaocong Chen[1], Jiayi Lu[1], Zhen Qin[2], Xiaoqing Hu[3], Caicai Zhang[1]
[1]*The Hong Kong Polytechnic University*, [2]*The Hong Kong University of Science and Technology*, [3]*The University of Hong Kong*
xiaocong.chen@polyu.edu.hk, jiayi-rachel.lu@polyu.edu.hk, hmzqin@ust.hk, xiaoqinghu@hku.hk, caicai.zhang@polyu.edu.hk

Perceptual learning of non-native speech sounds involves the formation of new phonological categories in long-term memory, which is supported by memory consolidation processes [1]. Recent studies indicate that sleep could facilitate memory consolidation during the perceptual learning of non-native speech sound categories [2][3]. Intriguingly, a new technique, known as targeted memory reactivation (TMR), has been reported to boost the reactivation of newly learned information during sleep by presenting external sensory cues associated with priorly learned information during sleep [4]. However, the neural oscillatory activities induced by TMR in relation to the behavioral performance of memory consolidation are not well understood. Besides, TMR has been successfully applied to vocabulary learning [5] and grammar learning [6], but the application of this new technique in the perceptual learning of non-native speech sounds remains unexplored. To address these issues, the current study employed TMR in the perceptual learning of Cantonese level tones by native Mandarin speakers and examined the TMR benefits and the associated underlying neural activities.

Sixty-two native Mandarin speakers who were naïve to Cantonese were recruited to complete the perceptual learning of three Cantonese level tones. The experiment consisted of three sessions: (a) pre-TMR training and assessment session, where participants received training on the three Cantonese level tones and were assessed by a forced-choice identification task (identifying the correct tonal category among the three level tones upon hearing a tonal syllable); (b) a 90-min nap session with TMR, where half of the learned auditory tonal syllables (i.e., cues) were played to the participants during the slow-wave sleep (SWS) stage while their sleep EEG signals were recorded (the cued and uncued syllables were counterbalanced across the participants); (c) post-TMR assessment session, where participants' learning of the Cantonese level tones were re-assessed using the same forced-choice identification task as in the pre-TMR training session.

Although our focus is the individual differences in the brain oscillatory patterns as they contribute to the TMR benefits, we first analyzed the learning effect and behavioral changes related to TMR cues. By comparing the accuracy of the initial and final block of the pre-TMR training task, we observed significant accuracy improvement after tonal training, confirming the occurrence of perceptual learning. Intriguingly, the logistic mixed-effect regression analysis for the pre- and post-TMR identification tasks revealed no significant difference in the consolidation effects between the TMR-cued and uncued tonal syllables, indicating TMR benefits were not confined to specific cued items in tonal learning (see Fig 1). There was only a significant main effect of testing session, with worse performance in the post-TMR identification task than the pre-TMR identification task, indicating memory decay after sleep. The lack of cued syllable effects can be explained by the nature of the training task, which requires participants to categorize tonal categories in an abstract manner irrespective of syllables, consistent with the formation of abstract tonal representations.

Importantly, we compared the sleep EEG data between participants with higher TMR and those with lower TMR benefits. The phase analysis showed that the phase angles at the onset of the auditory cues for the higher benefit group exhibited a significant non-uniform distribution in the parietal electrodes, which was not observed for participants with lower TMR benefits. Specifically, the trial-level cue onsets from the parietal electrodes among the higher benefit group were preferentially coupled to the slow oscillatory up-states (see Fig 2). In addition, the time-frequency analysis revealed higher post-cue delta-band (1-4 Hz) and theta-band (4-8 Hz) power for the higher benefit group than the lower benefit group (see Fig 3). These suggest that the degree of TMR benefits may be associated with the relative coupling between the cue onset and the SO up-state, and the increase in the post-cue delta-theta power, consistent with recent findings [7][8]. These findings shed some new light on the neural oscillatory mechanism of memory reactivation and consolidation during sleep in speech sound learning, and also demonstrate some potential value of TMR for assisting in the perceptual learning of non-native speech sounds by second language learners.

Proceedings of the Second International Conference on Tone and Intonation, pages 17–18
November 18–20, 2023.

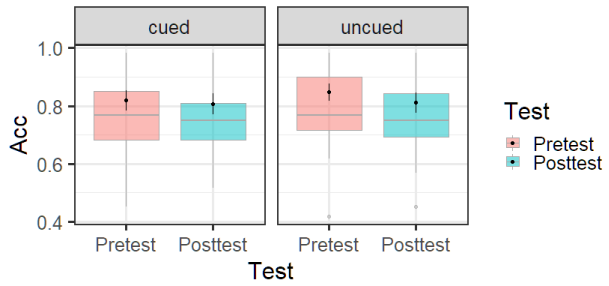


Figure 1: *Accuracy of the pretest and posttest for cued and uncued syllables.*

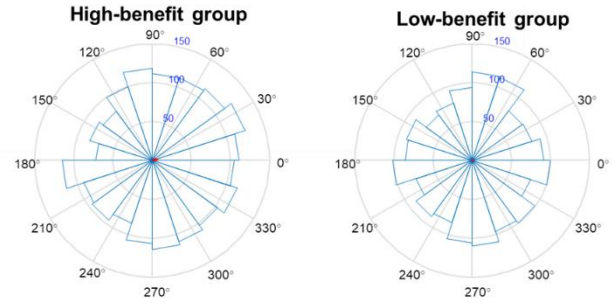


Figure 2: *Phase angle distributions of the cue onset in PZ for high- and low-TMR-benefit groups*

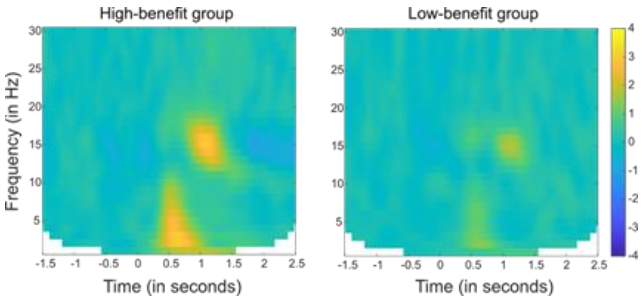


Figure 3: A*verage time-frequency power of nine electrodes for high- and low-TMR-benefit groups*

References

[1] Earle, F. S., & Myers, E. B. 2014. Building phonetic categories: An argument for the role of sleep. *Frontiers in Psychology*, *5*, 1192.

[2] Earle, F. S., & Myers, E. B. 2015. Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *Journal of Acoustical Society of America*, *137* (1), EL91–EL97.

[3] Qin, Z., & Zhang, C.. 2019. The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PLoS One*, *14*, e0221498.

[4] Hu, X., Cheng, L. Y., Chiu, M. H., & Paller, K. A. 2020. Promoting memory consolidation during sleep: A meta-analysis of targeted memory reactivation. *Psychological Bulletin*, *146*(3), 218–244.

[5] Schreiner, T., & Rasch, B. 2015. Boosting vocabulary learning by verbal cueing during sleep. *Cerebral Cortex*, *25*(11), 4169–4179.

[6] Batterink, L. J., & & Paller, K. A. 2017. Sleep-based memory processing facilitates grammatical generalization: Evidence from targeted memory reactivation. *Brain and Language.*, *167*, 83–93.

[7] Xia, T., Antony, J. W., Paller, K. A., & Hu, X. 2022. Targeted memory reactivation during sleep influences social bias as a function of slow-oscillation phase and delta power. *Psychophysiology*, e14224.

[8] Xia, T., Yao, Z., Guo, X., Liu, J., Chen, D., Liu, Q., Paller, K. A., & Hu, X. 2023. Updating memories of unwanted emotions during human sleep. *Current Biology*, *33*(2), 309-320.

Daytime Naps Consolidate Cantonese Tone Learning Through Talker Generalization

Ruofan Wu[1, 2], Zhen Qin[1], & Caicai Zhang[2]
[1]*Hong Kong University of Science and Technology*, [2]*The Hong Kong Polytechnic University*
ruofan-ann.wu@connect.polyu.hk, hmzqin@ust.hk, caicai.zhang@polyu.edu.hk

This study investigates whether nap-meditated memory consolidation of Cantonese tones promotes generalization to a novel talker (i.e., talker generalization). Previous research has consistently shown that post-training sleep is beneficial for the consolidation of newly learned linguistic knowledge [1]. This sleep-dependent consolidation effect was found to promote the generalization of non-native sounds (e.g., lexical tones) to a novel talker [2,3]. Importantly, recent studies showed an effect of prior knowledge on memory consolidation, which suggests that new information that is consistent with the existing knowledge is better consolidated than that which is not [4]. Lexical tones, with pitch variations signaling word identity, show high variability across talkers and may pose difficulty to second-language learners. For instance, Mandarin speakers encode pitch contour differences in their native contour-tone system and are found to have difficulties learning Cantonese level-level tonal contrasts cued by pitch height differences [5]. Therefore, the present study investigates whether daytime naps help Mandarin speakers consolidate Cantonese tones through the promotion of talker generalization (***RQ1***); whether the nap-mediated consolidation effect is dependent on tonal contrasts, given Mandarin speakers' prior knowledge of pitch contour cueing contour-level contrasts (***RQ2***).

Eighty-eight Mandarin native speakers, who are from North China and don't speak tonal languages other than Chinese, were recruited for a nap group and a control (non-nap) group (see Fig.1 for experimental procedures). They first completed language background questionnaires and tests of pitch and musical aptitudes, cognitive batteries (e.g., memory span), and sleep surveys to ensure that the two groups had matched pitch processing sensitivity, cognitive abilities, and normal sleep patterns. After that, they were trained with two Cantonese contour-level (T5-T6) tonal contrasts (see Fig. 2; four words making up two pairs) and two level-level (T3-T6) tonal contrasts, respectively (240 trials each tonal pair, order counterbalanced), followed immediately with the 1st tone identification (ID) task using the stimuli produced by the trained (male) talker (80 trials for each tonal pair). Participants, matched in nap habitual routines, were then pseudo-randomly assigned to either the nap group, who napped for 1.5 hours with brain EEG activity recorded, or the non-nap group, who rested and stayed awake by watching silent documentaries for 1.5 hours. After the manipulation of the nap session, the participants were tested again with the trained talker 2nd ID task, and then finally a novel (female) talker 3rd ID task (see Fig. 2 for the stimuli produced by trained vs. novel talkers).

Two sets of mixed-effects models were conducted on response accuracy of tone identification across sessions. The model on the results of ID 1 and ID 2 (deviation coding; ID 1:-0.5, ID 2:0.5; see Fig. 3) revealed that comparing identification accuracy of trained talker before and after the nap, there were main effects of Tone ($\beta = -0.65$, SE = 0.03, $z = -19.20$, $p < .001$) and Session ($\beta = 0.13$, SE = 0.05, $z = 2.41$, $p = 0.02$), but no interaction was found between Group and the other two factors. The model on the results of ID 2 and ID 3 (deviation coding; ID 2: -0.5, ID 3:0.5) revealed a 3-way interaction between Group, Session, and Tone ($\beta = 0.30$, SE = 0.13, $z = 2.32$, $p = 0.02$). A post-hoc analysis showed that, as illustrated in the bottom panel of Fig.3, only the nap group had an interaction between tone and session ($\beta = -0.20$, SE = 0.09, $z = -2.34$, $p = .02$), but not the non-nap group ($\beta = 0.10$, SE = 0.10, $z = 1.02$, $p = .31$). The finding indicates that the nap group showed a smaller ID difference between the trained and novel talkers for contour tones than for level tones, whereas the non-nap group did not show the effect. We further analyzed individual participants' sleep-related EEG activities on the talker generalization effect (i.e., ID3-ID2 difference) for contour and level tones in the nap group [6]. The results (see Fig. 4) showed that the N2 fast spindle density ($\beta = 0.48$, SE = 0.23, $z = 2.14$, $p = 0.03$) and N3 slow-wave sleep percentage ($\beta = 0.19$, SE = 0.07, $z = 2.80$, $p = .005$) positively predicted the talker generalization effect in the contour tones only [4,6].

Aligned with previous studies on overnight consolidation [2,3], the findings suggest that daytime naps benefit the nap participants' tone consolidation by promoting talker generalization. The performance depends on the sleep spindles and slow-wave sleep, which are the brain activities indexing the hippocampal-neocortical cycle underlying memory consolidation processes [4,6], providing further evidence supporting the beneficial role of naps. The finding of tone effect might be attributed to the participants' prior tonal knowledge of pitch contour that cues contour-level tonal contrasts, which was found to be prioritized during sleep-mediated memory consolidation [4].

Proceedings of the Second International Conference on Tone and Intonation, pages 19–20
November 18–20, 2023.

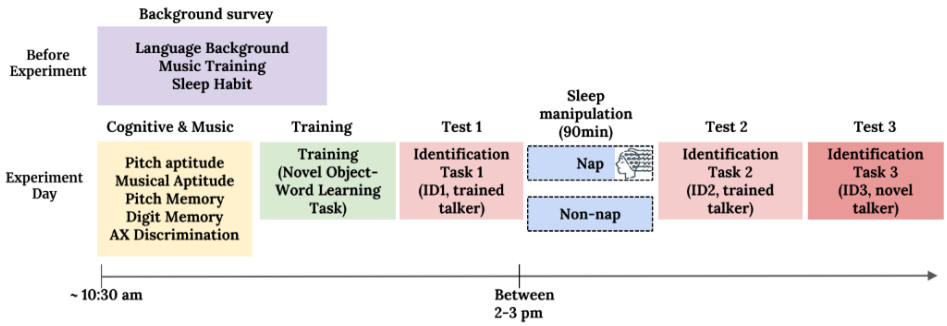


Figure 1: *An overview of the experimental procedure.*

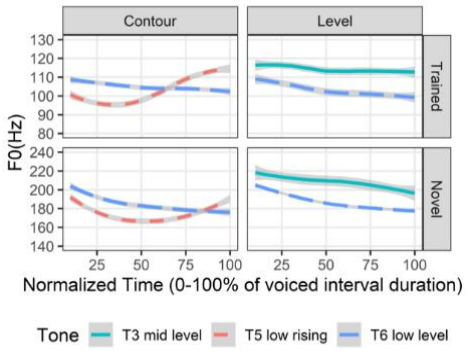


Figure 2: *Time-normalized F0 tracks of a Cantonese contour-level (T5-T6, left, primarily cued by pitch contour differences) and level-level (T3-T6, right, primarily cued by pitch height differences) tonal contrast, produced by a trained male (top) and a novel female talker (bottom).*

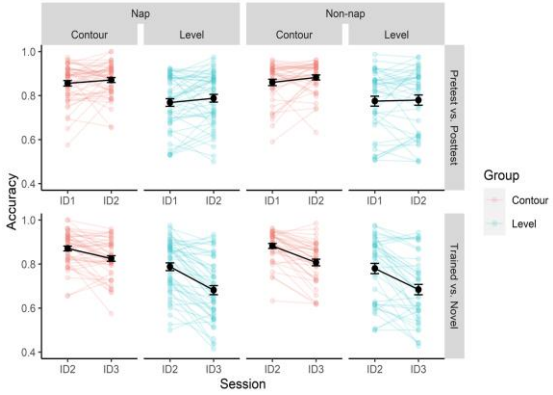


Figure 3: *The identification accuracy by the nap group (left) and the non-nap group (right) of a trained male talker (top) and a novel female talker (bottom) before and after the nap manipulation.*

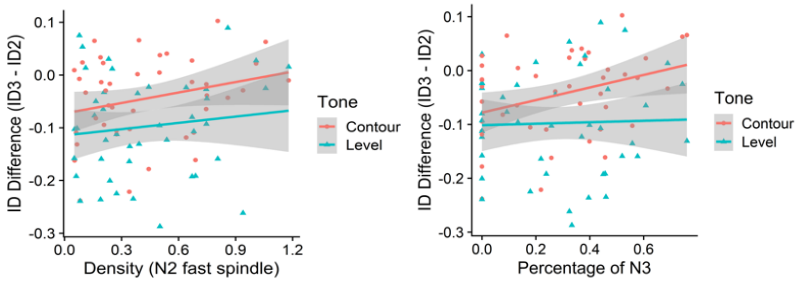


Figure 4: *The predictive relationship between N2 fast spindle density (left) / N3 percentage (right) and the difference between identification accuracy of post-nap trained and novel talkers.*

References

[1] M. H. Davis, A. M. Di Betta, M. J. E. Macdonald, and M. G. Gaskell, "Learning and Consolidation of Novel Spoken Words," *J. Cogn. Neurosci.*, vol. 21, no. 4, pp. 803–820, Apr. 2009, doi: 10.1162/jocn.2009.21059.

[2] Z. Qin and C. Zhang, "The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts," *PLOS ONE*, vol. 14, no. 12, p. e0221498, Dec. 2019, doi: 10.1371/journal.pone.0221498.

[3] F. S. Earle and E. B. Myers, "Overnight consolidation promotes generalization across talkers in the identification of non-native speech sounds," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. EL91–EL97, Jan. 2015, doi: 10.1121/1.4903918.

[4] N. Hennies, M. A. Lambon Ralph, M. Kempkes, J. N. Cousins, and P. A. Lewis, "Sleep Spindle Density Predicts the Effect of Prior Knowledge on Memory Consolidation," *J. Neurosci.*, vol. 36, no. 13, pp. 3799–3810, Mar. 2016, doi: 10.1523/JNEUROSCI.3162-15.2016.

[5] Z. Qin and A. Jongman, "Does Second Language Experience Modulate Perception of Tones in a Third Language?," *Lang. Speech*, vol. 59, no. 3, pp. 318–338, Sep. 2016, doi: 10.1177/0023830915590191.

[6] S. Studte, E. Bridger, and A. Mecklinger, "Sleep spindles during a nap correlate with post sleep memory performance for highly rewarded word-pairs," *Brain Lang.*, vol. 167, pp. 28–35, Apr. 2017, doi: 10.1016/j.bandl.2016.03.003.

Tonal contrast in Drenjongke (Bhutia): an Electroglottograph study

Seunghun J. Lee[1,2] , Julián Villegas[3] & Kunzang Namgyal[4]
[1]*International Christian University,* [2]*IIT Guwahati,* [3]*University of Aizu,* [4]*Sikkim University*
seunghun@icu.ac.jp, julian@u-aizu.ac.jp, knamgyal@cus.ac.in

Introduction This paper explores the tonal contrast in Drenjongke with electroglottographic (EGG) data. Drenjongke (Bhutia), a Tibeto-Burman language spoken in Sikkim, India, is a two-tone language (low (L) and high (H) tone). The two tones are lexical in vowels and sonorant-initial syllables, but post-lexical in obstruent-initial syllables. In Lee et al. (2018), acoustic data showed that f0-based contrast is realized on the sonorant onset, but not in the following vowel. In Lee et al. (2019), it was found that devoiced plosives are associated with a higher F1, most probably due to the jaw opening, in addition to lower F0 in the vocalic portion following the plosive.

Research questions We report EGG data to investigate whether the previously tonal contrast observed in the acoustic signal is also observable in the EGG, and whether the L-tone induced in devoiced consonants is realized with breathy phonation.

Methods Recordings of 12 Drenjongke speakers are analyzed for the four laryngeal types. Both the consonantal and the vocalic components of acoustic and EGG signals were annotated. Eggnog (Villegas 2020) was used to extract open quotient—OQ values. From 2439 tokens, 196 tokens (8%) were discarded because of time-misalignments between EGG and audio recordings or because of problems when extracting OQ from the EGG signal.

Analysis Data were analyzed using a Generalized Mixed-Effect Model. OQ was logit transformed, scaled, and centered. We computed the mean of the transformed values per speaker and tone (grouping all repetitions). To observe the evolution of OQ throughout a segment, the analyses were performed at each of the time terciles. Only short vowels [a, i, y, u, ʻe, ø, o, æ] were used in the analyses.

Results Confirming our first research question, tone had a significant effect on OQ throughout the whole vowel. Low tones presented higher OQ in comparison to High tones as shown in Figure 1. The analysis of sonorants was performed separating the consonant [ŋ, ɲ, n, m, l] from the vowel segment [a]. Significantly lower values of OQ were observed for High tones, but only in the middle and last part of the consonant as summarized in Figure 2. For obstruent onsets, four laryngeal contrasts were considered: ʻVoiceless,ʼ ʻAspirated,ʼ ʻVoiced,ʼ and ʻDevoicedʼ before an [a]. When the effect of Tone was found significant, a post-hoc analysis based on Tukeyʼs multiple comparisons was performed to find significant OQ differences between the laryngeal contrasts. Significant differences in OQ were found at the beginning of the vowel as shown in Fig. 3. Considering these results with those of Fig. 1, it seems that L-tone induces devoicing, and that such phonation is produced breathier than voiced and voiceless phonation.

Discussion In general, the OQ values were higher than 50% in this corpus suggesting that phonemic contrast between creaky and breathy tones is unlikely. Additionally, an increase on the value of OQ at the end of the segments was observed. In the presented results differences (if any) were neutralized towards the end of the segments with exception of the vowels in isolation. The absence of significant differences at the beginning of the consonant in sonorants may be explained the lack of realization of the consonant. Interestingly, the OQ contrast gets neutralized in the vowel as previously observed in Lee et al. (2018) on the analysis of the acoustic signal of the same corpus. An additional study of OQ with actual words along with a perceptual study are currently being performed, these studies could elucidate whether OQ plays phonemic role in Drenjongke.

Proceedings of the Second International Conference on Tone and Intonation, pages 21–22
November 18–20, 2023.

References

Lee, Seunghun J., Hyun Kyung Hwang, Tomoko Monou, Shigeto Kawahara (2018) The phonetic realization of tonal contrast in Dränjongke. *Proc. of TAL2018*: 217–221.

Lee, Seunghun J., Shigeto Kawahara, Céleste Guillemot and Tomoko Monou (2019) Acoustics of the Four-way Laryngeal Contrast in Drenjongke (Bhutia): Observations and Implications. *J. of the Phonetic Soc. of Japan* 23: 65-75.

J. Villegas, "Eggnog," 2020. Available [June 26, 2023] from https://bitbucket.org/julovi/eggnog.

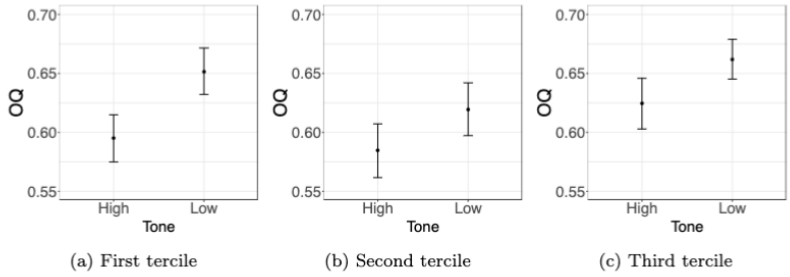


Figure 1. Effect of tone on OQ for vowels in isolation. Low tones have higher OQ compared to High tones. Unless otherwise noted, error bars show 95% CI.

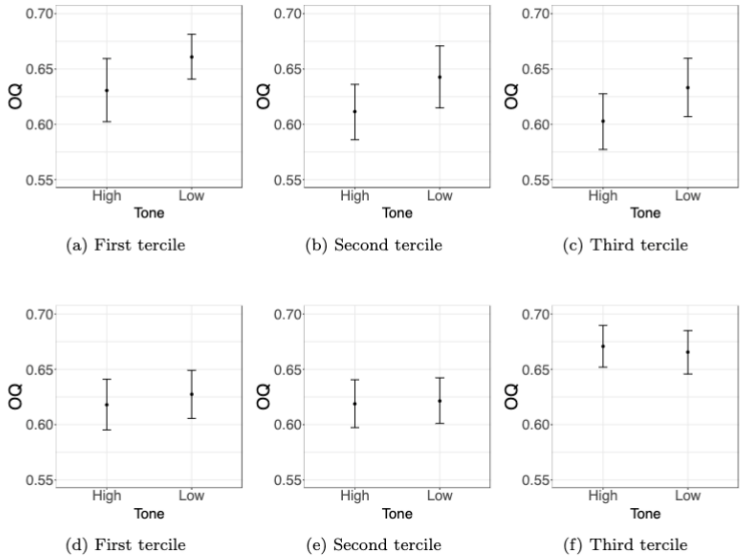


Figure 2. OQ on sonorants. Consonant and vowel segments in the top and bottom row, respectively.

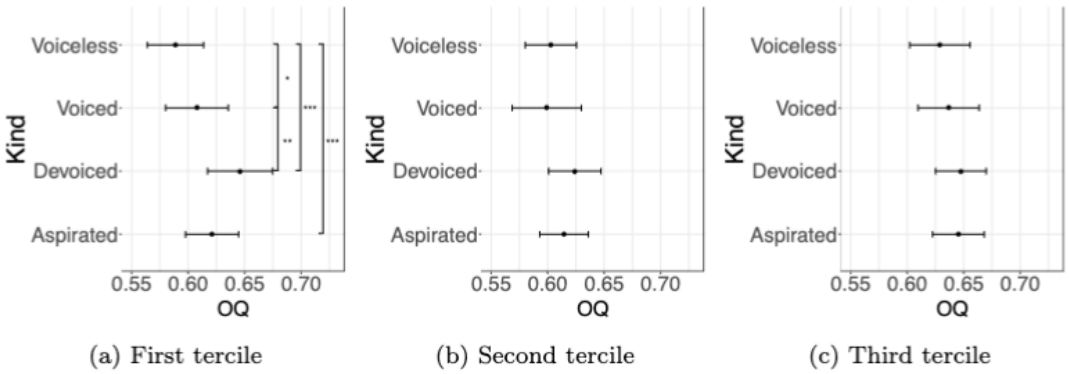


Figure 3. OQ on the vowel portion of laryngeal contrasts. Significant differences between laryngeal contrasts are denoted with brackets and asterisks ("***"< .001, "**"= 0.01, "*"=0.05).

Polite Tones of Voice in Transition: Investigating Speech Production and Perception in Thai Speakers of Different Generations

Sujinat Jitwiriyanont[1] & Pavadee Saisuwan[2]
[1 2] *Department of Linguistics and Center of Excellence in Southeast Asian Linguistics, Faculty of Arts, Chulalongkorn University*
Sujinat.j@chula.ac.th, Pavadee.s@chula.ac.th

Politeness in speech, a nuanced and culturally embedded aspect of human communication, undergoes intriguing transformations across different generations. In this study, we delve into the intriguing realm of polite speech production and perception among Thai speakers spanning three generations: Generation X (born between 1965 and 1980), Generation Y (born between 1981 and 1996), and Generation Z (born between 1997 and 2012). Previous studies [1, 2, 3, 4, 5] have laid the foundation for comprehending polite speech by investigating various acoustic parameters, consistently emphasizing the importance of pitch, alongside other phonetic features, in the manifestation of politeness. To build upon this foundation, our study narrows its focus to the pitch-related parameters that have consistently demonstrated significance in previous research. Specifically, we analyze the average fundamental frequency (F0) as a measure of long-term F0 and the standard deviation of F0 as a measure of F0 variability. We hypothesize that there will be generational differences in polite speech production and perception among Thai speakers. Specifically, we expect to find significant variations in long-term fundamental frequency and F0 variability, across the three generations. By conducting this study, we aim to contribute novel insights to the understanding of polite speech in the context of Thai language and culture.

In the production experiment, 60 participants (equally distributed between males and females, with 20 participants from each generation) engaged in scripted conversations with different addressees (higher status, equal status, and lower status) and performed various speech acts (inquiry, request, persuasion). Every sentence in the script incorporated a polite participle, ensuring a consistent expression of politeness in participants' speech. Using the same script for all participants controlled for variations in lexical tones, allowing for a focused examination of politeness strategies and their impact on acoustic measures. Acoustic analyses measured the mean of F0 (long-term F0) and the standard deviation of F0 (F0 variability), providing insights into the production patterns of polite speech. Participants' ages were verified to ensure accurate generational categorization.

The production experiment, analyzed using multiple linear regression, found that Generation Z speakers had significantly higher long-term F0 compared to Generation X and Y speakers ($p < .001$). Addressee status did not significantly affect long-term F0 ($p = 0.98$). Moreover, the analysis revealed no significant differences in F0 variability across generations ($p = .23$) and addressee statuses ($p = .94$). These results strongly support the higher long-term F0 in Generation Z compared to other generations, while addressee status did not have a significant impact on F0 variation. Additionally, there were no significant differences in F0 variability among generations and addressee statuses.

The perception experiment involved 60 participants (equal gender distribution, with 20 participants from each generation) who rated their perceived level of politeness for recorded sentences. Notably, 60% of the participants in the perception experiment were the same individuals who had participated in the production experiment. Stimuli comprised two sets: one tested long-term F0, varying in high, mid, and low levels, and the other tested F0 variability, varying in high, mid, and low levels. Ratings were provided on a 7-point Likert scale, capturing the participants' subjective perception of politeness.

In the perception experiment, logistic mixed-effect models were used to analyze the data. Long-term F0 did not significantly influence perceived politeness. However, generation and gender had significant effects. Generation Y and Z participants rated sentences as more polite compared to Generation X participants, and males perceived sentences as more polite than females. Addressee status significantly impacted perception, with lower addressee status resulting in less perceived politeness ($p = .03$). F0 variability demonstrated significant effects of generation ($p < 0.02$) and addressee status ($p = .005$ for equal status, and $p = .01$ for lower status). Generation Z participants showed higher sensitivity to F0 variability, and sentences produced for higher addressee status were perceived as more polite when accompanied by higher F0 variability. Interaction effects of F0 variability, generation, gender, and addressee status were also significant ($p < 0.05$).

Proceedings of the Second International Conference on Tone and Intonation, pages 23–24
November 18–20, 2023.

These findings highlight evolving polite speech patterns among Thai speakers across generations. Generation Z speakers exhibit a higher pitch in production, though not necessarily aligned with perceived politeness. Generation Z's higher pitch in speech production seems to be a generational marker rather than a deliberate politeness strategy. They consistently use a high pitch regardless of the addressee's status. Notably, the analysis found no significant relationship between long-term F0 and perceived politeness. F0 variability emerges as an important factor influencing politeness perception. Understanding these shifting trends in polite tones is crucial for effective communication and intergenerational understanding in the Thai context. The study also highlights the significance of analyzing pitch characteristics in understanding socio-cultural issues more broadly.

References

[1] Chen, A., Carlos, G., & Toni, R. 2004. Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech*, 47(4). 311–349.

[2] Devís, E. H. & Francisco J. C. S. (2014). The intonation of mitigating politeness in Catalan. *Journal of Politeness Research*, 10(1). 127–149.

[3] Idemaru, K., Winter, B., & Brown, L. (2019). Cross-cultural multimodal politeness: The phonetics of Japanese deferential speech in comparison to Korean. *Intercultural Pragmatics*, 16(5). 517-555.

[4] Loveday, L. 1981. Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech*, 24(1). 71–89.

[5] Winter, B. & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics*, 40(6). 808–815.

The prosody of polar response particles in German and Dutch

Sophie Repp & Christiane Ulbrich
University of Cologne
sophie.repp@uni-koeln.de, christiane.ulbrich@uni-koeln.de

Response particles like English *yes* and *no* may express that the proposition introduced by an antecedent utterance is true or false, and they may express that the polarity of the response is positive or negative. These functions come apart after negative antecedents (NegA): A statement like *Tim didn't mow the lawn* in principle may be affirmed with *yes, he didn't* or *no, he didn't* (*yes* → true, *no* → negative polarity), and it may be rejected by *yes, he did* or *no, he did* (*yes* → positive polarity, *no* → false). Thus, *yes* and *no* in principle are ambiguous after NegA. This is also true for languages that additionally have specialized particles, e.g., for rejecting negative antecedents like German *doch* or Dutch *jawel*. Typically, languages show graded preferences for the truth- or the polarity-signaling function of *yes/no* [3][7][8]. The same holds for individual speakers: there usually is considerable inter-individual variation [3][7]. Written acceptability studies have shown that in German (G) most participants rate a truth-signaling particle after NegA more acceptable (*yes, he didn't*) [3], whereas in Dutch (D) most participants rate a polarity-signaling particle more acceptable (*no, he didn't*) [7]. Still, in both languages, the non-preferred particle is fairly or (for some speakers) equally acceptable. These findings raise the question what particle(s) speakers use in production, and if they use different prosodic means to mark the different functions of a particle.

This study investigates the choice and realization of response particles in oral production in G and D. We conducted two translation-equivalent experiments (G: 48 ppl., D: 32, sex balanced). Participants took part in pseudo-dialogues embedded in a larger conversation (48 items). Based on what they knew about the true state of affairs from the conversation, they affirmed or rejected a positive or negative assertion of an interlocutor by using *ja/ja*, *nein/nee* or *doch/jawel* followed by a sentence of their choice. The experiment design was 2×2 design (factors: *antecedent* (NegA/PosA); *speech act* (rejection/affirmation)). We analyzed the acoustic characteristics of the particles in two comparisons. (i) We compared productions of *nein/nee* in rejections of PosA vs. affirmations of NegA; the former is the more marked discourse because rejecting is a face-threatening act and has been suggested to produce a conversational crisis [4]. (ii) We compared productions of *ja* in affirmations of PosA vs. NegA; the latter are more marked because the expressed polarity is negative [8]. The acoustic analysis of the particles was carried out to illuminate the integrity of vocal mechanisms (harmonics-to-noise ratio (HNR), smoothed cepstral peak prominence (CPPs), local jitter) and their interaction with duration (log), mean intensity, f0 mean/max/minimum and excursion, to find out whether certain characteristics bundle (i) in the expression of preferred vs. dispreferred speech acts, which might be associated with different speaker attitudes, and (ii) in the expression of marked structures.

Fig. 1 shows the results for particle choice. After PosA, the choice is clear-cut with *ja/ja* used in affirmations and *nein/nee* in rejections. In affirmations of NegA, G speakers produce truth-indicating *ja* more often than polarity-indicating *nein*, which also occurs. D speakers show the opposite pattern. In G and D, there is inter- and intraindividual variation (not shown). In rejections of NegA, *doch* and *jawel* are clearly preferred, *nein/nee* are infrequent. All results match the acceptability findings. The statistical analysis (LMMs) of the acoustics revealed significant bundles of phonetic features (Table 1). For *nein/nee* in rejections vs. affirmations, tonal measures in both languages were lower, and they align with longer particle duration, but only D speakers modify voice quality. For *ja* in affirmations of NegA vs. PosA, tonal measures were lower in G and D, and they align in duration of the silence after the particle. Voice quality is modified only by G speakers. Thus, for both particles, temporal measures show effects in both comparisons in both languages. To differentiate between the more and the less marked discourses, the temporal characteristics combine with similar tonal, but with different voice quality measures, depending on language and discourse/particle. The longer duration in the marked discourses might be due to increased vocal effort, but we found lower f0, which normally is not associated with higher articulatory effort [2]. The higher HNR and CPPs in the marked discourses might have distinct pragmatic sources for *ja* vs. *nein/nee*. For *nein/nee* in rejections, the lower f0 and higher HNR might reflect the negative attitude [5][6] of the face-threatening speech act. For *ja* in affirmations of NegA, higher CPPs might signal the negative polarity of the answer [1] and the lower f0 measures might reflect markedness of the discourse because low(er) tonal targets are less frequent in both languages.

Proceedings of the Second International Conference on Tone and Intonation, pages 25–26
November 18–20, 2023.

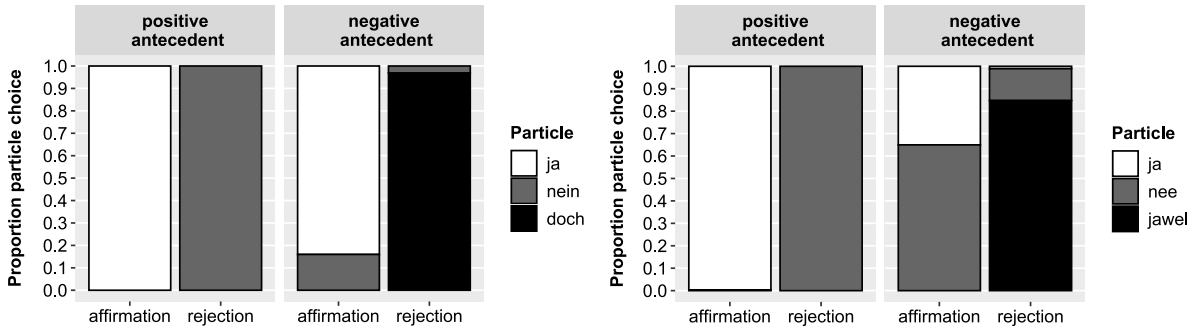


Figure 1: *Distribution of the German particles* ja, nein *and* doch *(left) and the Dutch particles* ja, nee, jawel *(right) by experimental condition*.

Table 1: *Overview of significant acoustic differences for* ja, nein/nee *in the different discourses.*

	NEIN/NEE in rejections vs. affirmations		**JA after NegA vs. PosA**	
	G	**D**	**G**	**D**
particle duration	rej > aff	rej > aff	NegA > PosA	-
duration of silence after particle	-	rej < aff	NegA < PosA	NegA < PosA
f0 maximum	-	-	-	-
f0 minimum	rej < aff (females)	-	-	NegA < PosA
f0 mean	-	rej < aff	NegA < PosA	-
f0 excursion	-	-	NegA < PosA	-
mean intensity	-	rej < aff		NegA < PosA
CPPs	-	-	NegA > PosA	-
HNR	-	rej > aff	NegA > PosA (females)	-
jitter	-	-	-	-

References

[1] Callier, P. 2013. *Linguistic Context and the Social Meaning of Voice Quality Variation.* Georgetown, DC: Georgetown University dissertation.

[2] Chen, A., Gussenhoven, C. & Rietveld, T. 2002. Language-specific uses of the effort code. In *Proceedings of the Speech Prosody*, 215-218.

[3] Claus, B., Meijer, M., Repp, S. & Krifka, M. 2017. Puzzling response particles: An experimental study on the G answering system. *Semantics and Pragmatics* 10, 19:EA.

[4] Farkas, D. F. & Bruce, K. B. 2010. On reacting to assertions and polar questions. *Journal of Semantics*, 27(1), 81-118.

[5] Gobl, C. & Ní Chasaide, A. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40. 189-212.

[6] Johnstone, T. & Scherer, K. R. 2000. Vocal communication of emotion. In Lewis, M. & Haviland, J. M. (Eds.) *Handbook of Emotions*. Guilford, New York, 220-235.

[7] Repp, S. Meijer, M. & Scherf, N. 2019. Responding to negative assertions in Germanic: On *yes* and *no* in English, Dutch and Swedish. In Espinal, M.T. et al. (Eds.) *Proceedings of Sinn und Bedeutung 23*. Vol. 2. Universitat Autònoma de Barcelona, Bellaterra, 267-285.

[8] Roelofsen, F. & Farkas, D. F. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language*, 359-414.

Spanish imperatives produced by proficient Chinese learners of Spanish: The differences in prenuclear pitch accent and boundary tones

Xiaotong Xi[1] & Peng Li[2]
[1]*Universitat Pompeu Fabra*, [2]*University of Olso*
xiaotong.xi@upf.edu, peng.li@iln.uio.no

According to the L2 Intonation Learning theory (LILt), the acquisition of L2 prosody can be influenced by a learner's L1 prosody [1]. Intonation languages like Spanish use pitch for pragmatic purposes, whereas tonal language like Mandarin realizes pitch specifications on lexical level to distinguish semantic meanings, and intonation does not largely distort the pitch contours of lexical tones [2]. As a result, when Mandarin speakers learn distinctive L2 Spanish intonational patterns, they may transfer the pitch specification from the lexical level to the sentence level while neglecting the pitch configurations required for accurate intonation [3]. This study focuses on how Mandarin learners of Spanish realize the intonation of Spanish imperatives. Spanish imperatives can function as commands or requests, characterized by a rise-fall or falling boundary tone. At the morpho-syntactic level, despite the typical imperative mood, wh- and yes-no questions in indicative mood can also function as requests, with a low boundary tone [4]. In contrast, Chinese imperatives (commands and requests) are mainly marked by a low/falling boundary tone, and the use of wh- or yes-no questions for conveying requests is infrequent [5]. In the prenuclear position, Spanish imperatives may show a pitch accent of L+<H*. However, L+<H* is not attested in Chinese. Based on the contrastive analysis between Chinese and Spanish, our first hypothesis is that Chinese students may realize imperative wh- and yes-no questions with falling or rising boundary tones, respectively, as they do with information-seeking questions [3]. Moreover, the F0 peak of L+<H* aligns with the post-accentual syllable, but Mandarin learners of Spanish consistently realize stressed syllables with a high pitch, both at the lexical level and in running speech [6]. Therefore, our second hypothesis is that Chinese students would not show the prenuclear pitch accent of L+<H* but rather produce the F0 peak within the stressed syllable.

A total of 16 Mandarin speakers (12 females) with advanced Spanish proficiency and 9 native Castilian Spanish speakers (6 females) participated in a discourse completion task. The task contained 12 contexts designed to elicit four types of sentences: imperative command (imp-comm), imperative request (imp-req), imperative wh- question (imp-wh), and imperative yes-no question (imp-y/n). We extracted 10 equidistant pitch points from each phoneme and generated time-normalized pitch contours using *z*-scored F0 values. For each of the 12 sentences, we built a Generalized Additive Mixed Model (GAMM) where the dependent variable was the normalized F0. The independent variables included a smooth for the interaction of speaker group (Chinese students vs. Spanish natives), a smooth for the interaction of gender (female vs. male) over time (normalized pitch points), and a random smooth for each individual speaker. Figure 1 illustrates the post-hoc comparisons between speaker groups across the four sentence types.

The GAMMs demonstrated significant differences in F0 contours between Chinese students and Spanish natives across all 12 sentences. The post-hoc comparisons are summarized as follows: Firstly, both Chinese students and Spanish natives produced similar low/falling boundary tones in imp-comm, imp-req, and imp-wh. Surprisingly, contrary to the expected low boundary tone reported in [4], Spanish natives produced a rising boundary tone in imp-y/n, possibly as a politeness strategy [7]. Chinese students, on the other hand, did not show consistent patterns for boundary tone, employing either a rising tone or a low falling tone. Secondly, as anticipated, Spanish natives produced an F0 peak at the post-accentual syllable in imp-wh and imp-y/n. However, Chinese students produced an F0 peak within the accented syllable. Thirdly, in imp-req with a Verb-Object structure, Mandarin speakers tended to realize the F0 peak in the accented syllable of the imperative verb. This may be attributed to the influence of their L1 Mandarin, where the predicate verb in imperative sentences is focused [8].

All in all, the results highlight that Mandarin speakers tend to exhibit L1 prosodic patterns in their L2 Spanish speech, which contributes further evidence to the crosslinguistic influence in speech prosody, supporting the predictions of the LILt. This finding underscores the importance of paying closer attention to the intonational patterns of Spanish during the learning process.

Proceedings of the Second International Conference on Tone and Intonation, pages 27–28
November 18–20, 2023.

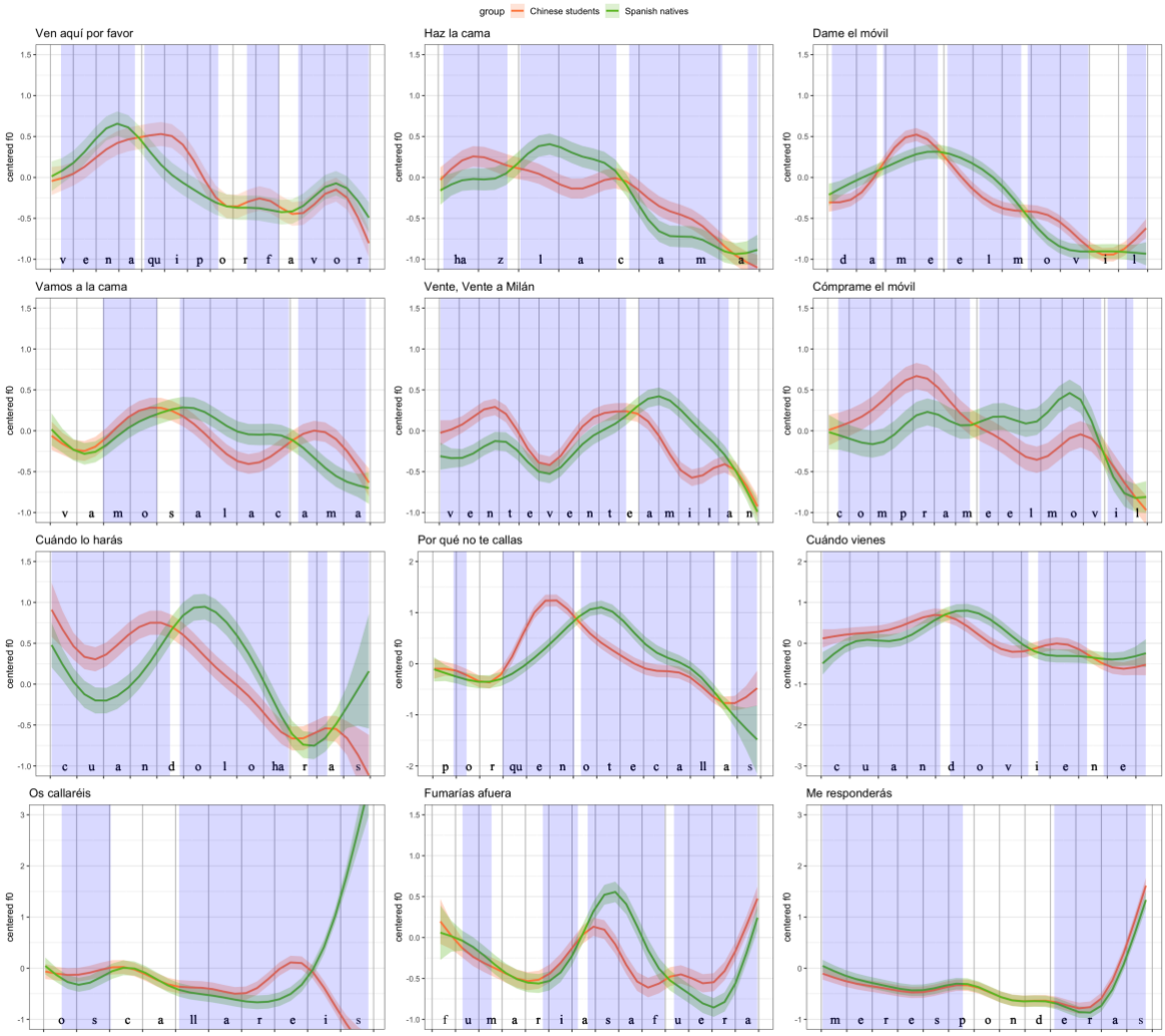


Figure 1: *Groupwise comparisons of F0 contours of the Spanish sentences estimated by GAMM for each sentence type, from top to bottom: imp-comm, imp-req, imp-wh, and imp-y/n. The shaded area of each contour paints the 95% Confidence Interval. The purple squared shades illustrate significant contrasts in intonation contours between Chinese students and Spanish natives.*

References

[1] I. Mennen, "Beyond Segments: Towards a L2 Intonation Learning Theory," in *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*, E. Delais-Roussarie, M. Avanzi, and S. Herment, Eds., Springer Link, pp. 171–188.

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

[3] P. Li and X. Xi, "Nuclear contours of Spanish echo questions produced by proficient Chinese learners of Spanish: a dynamic analysis," in *Proceedings of the 20th International Congress of the Phonetic Sciences*, 2023.

[4] E. Estebas-Vilaplana and P. Prieto, "Castilian Spanish intonation," in *Transcription of intonation of the Spanish language*, P. Prieto and P. Roseano, Eds., LINCOM Europa, 2010, pp. 17–48.

[5] P. Shi, "四种句子的语调变化[Intonation of four types of sentences]," *语言教学与研究 Language Teach. Linguist. Stud.*, vol. 2, pp. 71–81, 1980.

[6] P. Li and X. Xi, "Spanish lexical stress produced by proficient Mandarin learners of Spanish," in *Proceedings of the 4th International Symposium on Applied Phonetics*, ISCA, Sep. 2022, pp. 40–45. doi: 10.21437/ISAPh.2022-8.

[7] L. Astruc and M. D. M. Vanrell, "Intonational phonology and politeness in L1 and L2 Spanish," *Probus*, vol. 28, no. 1, 2016, doi: 10.1515/probus-2016-0005.

[8] F. Shi and X. Jiao, "普通话命令句语调的时长和音量分析[An analysis of duration and intensity of Mandarin imperatives]," *Chin. Lang. Learn.*, vol. 1, pp. 65–73, 2016.

The Interaction of Tonal and Metrical Prominence in the Pingding Variety of Chinese

Pingping Jia
Free University of Berlin
Pingping.jia2@gmail.com

The present study addresses the question of how metrical prominence is realized in tonal languages. The data come from a field study conducted in 2019 on the Pingding Variety of Chinese and have been published as a corpus at the Repository from the Free University of Berlin – Refubium (Jia 2021). Pingding is a variety of Jin Chinese spoken in an area surrounded by Mandarin varieties (CASS 2012). It has often been claimed that tonal languages do not have word stress or metrical prominence, and that lexical tone and word stress are mutually exclusive features (cf. the discussion in Hyman 2006; but see Hyman 2014 for counter-arguments; cf. Sui 2016, Feng 2016, Duanmu 2022 on Chinese). This study aims to demonstrate the interaction between metrical prominence and tonal prominence, as is visible in the form of tone deletion in metrically weak, but (underlyingly) tonally prominent syllables in the Pingding Variety. The generalizations on tone deletion receive a straightforward analysis if we assume (i) that metrical prominence is on the leftmost syllable of a phonological word, and (ii) that tones are ranked according to a scale of tonal prominence, (4), where contour tones are more prominent than level tones and high tones are more prominent than non-high tones.

Firstly, assuming that it is applied in Pingding Variety that a natural foot (disyllabic trochee) maps to a phonological word generally in Chinese (Feng 1998), then we can see that a disyllabic phonological word (a two-tone sequence as well) can be either a disyllabic lexeme or a disyllabic phrase formed by combining two monosyllabic lexemes. The Pingding data show that if, in a lexeme, the citation tone of a metrically non-prominent ("unstressed") syllable is more prominent than the tone of the metrically prominent ("stressed") syllable, the tone of the "unstressed" syllable is deleted; (1a). As a result, the citation tone pattern MF-HF is realized as MF-o (‘o’ referring to the tone that is deleted). However, tone deletion is only observed within lexemes but not in disyllabic phrases exhibiting the same tonal pattern; (1b). Therefore, tone deletion is only applied to the part of phonological words that map to lexemes, but not to the part that map to phrases.

Secondly, tone deletion is applied only if the "unstressed" syllable bears a citation tone that is more prominent than the citation tone of the "stressed" syllable. For example, in the citation tone sequence HF-MF, where the HF on the "stressed" syllable is more prominent than the MF on the "unstressed" syllable, both tones are preserved (but the sandhi process applies) in phonological words, no matter they are lexemes or phrases; (2).

Thirdly, the same phenomenon can be observed in trisyllabic sequences and even longer monomorphemic lexemes, deleting (all and only those) tones that are more prominent than the tone of the "stressed" syllable; (3). This finding seems to imply that the structural prominence triggering tone deletion in lexemes might be word accent rather than the strong syllable in a trochaic disyllabic foot.

The tonal prominence hierarchy observed in this variety is compatible with standard assumptions on the cognitive saliency of high tones and contour tones, which are generally more perceptually salient than low tones and level tones, see (4) (Jiang-King 1996; Jiang-King 1999; de Lacy 2002; Zhang 2007). The data from Pingding thus constitute another piece of evidence that both lexical tone and metrical prominence at the level of the phonological word can coexist in one phonological system, where metrical prominence can be perceived by the speakers as tonal prominence (and is thus learnable). If underlying tonal prominence does not align with (surface) metrical prominence, tone deletion applies, with metrical prominence winning over tonal faithfulness. The analysis presented in this study for the data is cast in the framework of Optimality Theory.

Proceedings of the Second International Conference on Tone and Intonation, pages 29–30

Examples

(1) a. Lexeme MF-HF→MF-o (tone deletion)

yao-gu	[iɔ^{42}ku$^{53\text{-}22}$]	"waist-drum"	Noun
xin-ku	[ɕiəŋ^{42}k^{h}u$^{53\text{-}22}$]	"toilsome"	Adj.

b. Phrase MF-HF→MF-HF (no tone deletion)

zhe-yu	[tʂɤr^{42}y^{53}]	cover + rain	"to keep out the rain"	Verb + Obj.
zhua-jin	[tsuɑ^{42} tɕiəŋ^{53}]	hold + tight	"grasp firmly"	Verb + Adv.

(2) HF-MF→HH-MF (no tone deletion, but sandhi in the form of contour dissimilation)

a. Lexeme

shou-jin	[ʂəo $^{53\text{-}55}$ tɕiəŋ^{42}]	"towel"	Noun
pu-tong	[p^{h}u$^{53\text{-}55}$t^{h}uəŋ^{42}]	"normal"	adj.

b. Phrase

yan-hua	[iæ̃$^{53\text{-}55}$xuɑ^{42}]	eye + dim	"have dim eyesight"	Subj. + Predicate
qi-feng	[tɕhi$^{53\text{-}55}$fəŋ^{42}]	start + wind	"gust"	Verb + Obj.

(3) Tone deletion in trisyllabic or longer segments

yang-gao-feng [iɑŋ^{22}kɔ$^{42\text{-}22}$fəŋ$^{42\text{-}22}$] "epilepsy"

LL-MF-MF → LL-o-o

duo-lun-duo [tuɤ^{42}luəŋ^{22}tuɤ$^{42\text{-}22}$] "Toronto"

MF-LL-MF → MF-LL-o

xi-ma-la-ya [ɕi$^{53\text{-}42}$mɑ$^{53\text{-}22}$lɑ$^{42\text{-}22}$iɑ$^{53\text{-}22}$] "Himalayas"

HF*-HF-MF-HF → MF-o-o-o *sandhi rule (HF→MF/__HF) is applied before tone deletion

(4) Tonal prominence hierarchy

HF>>MF>>LF/LFq>>LL/Lq

(HF: high falling tone; MF: mid falling tone; LF: low falling tone; LFq: low falling checked tone; LL: low level tone; Lq: low level checked tone)

References

[1] Duanmu, S. 2022. Evidence for Stress and Metrical Structure in Chinese. In *The Cambridge handbook of Chinese linguistics*, 361–382. Cambridge University Press.

[2] Feng, S. 1998. Discussion on default foot in Chinese (论汉语的"自然音步 "). *Studies of the Chinese Language (中国语文)* 1998(1). 40–47.

[3] Feng, S. 2016. Mandarin Chinese is a stress language (北京话是一个重音语言). *Linguistic Sciences (语言科学)* 15(5). 449–473.

[4] Hyman, L. 2006. Word-prosodic typology. *Phonology* 23(02). 225–257.

[5] Hyman, L. 2014. Do All Languages Have Word Accent. In Harry van der Hulst (Ed.), *Word Stress. Theoretical and Typological Issues*, 56–82. Cambridge: Cambridge University Press.

[6] Jia, P. 2021. A comprehensive corpus of three varieties of Jin Dialects, Chinese (Yu Dialect, Pingding Dialect, Jiaoqu Dialect). https://refubium.fu-berlin.de/handle/fub188/31554.

[7] Jiang-King, P. 1996. *An optimality account of tone-vowel interaction in Northern Min*. The University of British Columbia dissertation.

[8] Jiang-King, P. 1999. Sonority constraints on tonal distributions across Chinese dialects. In *Proceedings of WCCFL 17*, 332–346. Stanford University: CSLI.

[9] De Lacy, P. 2002. The interaction of tone and stress in Optimality Theory. *Phonology* 19(01). 1–32.

[10] Sui, Y. 2016. The Interaction of Metrical Structure and Tone in Standard Chinese. In Jeffrey Heinz, Rob Goedemans & Harry van der Hulst (Eds.), *Dimensions of Phonological Stress*, 101–122. Cambridge: Cambridge University Press.

[11] CASS. The Institute of Linguistics from CASS, The Institute of Ethnology and Anthropology from CASS & City University of Hong Kong. 2012. *Language atlas of China (中国语言地图集)*. 2nd edn. Beijing: The Commercial Press (商务印书馆).

[12] Zhang, J. 2007. A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*. Vol. 16.

The Attractivity of Average Speech Rhythm in Mandarin Chinese

Gyong Min Oh, Chun Wang and Constantijn Kaland
University of Cologne
goh@smail.uni-koeln.de, cwang8@smail.uni-koeln.de, ckaland@uni-koeln.de

The current study explores the perceptual preferences of listeners concerning speech rhythm [1][2] and investigate the so-called "averaging effect" [3] of speech. The central hypothesis is that listeners will prefer speech characterized by an average temporal pattern [4] across different speakers. This research is built upon prior studies that have examined the aesthetic nature of the averaging effect in diverse domains, such as faces [5], music, and voice [6], revealing a general tendency among observers to favor averages in the performance or appearance of various human and non-human phenomena. Additionally, this study extends a study conducted with Dutch speakers that yielded very promising results. [7] The current study investigates whether the "averaging affect" [8] is indeed prevalent in a domain that is yet to be investigated, which is Mandarin Chinese.

The study was split into two steps, data collection and manipulation, and two forced-choice perceptual tests in the form of a survey. The recordings of eight different native standard Mandarin (Pu Tong Hua) speakers saying the same first phrase from "North Wind and the Sun" were collected. Then, each of the phrases were broken down to syllables and significant pauses, which resulted in a total of 24 intervals. Each interval's mean across the eight speakers were calculated, and the speaker that deviated the furthest and least from the mean were selected as a carrier phrase for two separate surveys. A carrier phrase was chosen so that the participants choices are not influenced by anything other than the rhythm of each stimulus.

The two carrier phrases were then manipulated so that the duration of each syllable identically matched the remaining seven speakers to create seven stimuli. The eighth and final stimulus was one that was manipulated to match the recalculated mean duration, omitting the carrier phrase. Two carrier phrases were selected instead of one to see if the results could be replicated under two different circumstances, especially as the degree of manipulation would be greater in the data set created by the carrier phrase with the furthest mean and therefore sound more unnatural. This was specifically included due to a concern of the previous study conducted with Dutch speakers [8] where the carrier phrase could have been chosen due to it sounding more natural, as it underwent the least manipulation.

With the completed manipulated data set, the eight rhythm-varying stimuli were presented in pairs in a force choice survey, making both surveys consist of sixty four questions. Twenty five native Mandarin speakers between the ages of eighteen to thirty without any hearing issues were then gathered for each of the perception tests and were instructed to compare and select which stimuli in the pair they found more attractive in terms of rhythm. Participants who did not assign their age or took too short to answer the questions (e.g., the decision time was shorter than the duration of the stimuli) were excluded from the analysis. Of the fifty participants, twenty results per survey were analyzed. The results are shown in Table 1. Then, participants' responses were analyzed using a choice ratio, which is calculated by how many times the target stimulus was chosen out of the total number of times it was compared. Because an audio is compared with the rest of the seven audios once as option A and once as B in the forced choice perception survey, the total number of times compared was 14.

As seen in Figure 1, the findings indicated that rhythms with syllable durations that deviated further from the mean were generally perceived as less attractive. The red circle refers to the stimulus with the smallest deviation from the mean and the blue to the stimulus with the largest deviation from the mean. Conversely, when syllable duration was only slightly deviating from the mean, it tended to be rated as more attractive. Thus, the study contributes to our understanding of the generalizability of the averaging effect in the evaluation of temporal variability in speech, specifically in the context of Mandarin. The results of this study emphasize the significance of rhythmic patterns in speech and the way that these patterns can shape the listener's experience of the speech. This could potentially have new implications for the role of phonetics in psychology. Additionally, the findings suggest that speakers must strike a balance between variability and uniformity in syllable duration to create an engaging and flowing speech rhythm. The study underscores the value of investigating perceptual preferences in diverse domains, which may facilitate a more comprehensive understanding of the underlying cognitive mechanisms that guide human behavior, their perception of attractiveness, and decision-making.

Proceedings of the Second International Conference on Tone and Intonation, pages 31–32
November 18–20, 2023.

A \ B	M_01	F_02	M_02	F_01	F_03	M_03	F_04	Mean	sum A
M_01		18/2	12/8	4/16	4/16	6/14	5/15	5/15	54
F_02	4/16		3/17	1/19	2/18	1/19	2/18	1/19	14
M_02	4/16	14/6		2/18	3/17	4/16	2/18	3/17	32
F_01	15/5	17/3	15/5		10/10	9/11	9/11	2/18	77
F_03	12/8	16/4	20/0	8/12		11/9	8/12	5/15	80
M_03	16/4	19/1	14/6	15/5	12/8		10/10	6/14	92
F_04	13/7	18/2	16/4	8/12	13/7	9/11		5/15	82
Mean	13/7	19/1	17/3	15/5	16/4	9/11	15/5		104
sum B	63	19	43	87	80	91	89	113	1120

A \ B	F_02	M_02	F_01	F_03	M_03	F_06	F_04	Mean	sum A
F_02		4/16	1/19	1/19	4/16	1/19	2/18	2/18	15
M_02	15/5		2/18	1/19	4/16	0/20	2/18	3/17	27
F_01	20/0	19/1		12/8	12/8	9/11	14/6	9/11	95
F_03	17/3	19/1	3/17		15/5	6/14	11/9	12/8	83
M_03	19/1	16/4	9/11	10/10		6/14	6/14	7/13	73
F_06	18/2	16/4	8/12	13/7	17/3		12/8	10/10	94
F_04	19/1	16/4	4/16	5/15	12/8	5/15		8/12	69
Mean	20/0	17/3	7/13	13/7	15/5	9/11	9/11		90
sum B	12	33	106	85	61	104	84	89	1120

Table 1: *The **choice counts of stimuli in two surveys***

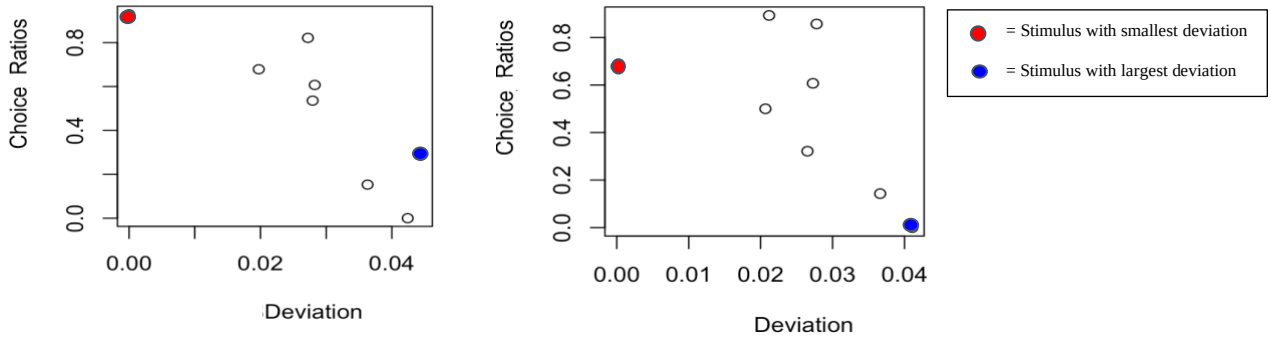


Figure 1: *The distribution of stimuli in two surveys*

References

[1] Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica, 66(1-2)*, 46-63.

[2] Díaz Romero, C.E (2020). On the Concept of Rhythm in Phonology. In Ö. Özturk (Ed.), Social Science Conferences 2020 Spring (pp.63-78).

[3] Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological science, 1(2)*, 115-121.

[4] Nespor, M. (1990). On the rhythm parameter in phonology. *Logical issues in language acquisition, 157-175.*

[5] Langlois, J. H., Roggman, L. A., & Musselman, L. (1994). What is average and what is not average about attractive faces?. *Psychological science, 5(4)*, 214-220.

[6] Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., ... & Belin, P. (2010). Vocal attractiveness increases by averaging. *Current biology, 20(2)*, 116-120.

[7] Kaland, C., & Swerts, M. (under review). The attractiveness of average speech rhythms: Revisiting the average effect from a crosslinguistic perspective.

[8] Swerts, M., & Kaland, C. (2023). Mean Rhythm. Listeners Prefer Speech with Quantitatively Average Syllable Durations [Preprint]. SSRN. https://doi.org/10.2139/ssrn.4472289

The phonetics and pragmatics of H* and L+H* in British English

Jiseung Kim, Na Hu, Riccardo Orrico, Stella Gryllia, Amalia Arvaniti
Radboud University
jiseung.kim@ru.nl; na.hu@ru.nl; riccardo.orrico@ru.nl;
stella.gryllia@ru.nl; amalia.arvaniti@ru.nl

In English intonation, a distinction is often posited between two pitch accents, H* and L+H*: H* is realized as high pitch and encodes new information, while L+H* is realized as rising pitch and used contrastively [1]. However, the distinction is not generally accepted: for instance, [2] argues that L+H* is just an emphatic rendition of H*, not a distinct accentual category. Further, intonation is subject to dialectal variation [3], and descriptions of Southern British English (SBE) intonation, the variety investigated here, usually argue for the presence of only falls, with high falls largely corresponding to L+H* and low falls to H* (e.g., [4]). Based on the above, our aim was twofold: first, to provide empirical evidence regarding the shape of the accents in SBE, and second to test the hypothesis stemming from [1] regarding their function. If H* and L+H* are phonologically distinct, each accent should have a distinct form that should largely correspond to a specific function related to information-structure.

We investigated this hypothesis by examining 2,450 accents elicited from 8 native speakers of SBE producing unscripted speech in 3 tasks: each participant created and narrated three short stories, and pairs of speakers participated in a map task and an informal discussion. We used unscripted speech because it is less controlled than lab speech but reflects more realistically how individual speakers encode information structure relative to lab-elicited scripted speech where speakers are often reminded of intended differences in meaning – either explicitly or implicitly. The data were annotated separately for the phonetic and information-structure dimensions to avoid phonetic classification being guided by pragmatic meaning and vice versa. The phonetic annotation was based on f0 shape only: accents were annotated as L+H* if they included a deliberate f0 dip at the onset of the accented syllable, and as H* otherwise. The pragmatic annotation was based on orthographic transcripts only: items were annotated as corrective if they were an explicit correction of a previously mentioned item, and as contrastive if they were part of an implicit set of alternatives. Next, each item accented with H* or L+H* (based on the phonetic criteria above) received a separate pragmatic label: *corrective* or *contrastive* if marked as such in the orthographic transcript, and *new* otherwise (see Fig. 1). A Generalized Additive Mixed-effects Model (GAMM) fitted the normalized f0 curves of the annotated items using the *mgcv* [6] and *itsadug* [7] packages in R [8], with accent (H*, L+H*) and pragmatics (new, contrastive, corrective) as fixed intercepts and smooth terms, and speaker as a factor smooth (random intercept and slope). Functional Principal Component Analysis (FPCA; [5]) was used to normalize for time (using the onset of the accented vowel as a temporal landmark), and for speaker.

The smooth curves in Fig. 2A showed that both accentual falls (H*s) and rise-falls (L+H*s) were attested in these SBE unscripted data. The range of significant difference between the two accentual curves (line a) included the accented vowel, which is expected to carry the bulk of the difference between the two accents. On the other hand, the difference between the pragmatic conditions (Fig. 2B) was significant only between new and corrective items for a range that included the accented vowel (line b), but not between new and contrastive, or between contrastive and corrective (for which the difference was significant but for a range not inclusive of the accented vowel; line c). The lack of clear delineation between the F0 curves of new vs. contrastive items, and between contrastive vs. corrective items indicates that the information-structure based distinction between H* and L+H* posited by [1] for American English does not hold for SBE. This could be a variety-specific difference, but it also raises the possibility that there are several degrees of contrastivity [cf. 9, 10] than previously assumed in intonation studies, where the distinction between new vs. contrastive information is often seen as binary. Further, our data indicate that in SBE, L+H* is also used to mark unexpectedness (Fig. 1), and intensify the meaning of items such as *really, as well* etc.: an additional indication that the function of L+H* is not exclusively related to information structure.

In conclusion, by separating the shape-based and meaning-based annotation procedures, the current study showed that H* and L+H* are distinct in SBE in terms of f0 shape but the two shapes do not map one-to-one with information-structure. Moreover, our findings underline the importance of using different types of speaking tasks, a practice that allows for the observation of greater variation of f0 shape and accent function in discourse. Lastly, the data-driven approach taken in the current analysis can be applied to other types of contrasts lacking clear evidence for a phonological distinction.

Proceedings of the Second International Conference on Tone and Intonation, pages 33–34
November 18–20, 2023.

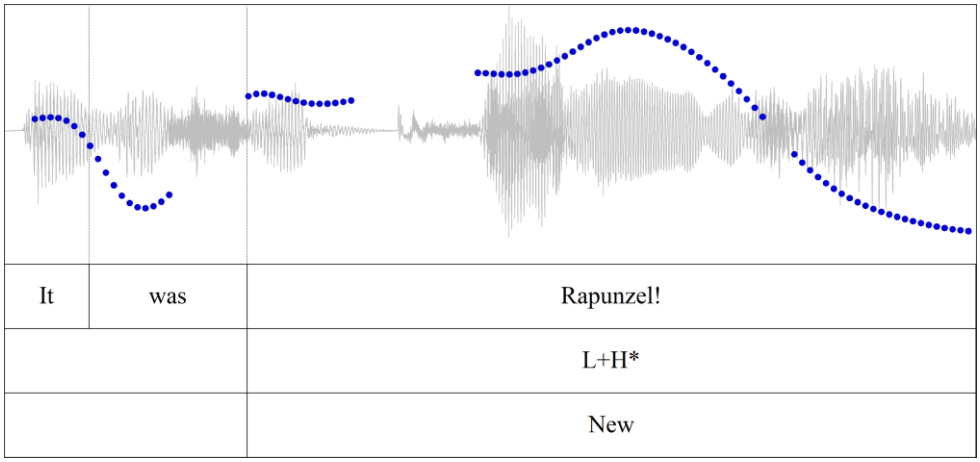


Figure 1: *A sample utterance from the storytelling task showing the phonetic and pragmatic annotation tiers.*

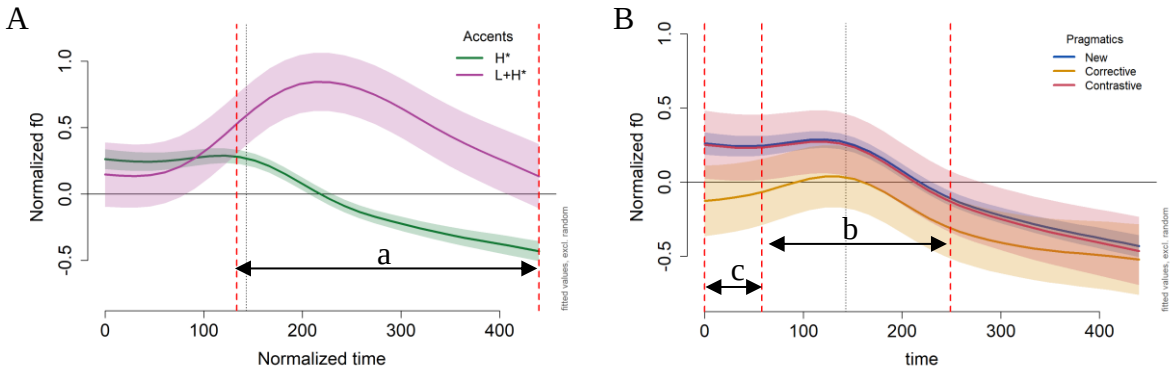


Figure 2: *The smooth plots from the GAMM, showing two accent smooths (left) with the range of significant difference (marked with red vertical lines) between 133 – 439 ms (line a), and three pragmatics smooths (right) with the ranges of significant difference between 0 – 249 ms for new vs. corrective accents (line b) and 0 – 58 ms for contrastive vs. corrective accents (line c). The gray vertical lines indicate the onset of the accented vowel.*

References

[1] Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In: P. Cohen, J. Morgan, & M. Pollack (eds.), *Intentions in Communication*, 271–311.

[2] Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

[3] Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. In: J. Cole & J. I. Hualde (eds.), *Papers in Laboratory Phonology 9*, pp. 547–576.

[4] Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

[5] Gubian, M., Torreira, F., & Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, 49, 16–40.

[6] Wood, S. (2017). *Generalized Additive Models: An Introduction with R*, 2nd edition. Chapman and Hall/CRC.

[7] van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2022). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.4.1.

[8] R Core Team, (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, [Online]. Available: http://www.r-project.org/.

[9] Cruschina, S. (2021). The greater the contrast, the greater the potential: On the effects of focus in syntax. *Glossa: A Journal of General Linguistics*, 6(1), 3. doi: https://doi.org/10.5334/gjgl.1100.

[10] Borràs-Comes, J., Vanrell, M. M., & Prieto, P. (2014). The role of pitch range in establishing intonational contrasts. *Journal of the International Phonetic Association*, 44(1), 1–20.

The prosody of contrastive focus and VERUM focus in rejections

Heiko Seeliger[1] & Sophie Repp[1]
[1]*University of Cologne*
<heiko.seeliger, sophie.repp>@uni-koeln.de

Rejections are speech acts by which a speaker signals that they do not accept a proposition *p* associated with the previous utterance into the common ground. Rejections may simply negate the truth of *p*, e.g., by *No!*, and then the speaker may add a correction. Rejections may also directly correct the previous utterance. Hence, corrections are a subtype of rejections. Corrections typically contain (narrow) contrastive focus, or, if the rejected proposition is negative ($\neg p$), they may also contain so-called VERUM focus, by which a speaker highlights the truth of *p*.

In German, contrastive narrow focus has been proposed to be marked with increased prosodic prominence of the accent on the focused expression compared to non-contrastive narrow focus [1, 2, 3]. However, the manipulation of the focus structure of the target utterance involved a confounded manipulation of contrast and correction: Narrow non-contrastive focus was elicited in assertions (e.g., A: *What does Nina want to tailor?* B: *Nina wants to tailor blouses*$_{Foc}$), and contrastive focus in corrections (e.g., *Does Nina want to tailor trousers?* B: *Nina ... blouses*$_{Foc}$). Thus, it is unclear if the observed prominence increase is a result of contrast marking or of speech act marking. The present study aims to disentangle the effects of contrast and of speech act by investigating focus types in rejections only: If we find effects of contrast, contrast marking is not dependent on speech act marking.

The second aim of this study is to investigate the prosody of German rejections more generally. For other languages (e.g., English and Catalan), rejections have been observed to be marked by the so-called contradiction contour (a rise-fall-rise), but this contour is also associated with other meaning components (e.g., obviousness). For German, no such contradiction contour has been described but there is evidence on the prosody of VERUM focus, which is marked by an accent on the finite verb, with different accent types depending on discourse context [4, 5, 6]. For perception, [7] found that for the marking of VERUM in utterances rejecting negative assertions L*+H is most appropriate, L+H* less appropriate and H* least appropriate. [7] also found that prosodic marking of lexical contrast in corrections is judged as most appropriate with L+H* and least appropriate with L*+H.

We present evidence from a production study (24 participants; 1116 utterances), which compared rejections whose information structure was manipulated by the context in a triadic pseudo-dialogue. The target sentences were transitive with an auxiliary, a lexical verb, an adverb and a modal particle (signaling that the addressee should already be aware of the truth of the sentence); see Table 1. There were four conditions: the object and lexical verb were either both new (O_NV_N = broad focus) or both given (O_GV_G), or there was narrow contrastive focus on either the object (O_CV_G) or the lexical verb (O_GV_C). In broad-focus assertions, the nuclear accent in a transitive sentence would be on the object.

Fig. 1 shows nuclear accent locations by condition. In O_NV_N, most nuclear accents were on the object, but there was also a high proportion of VERUM focus marking on the auxiliary (19% of utterances) and late nuclear accents on the lexical verb (11%). In O_GV_G, most utterances contained VERUM focus marking (52%) or an accent on *doch* (31%). Finally, contrastive focus on the object or the lexical verb attracted the nuclear accent to that element, with a stronger tendency for the object (97%) than for the lexical verb (79%). VERUM focus was very rare in either narrow-focus condition.

Regarding accent types, we focus here on L+H*, which was associated with contrast, as expected. The proportion of L+H* (relative to H*) significantly increased for contrastive relative to new objects (mixed models; $p < 0.001$), and L+H* was by far the most common accent type on contrastive verbs (Fig. 3). However, L+H* was also more common than H* on new objects, and commonly occurred with VERUM focus (in line with [7]). Comparing accented objects in broad focus (O_NV_N) and contrastive focus, the stressed syllables of contrastive objects were significantly longer than those of new objects ($p < 0.001$). Pitch excursion was higher, but not significantly so (Fig. 2).

Overall, we find accentuation patterns in rejections that are different from what is known for assertions: In the absence of lexical contrast, there frequently is VERUM focus marking as well as accentuation of the modal particle *doch* (an interesting finding because unaccented and accented *doch* have been argued to have different meanings [8]). As for the relationship between contrast and speech act marking (corrections), we find that contrast marking is independent of speech act marking.

Proceedings of the Second International Conference on Tone and Intonation, pages 35–36
November 18–20, 2023.

Table 1: *Overview of experimental contexts for one item*

O_NV_N:	I don't think that Nina wants to make clothes herself in that workshop.
O_CV_G / O_GV_C:	I think that Nina wants to {tailor trousers / embroider blouses}.
O_GV_G:	I hope Nina doesn't want to tailor blouses in that workshop.
Target:	Nina will da doch Blusen schneidern!
	Nina wants there MP blouses tailor
	'Nina (does) want(s) to tailor blouses there!'

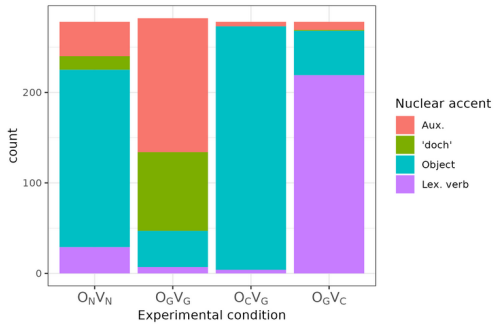


Figure 1: *Nuclear accent location by experimental condition.*

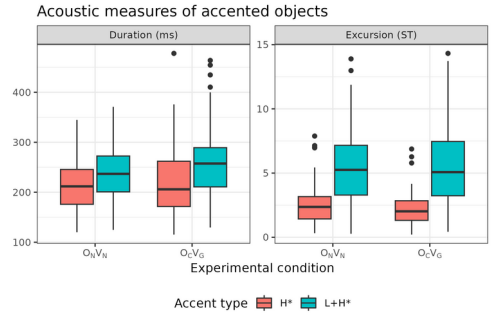


Figure 2: *Acoustic measures of accented objects, broad vs. contrastive object focus.*

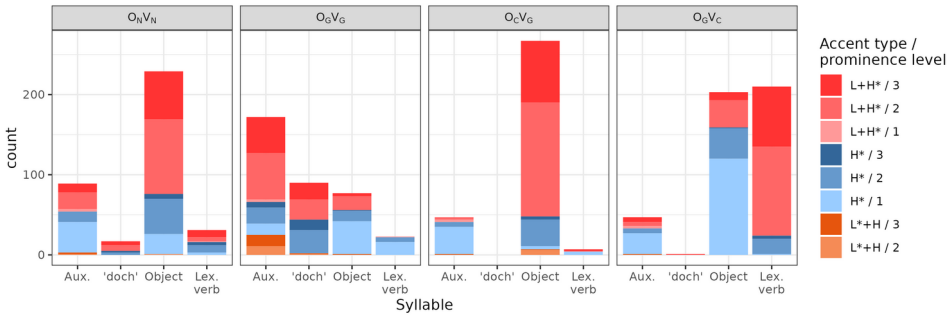


Figure 3: *Combinations of accent type and prominence levels, most common accent types only.*

References

[1] Baumann, S., Grice, M. & Steindamm, S. 2006. Prosodic marking of focus domains – categorical or gradient? *Proceedings of Speech Prosody 2006,* 301-304.

[2] Baumann, S., Becker, J., Grice, M. & Mücke, D. 2007. Tonal and articulatory marking of focus in German. *Proceedings of 16th ICPhS*, 1029-1032.

[3] Baumann, S. & Röhr, C. 2015. The perceptual prominence of pitch accent types in German. *Proceedings of 18th ICPhS*, 298, 1-5.

[4] Turco, G., Braun, B. & Dimroth, C. 2014. When contrasting polarity, the Dutch use particles, Germans intonation. *Journal of Pragmatics, 62*, 94-106

[5] Turco, G., Dimroth, C. & Braun, B. 2013. Intonational means to mark verum focus in German and French. *Language and Speech, 56*, 461-491.

[6] Seeliger, H. & Repp, S. 2023. Information-structural surprises? Contrast, givenness and (the lack of) accent shift and deaccentuation in non-assertive speech acts. *Laboratory Phonology, 14*(1), 1-46.

[7] Röhr, C. T., Grice, M. & Baumann, S. 2023. Intonational preferences for lexical contrast and verum focus. *Proceedings of 20th ICPhS*.

[8] Egg, M. & Zimmermann, M. 2012. Stressed out! Accented discourse particles - the case of DOCH. *Proceedings of Sinn und Bedeutung 16*, 225-238.

The Effects of Tone Types and Bilingual Experiences on Attentional Control in Cantonese Tone Dichotic Listening Task

Yuqi Wang, Zhen Qin
The Hong Kong University of Science and Technology
ywangqi@connect.ust.hk, hmzqin@ust.hk

Bilingual individuals are suggested to outperform monolinguals in attentional control by the practice of focusing on one language and suppressing unused languages. Soveri et al. [1] used the Forced-attention Dichotic Listening (FADL) task to test how monolinguals and bilinguals showed different performances in consonant processing. Participants were simultaneously exposed to two stimuli and to respond according to conditions. The non-forced (NF) condition asked for a clearer stimulus and introduced a right-ear advantage (REA). Both forced conditions (forced-left, FL; forced-right, FR) required attentional control with differed demands by focusing on and reporting stimuli in the instructed ear. The FL called for higher demands, as its incongruent pattern (i.e., left-ear advantage, LEA) with the NF, than the FR. Bilinguals performed better than monolinguals, with greater accuracy improvements in instructed ears from NF to forced conditions (e.g., in the left ear from NF to FL condition) [1]. However, the between-group design overlooked individual bilingual experiences, which should be a continuous variable representing L1-dominant (monolingual-like) to balanced bilinguals [2]. Additionally, the ear advantage of tone processing in the NF differs from that of segments (e.g., consonants) and varies with tone types [3]. For example, Cantonese tones triggered an overall LEA. But contour tones processing showed a stronger REA compared to level tones [4]. Therefore, this study explored Cantonese-English bilinguals' attention control abilities for two types of Cantonese tones and the impact of individuals' bilingual experience on the FADL task.

This study recruited 60 Cantonese-English bilingual participants between 18 and 25 in Hong Kong. As shown in Fig.1, the participants completed a language history questionnaire [5], a tone training, and the target FADL task. The Multilingual Language Diversity (MLD) score from the questionnaire measured bilingual experiences, as it considers dominance and proficiency of all languages (max:4) the participants learned. The stimuli were produced by a speaker who did not merge Cantonese tones, and all stimuli were normalized at 511 ms and 70 dB. The tone training involved identifying contour and level tones, with feedback to enhance the mapping of tones and labels (T1-T6). In the FADL task, participants were presented with dichotic stimuli of contour tone pairs (e.g., /ji2/ 'chair' vs. /ji4/ 'son') and level tone pairs (e.g., /ji1/ 'doctor' vs. /ji3/ 'meaning'). In each trial, participants were required to report the tone played to the instructed ear in each attention condition (NF, FL, and FR) by pressing keys (1-6).

For the tone type effect of attention control abilities, the accuracy of ears (left, right) was compared across conditions (NF, FL, FR) and tone types (contour, level). Results of an ANOVA showed an interaction of tone type, ear, and condition ($F(2,708) = 15.1$, $p<.001$). Post-hoc analysis showed no significant difference in ear advantage for both tone types but a larger instructed-ear advantage contour tone than the level tone (see Fig.2). This greater instructed-ear advantage improvement from NF to FL or FR suggested better control abilities for contour tone. The difficulties in distracted conditions can be smaller for contour tones since more cues (i.e., both pitch height and direction) are provided than level tones [4].

For the bilingual effect, accuracy in the instructed ear was compared between the forced condition corresponding to the instructed ear (FL or FR) and the baseline condition (NF) in two mixed-effects models. For the right-ear accuracy, the interaction of condition (FR vs. NF) and MLD was significant ($z = 8.2$, $p<.001$), indicating that participants with higher MLD (i.e., more balanced usage and higher proficiency) greatly improved the REA in the FR condition for both tone types (see Fig.3). The model on left-ear accuracy did not reveal any significant main effect or interaction of MLD (see Fig.4). Two forced conditions varied in attentional demands for their congruency with the NF. The higher-level bilinguals (with higher MLD) can better manage the limited attentional resources to focus on the incongruent ear than the lower-level bilinguals (with lower MLD) in the high-demand (FR) condition. In the FL, all bilinguals performed similarly since the demand was easy to meet regardless of more or less bilingual experience [6]. The findings suggested the bilingual advantage in attentional control [1] but further specified the modulation of task demand with the consideration of gradient bilingual experience [2]. The bilingual effect needs to be further investigated for bilinguals with different dominant language pairs (e.g., two Chinese dialects).

Proceedings of the Second International Conference on Tone and Intonation, pages 37–38
November 18–20, 2023.

Figures

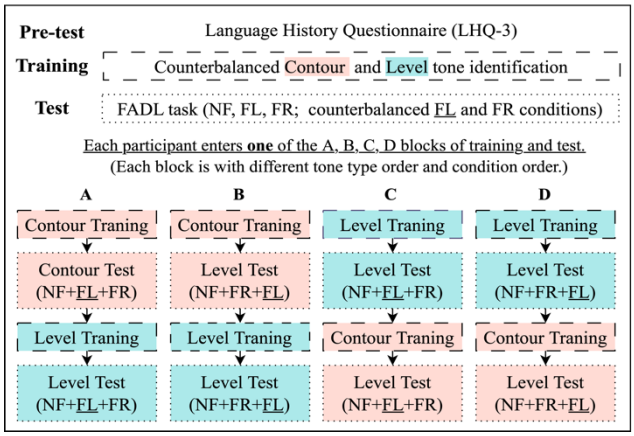


Figure 1: Overview of the procedure

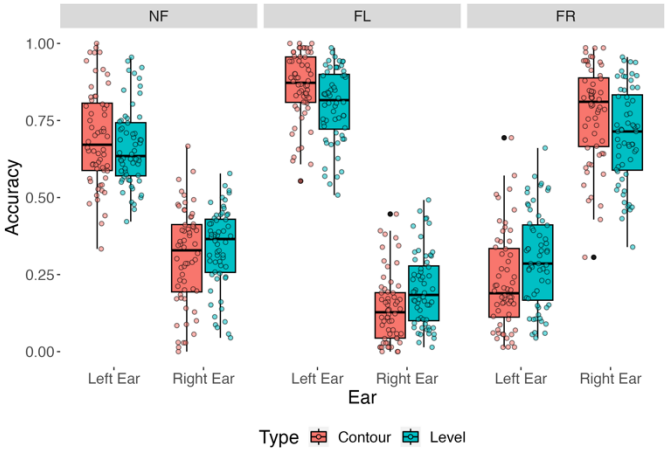


Figure 2: Left and right ear accuracy across NF, FL, FR conditions for contour (red) and level (green) tones

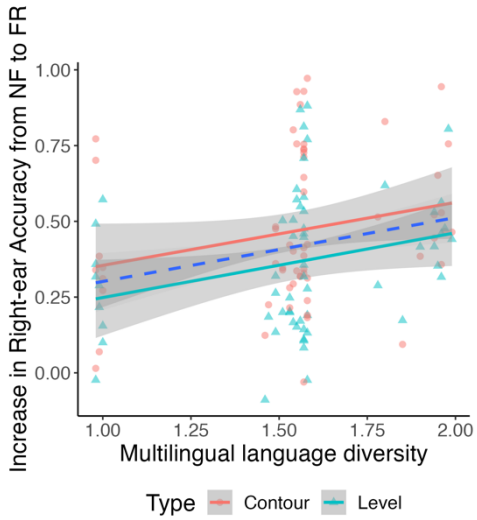


Figure 3: Right-ear accuracy improvement from NF to FR (incongruent) among bilinguals with different MLD scores in contour tone (red), level tone (green) , and regardless of type (blue dash line)

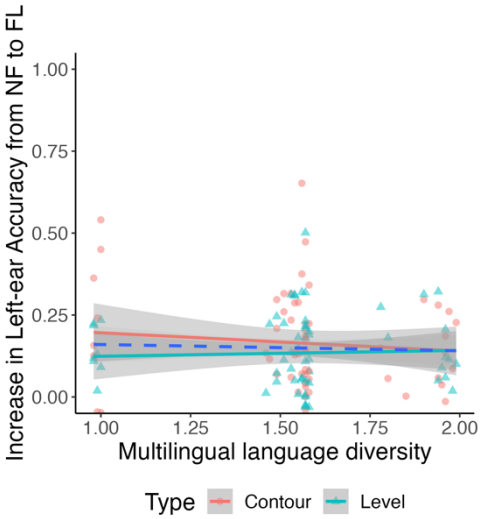


Figure 4: Left-ear accuracy improvement from NF to FL (congruent) among bilinguals with different MLD scores in contour tone (red), level tone (green) and regardless of type (blue dash line)

References

[1] Soveri, A., Laine, M., Hämäläinen, H., & Hugdahl, K. (2011). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism: Language and Cognition, 14*(3), 371-378.

[2] Privitera, A. J., & Weekes, B. (2022). Scoping review of research practices in the investigation of bilingual effects on inhibition and attention in young people. *International Journal of Bilingualism*, 13670069221121498.

[3] Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences, 103*(51), 19558-19563.

[4] Jia, S., Tsang, Y. K., Huang, J., & Chen, H. C. (2013). Right hemisphere advantage in processing Cantonese level and contour tones: evidence from dichotic listening. *Neuroscience letters, 556*, 135-139.

[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), 938-944.

[6] QU, L., LOW, J., ZHANG, T., LI, H., & ZELAZO, P. (2016). Bilingual advantage in executive control when task demands are considered. *Bilingualism: Language and Cognition, 19*(2), 277-293.

Vocative Intonation in Bulgarian

Bistra Andreeva[1] & Snezhina Dimitrova[2]
[1]*Saarland University,* [2]*Sofia University "St. Kliment Ohridski"*
andreeva@lst.uni-saarland.de, snezhina@uni-sofia.bg

The present study investigates the intonation of Bulgarian vocative contours. The materials were ten Bulgarian names 2 to 4 syllables long with stress on the final or the penultimate syllable. The names included mostly sonorants, e.g. *Marinela*. Data were elicited from 10 female native speakers of Standard Bulgarian.

Following the methodology used for other languages [1, 2], calling contours were elicited using four different contexts within a larger Discourse Completion Task: (1) addressing a person and asking a question (*Маринела*, къде отиваш? – '*Marinela*, where are you going?'), (2) scolding a person (*Маринела*, прекаляваш! – '*Marinela*, you are going too far!'), (3) threatening a person (*Маринела*, внимавай! – '*Marinela*, behave!'), and (4) calling a person's name to attract their attention (*Маринела!* – '*Marinela!*'). The speakers produced three tokens of each name in each context.

From a pragmatic and a functional point of view we found four different vocative intonation tunes roughly corresponding to the four contexts above which we distinguish as follows: (1) neutral vocative, (2) insistent vocative, (3) challenging chant, and (4) vocative chant. We analysed the four tunes following the ToBI conventions for Bulgarian which comprise an inventory of pitch accents (L*, L*+H, L+H*, H*, H+!H*), phrase accents (L-, H- and !H-) and boundary tones (%H, L%, H%, !H%).

The pitch accent for the neutral vocative tune is L+H* followed by a low phrase accent at the end of the intermediate phrase boundary (see Figure 1).

The same pitch accent L+H* (with delayed peak) is used in the insistent vocative (see Figure 2), but the nuclear syllable [ˈnɛ] and the final syllable [lɐ] of the name *Marinela* in this example are lengthened, and the name is given a separate intonation phrase ending in L-%. The lengthening which causes the peak delay distinguishes the insistent from the neutral vocative. The phonology of the Bulgarian insistent call thus shares many similarities to the urgent call described for languages like Polish [1], Romanian [3], and German [2].

In the challenging and vocative chants, the final syllable is lengthened, and has increased intensity. The tune of the challenging chant is analyzed as consisting of L* on the lexically stressed syllable [ˈnɛ] and the edge tones H-L% – a gradual rise whose peak is reached on the last syllable, followed by a fall (see Figure 3).

The tune of the vocative chant is a rising pitch movement followed by a sustained high to mid plateau (see Figure 4). Similar tunes have been attested in many languages [1], and are often characterized by additional prosodic modulations: stress shift, vowel lengthening, vowel insertion, promotion of reduced vowels to full (for an overview see [4]). In Bulgarian, the high target of the pitch accent L+H* is associated with the lexically stressed syllable [ˈnɛ] of *Marinela*. The final syllable is lengthened, has higher intensity, and an F0 change, and also its vowel does not undergo complete reduction, which contradicts the phonological vowel reduction pattern in Bulgarian [5]. These characteristics of the final syllable make it perceptually prominent. This is why we analyse it as being marked by a(n additional) pitch accent. The high target of the additionally inserted pitch accent on the final syllable is scaled (!H*). The tune ends in a mid plateau (H-%).

If the lexically stressed syllable is the last one in the proper name, then the tune of the vocative chant has to be realized on it. Therefore, the metrical structure is adjusted through the insertion of another syllable and a metrical beat on this additional syllable in the intonation phrase, which enables the realization of the communicatively relevant tune (see Figure 5).

Proceedings of the Second International Conference on Tone and Intonation, pages 39–40
November 18–20, 2023.

Details of intra- and inter-speaker variation are discussed. This research contributes to the study of Bulgarian vocative intonation, thereby situating it within the intonation system of Contemporary Standard Bulgarian. It also adds to our expanding understanding of vocative intonation gained through analyses conducted in recent years within the Autosegmental-Metrical model of intonational phonology.

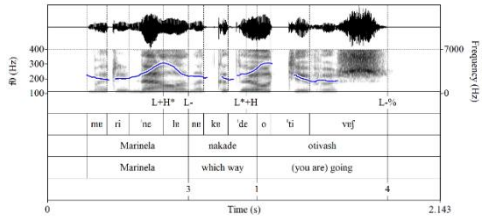


Figure 1: *Waveform, spectrogram, and F0 contour of the utterance 'Маринела, накъде отиваш?' ('Marinela, which way are you going?'); neutral vocative.*

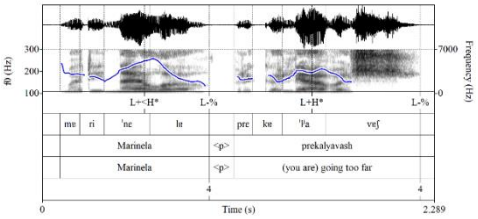


Figure 2: *Waveform, spectrogram, and F0 contour of the utterance 'Маринела, прекаляваш!' ('Marinela, you are going too far!'); insistent vocative.*

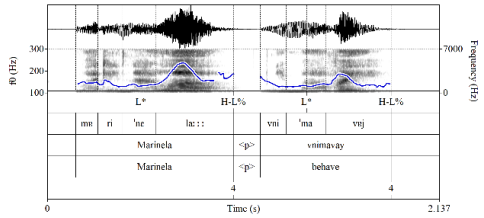


Figure 3: *Waveform, spectrogram, and F0 contour of the utterance 'Маринела, внимавай!' ('Marinela, behave!'); challenging chant.*

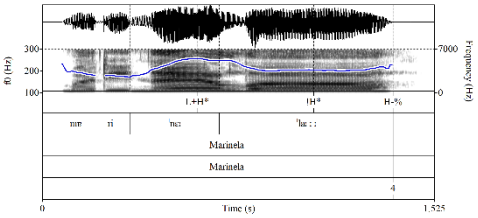


Figure 4: *Waveform, spectrogram, and F0 contour of the utterance 'Маринела!' ('Marinela!' - proper name); vocative chant.*

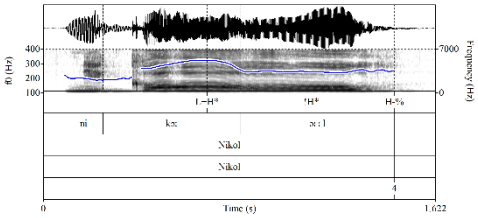


Figure 5: *Waveform, spectrogram, and F0 contour of the utterance 'Никол!' ('Nikol!' - proper name); vocative chant.*

Acknowledgements:
This research was funded by the Bulgarian National Science Fund (Project Кп-06-Н40/11) and the German Research Foundation (DFG Project 491553503).

References:

[1] Amalia, A., Żygis, M., and Jaskuëa, M. 2017. The phonetics and phonology of the Polish calling melodies. *Phonetica* 73, 338–61.
[2] Quiroz, S. I., & Żygis, M. 2017. The vocative chant and beyond. German calling melodies under routine and urgent contexts. In Lacerda, F. (ed.), *Proceedings of Interspeech* 2017, 1208–1212.
[3] Jitcă, D., Apopei, V., Păduraru, P., & Marusca, S. 2015. Transcription of Romanian intonation, In Frota, S., & Prieto, P. (eds.), *Intonational Variation in Romance*. Oxford: Oxford University Press, 284–316.
[4] Sóskuthy, M., and Roettger, T. B. 2020. When the tune shapes morphology: The origins of vocatives, *Journal of Language Evolution*, 5(2): 140–155.
[5] Ternes, E., & Vladimirova-Buhtz, T. 1999. Bulgarian. In *Handbook of the In-ternational Phonetic Association*, Cambridge: Cambridge University Press, 55–57.

Ethnicity and intonational variation in Singapore English child-directed speech

Adam J. Chong[1], Jasper H. Sim[2] & Brechtje Post[3]
[1]Queen Mary University of London, [2]National Institute of Education/Nanyang Technological University, [3]University of Cambridge
a.chong@qmul.ac.uk, jasper.sim@nie.edu.sg, bmbp2@cam.ac.uk

Singapore English (SgE) is a contact variety of English [1] situated in a complex multilingual setting. SgE intonational structure has been argued to consist of prosodic units that typically consist of a single content word and any preceding function words (Accentual Phrases: AP), with a L(ow) tone at the left edge and a H(igh) tone at the right [2]. Previous work on SgE intonation has largely concentrated on productions of ethnically Chinese speakers [3], with the intonation of non-Chinese speakers (e.g., Malay and Indian) still under-examined. Further, most work on ethnicity-related differences, and in fact most sociolinguistic variation, in SgE has largely focussed on segmental features [4,5,6]. Recent work by [7], however, found some evidence of ethnicity-related intonational differences in the speech of SgE-acquiring young children. Specifically, they found that while the global shape of f0 contours in SgE-children were similar, the relative scaling of f0 rises and falls across the utterance differed, with Malay children showing shallower rises than Chinese children, regardless of language dominance. [7] raised the possibility that these scaling differences could be at least partly explained by the influence of Malay intonational phonology [8]. These findings in children's speech raise the question of whether these differences are derived from caregiver input.

This exploratory study addresses the question to what extent apparently ethnicity-related variation in children's SgE intonation patterns can potentially be accounted for by their input. We examine intonational variation within the child-directed speech register of SgE, analysing the speech of 9 mothers from the same caregiver/child corpus in [7]: 3 English-dominant English-Chinese bilinguals (EC), 3 English-dominant English-Malay bilinguals (EM), and 3 Malay-dominant English-Malay bilinguals (MM). The dataset consisted of semi-spontaneous SVO declarative sentences (e.g. '*Mary is eating an orange'*) with stress-initial subjects and verbs that were elicited through an information gap activity between mother and child. We focus here on intonational patterns in utterance-initial and medial APs (i.e. subject and verb, including auxiliary) where tonal melodies are not affected by utterance-final boundary tones. In total, 280 sentences were analysed across 9 speakers. Time-normalized f0 measures were extracted over each syllable (10 points/syllable) using a custom Praat [9] script.

First, we observed that while Chinese mothers showed fairly uniform rises over the subject, Malay mothers, especially MM, sometimes showed late peaks where the H tone of the first AP was realized on the following auxiliary verb instead of the final syllable of the subject (Fig. 1). This is a pattern not previously observed in Chinese SgE adults [3] nor in children's productions [7]. Next, we examined the scaling of the LH rises (Fig. 2) over the first (subject) and second AP (auxiliary and verb), focusing only on cases where the rises were contained within a prototypical AP as postulated by [2] (excluding cases with late peaks). LH (rise) ratios were calculated by taking the semitone transformation of the ratio between the maximum and minimum F0 in each domain. The effects of ethnicity/language (EC vs. EM vs. MM), AP duration and syllable count, on rise ratios were tested using linear mixed-effects models. In both subject and verb APs, only duration showed a significant effect on rise ratios, with larger rises when with longer APs, echoing findings by [10]. Mothers' productions did not differ significantly based on ethnicity/language on these measures, despite numerical differences (Fig. 2).

Overall, our results firstly reveal a possible difference in tonal alignment and possibly prosodic parsing between Chinese and Malay mothers, with Malay mothers sometimes aligning the AP-final H tone of the first AP on a following auxiliary verb (vs. subject noun). Secondly, our analysis of the scaling of rises in initial and medial APs failed to reveal any effect of ethnicity/language dominance, contrary to [7]'s finding with children in the same corpus. It is possible any differences were masked due to the small sample size in the current analysis. Nevertheless, this finding points to the fact that any ethnicity-related differences in the children's speech are not likely just the result of in-task mimicking of caregiver input. Future work will examine other measures (e.g. tonal alignment) and adult-directed speech to examine whether children's patterns reflect speech community-wide norms.

Proceedings of the Second International Conference on Tone and Intonation, pages 41–42
November 18–20, 2023.

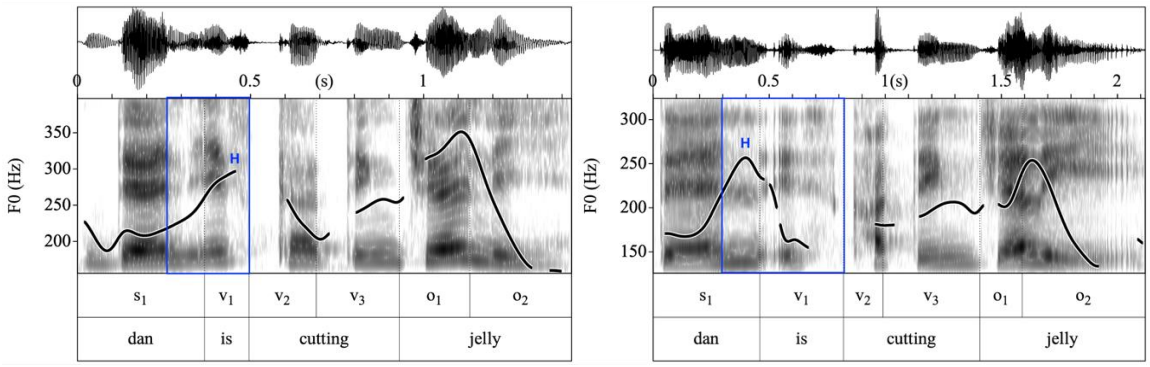


Figure 1. Pitch tracks and spectrograms of (L) a Malay speaker with a late peak - H tone - on the auxiliary and (R) a Chinese speaker with a peak at the end of the subject.

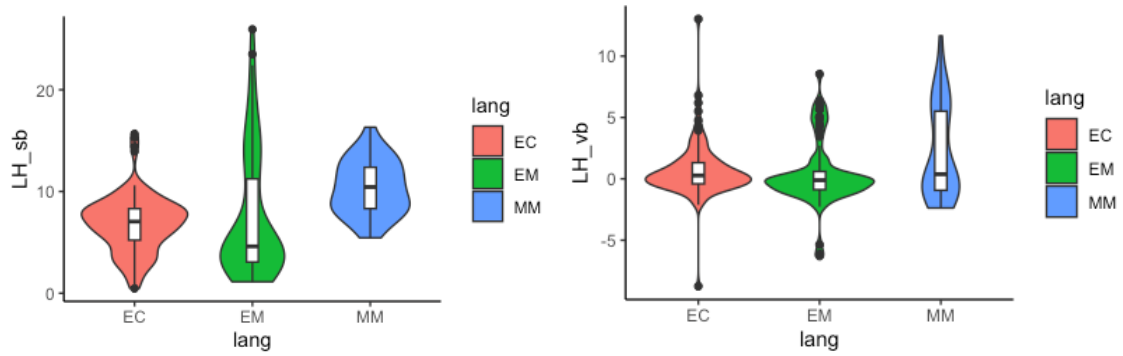


Figure 2. LH (rise) ratios (in semitones) on the (L) subject and (R) auxiliary and verb. EC = English-dominant Chinese, EM = English-dominant Malay and MM = Malay-dominant Malay speakers.

References:

[1] Grice, M., German, J.S., & Warren, P. 2020. Intonation systems across varieties of English. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language prosody* (pp. 285–302). Oxford: Oxford University Press.

[2] Chong, A.J. 2013. Towards a model of Singaporean English intonational phonology. *Proceedings of the Meeting on Acoustics*, 19, 1–9.

[3] Chong, A.J., & German, J.S. 2023. Prominence and intonation in Singapore English. *Journal of Phonetics*, 98, 101240.

[4] Leimgruber, J.R. 2013. *Singapore English: Structure, variation, and usage*. Cambridge: Cambridge University Press.

[5] Sim, J.H. 2019. "But you don't sound Malay!" Language dominance and variation in the accents of English-Malay bilinguals in Singapore. *English World-Wide*, 40, 79–108.

[6] Starr, R.L. & Balasubramaniam, B. 2019. Variation and change in English /r/ among Tamil Indian Singaporeans. *World Englishes*, 38, 630 – 643.

[7] Sim, J.H. & Post, B. 2021. Variation in pitch scaling in English of young simultaneous bilinguals in Singapore. *Poster presented at PAPE 2021*, Barcelona.

[8] Hamzah, D. and German, J.S. 2014. Intonational phonology and prosodic hierarchy in Malay. *Proceedings of Interspeech 2014*.

[9] Boersma, P., & Weenink, D. 2015. *Praat: doing phonetics by computer*. Computer program.

[10] Chong, A.J., & German, J.S. 2019. Variation in tonal realization in Singapore English intonation. *Proceedings of the 19th ICPhS*.

Polysyllabic tone sandhi and morphosyntax in Xiangshan Wu Chinese

Yibing Shi, Francis Nolan, Brechtje Post
University of Cambridge
ys538@cam.ac.uk, fjn1@cam.ac.uk, bmbp2@cam.ac.uk

Northern Wu Chinese dialects, which arguably provide the most interesting tone sandhi patterns across Chinese languages, show two contrasting sandhi processes that frequently correlate with different morphosyntactic structures (e.g., Shanghai [1], see a typology in [2]): (a) Left-dominant: rightward tone extension of the initial tone that usually applies to Lexical Compounds (LC), Modifier-Head phrases (MH), etc.; (b) Right-dominant: neutralisation of non-final tones that is usually found in structures such as Verb-Object phrases (VO). This study aims to give a detailed acoustic analysis of tone sandhi patterns regarding its interaction with morphosyntax in Xiangshan dialect, an under-studied Northern Wu Chinese dialect spoken in Xiangshan County, Zhejiang Province in China.

Xiangshan dialect has 6 lexical tones: 4 non-checked tones (HH, HL, LHL, LH) and 2 checked tones (Hq and LHq). The current analysis looks at non-checked tone combinations with an initial LHL underlying tone in disyllabic and trisyllabic contexts. The three morphosyntactic structures mentioned above (i.e., LC, MH, and VO) were selected for the disyllables. The trisyllables were left-branching Modifier-Head phrases with two different internal morphosyntactic structures: (a) [[Verb-Noun]-Noun], e.g., [[sell-flower]-person] *'a person who sell flowers'*; (b) [[Adjective-Noun]-Noun], e.g., [[yellow-melon]-soup] *'cucumber soup'*. A total of 30 frequently encountered tokens were selected, including 22 disyllables (6 LC, 10 MH, 6 VO) and 8 trisyllables (4 [[Vb-N]-N], 4 [Adj-N]-N]). There were 8 Xiangshan native speakers (mean age: 50, 4 female) who participated in the study, resulting in a total of 240 tokens. F0 values at 10 equidistant measurement points in each rhyme were obtained and normalised via z-scores of log-transformed raw data for each speaker. The F0 value at the first time point in each rhyme was discarded to eliminate possible F0 perturbations from onset consonants.

The findings of the study validate the presence of both left- and right-dominant tone sandhi patterns in disyllables in Xiangshan dialect. Specifically, disyllabic LC and MH exhibit a left-dominant sandhi pattern for all but the LHL-HL underlying tone combinations, as they converge to similar sandhi patterns regardless of the second underlying tone (i.e., sandhi1 and sandhi2 for LHL-HH/LHL/LH combinations in Figure 1). This suggests that the leftmost tone determines the whole contour for the entire disyllable. On the contrary, the VO structure in general demonstrates a right-dominant sandhi pattern, preserving the second tone whilst neutralising the first tone to either a L or a LM tone (see Figure 2), with the only exception sandhi1 for the LHL-HH underlying tone combination.

The trisyllables also show an asymmetry in the sandhi patterns correlated with morphosyntax. The sandhi of the [[Adjective-Noun]-Noun] structure mirror those of its disyllabic counterpart, both of which exhibit an overall rise-fall (sandhi1) or rise (sandhi2) contour across the entire sandhi domain, with the rise or rise-fall realised on the final syllable (see Figure 3). Essentially, the disyllabic domain within the trisyllable is eliminated, leading to a rightward tone extension over the whole trisyllable. The [[Verb-Noun]-Noun] structure, on the other hand, preserves the sandhi contours of the contained disyllables (see Figure 4), which could be accounted for as a cyclic application of sandhi. In this scenario, the right-dominant sandhi would initially be applied within the internal disyllabic domain of the initial two syllables, after which the tone features of the second syllable spread to the third syllable, showing a left-dominant sandhi across the boundary.

The asymmetry of left- vs. right-dominant tone sandhi has fostered abundant valuable theoretical works, such as Selkirk & Shen's edge-based approach [3], which aligns the left edge of a tonal domain (phonological word) with that of a lexical word, and Duanmu's stress-based account [4] which proposes a syntactic nonhead to be stressed and thus able to preserve its underlying tone. While the disyllabic data in Xiangshan dialect fit well in either theory, the trisyllabic patterns pose a challenge to the existing frameworks. Given that both trisyllabic structures are left-branching and surface as Modifier-Head phrases as a whole, the sandhi behaviours are predicted to be similar by previous frameworks, which, however, has not been attested here. Rather, Xiangshan seems to apply two distinct strategies to form tone sandhi domains, i.e., 'flattening' the internal syntactic structures and creating a single domain for [[Adjective-Noun]-Noun], or forming two tone sandhi domains cyclically according to morphosyntax of [Verb-Noun]-Noun].

Proceedings of the Second International Conference on Tone and Intonation, pages 43–44
November 18–20, 2023.

Figure 1: *Sandhi patterns for Lexical compounds & Modifier-Head phrases with LHL-σ2 underlying tones (UR)*

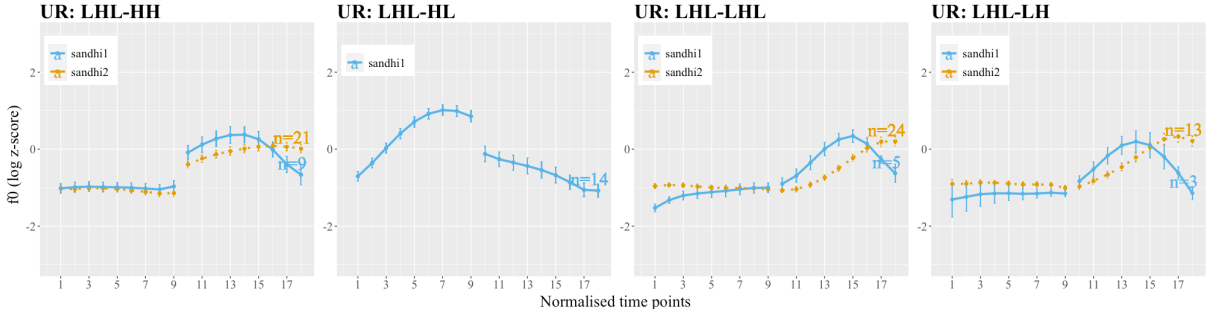


Figure 2: *Sandhi patterns for Verb-Object phrases with LHL-σ2 underlying tones (UR)*

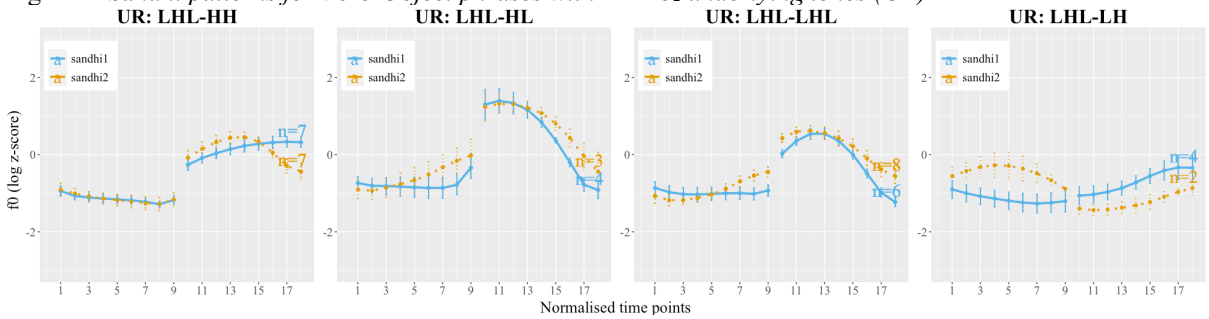


Figure 3: *Sandhi patterns for [[Adjective-Noun]-Noun] trisyllables with LHL-HH/LHL-HH/LHL underlying tones (right panel) & sandhi patterns for their contained disyllables (left panel)*

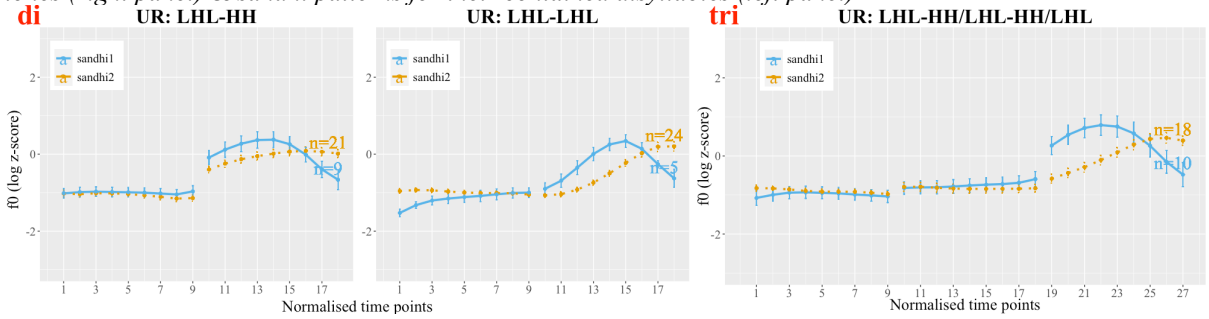
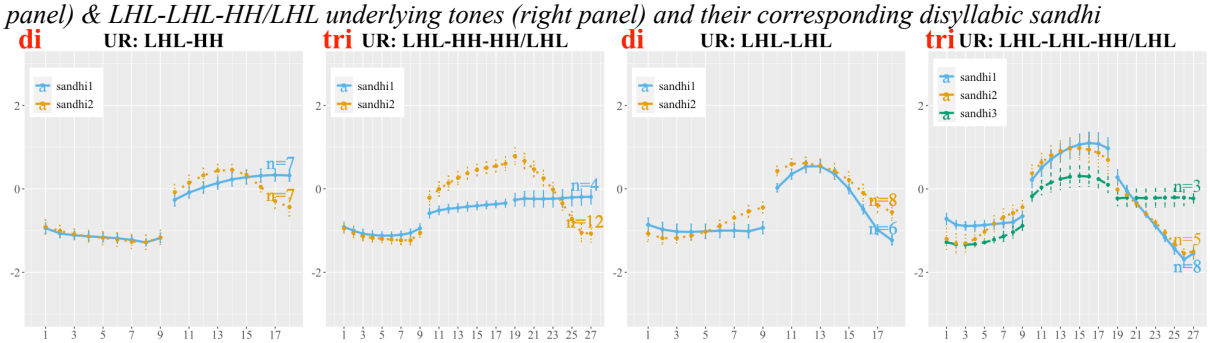


Figure 4: *Sandhi patterns for [[Verb-Noun]-Noun] trisyllables with LHL-HH-HH/LHL underlying tones (left panel) & LHL-LHL-HH/LHL underlying tones (right panel) and their corresponding disyllabic sandhi*



* Note: the sandhi outputs in the figures above, which are named as sandhi1, 2, etc., are random variants that do not seem to correlate with certain specific tokens.

References:

[1] Zhang, J., & Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. *Journal of Phonetics*, *54*, 169-201.

[2] Zhang, J. (2007). A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*, *16*, 259-302.

[3] Selkirk, E., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. *The phonology-syntax connection*, 313-337.

[4] Duanmu, S. (2005). The tone-syntax interface in Chinese: some recent controversies. In *Proceedings of the Symposium "Cross-Linguistic Studies of Tonal Phenomena*. 221-254. Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Stylised Tone and Intonation in Dog Directed Speech

Yu-Xian (Claire) Huang
University of Oxford
Yu-xian.huang@some.ox.ac.uk

Dogs were domesticated thousands of years ago (Vilá et al., 1997) and have played an important role in humans' life so much so that many dog lovers do not perceive themselves as dog owners but mothers and fathers and consider their dogs to be family members (Greenebaum, 2004). In order to encompass the significant influence of human-animal interactions in our community and cultures, Levinson (1982) suggested that we need more research on human animal communication and adopt a scientific approach.

In recent years, there has been growing interest in the special speech register human adults use when talking to dogs: dog-directed speech (DDS). DDS exhibits various linguistic features resembling infant- or child-directed speech (IDS or CDS), including short utterances, few declarative sentences, a preponderance of imperatives and questions, high-pitched voice, and wider pitch range (e.g. Hirsh-Pasek & Treiman,1982). Despite these similarities, DDS diverges from CDS in several ways, e.g. DDS is said to lack the hyperarticulation of vowels observed in CDS (Hirsh-Pasek & Treiman,1982; Burnham et al., 2002). While IDS-CDS have been extensively explored, research on DDS remains limited and there is scant documentation on tone and intonation in DDS cross-linguistically and cross-culturally.

To enhance our understanding of how sociophonetic, pragmatic, and typological language internal factors shape different interactional contexts, this study investigates DDS in American and British English and Taiwan Mandarin with regard to tone and intonation. By exploring how speakers from different cultures modulate their sociophonetic variables while interacting with dogs and whether the variables may be further constrained by typological factors such as tone, we learn about the influence of different recipients of speech and various scenarios in different timing and interactional contexts.

We recruited 9 American, 8 British, and 7 Taiwan Mandarin females, who were naïve to the purpose of the study, to interact with a college welfare dog for 10 minutes in the presence of the experimenter and 10 minutes without the researcher in the same room. At the end of the DDS session, speakers stayed in the room and spoke with the experimenter to elicit adult-directed speech (ADS). Participants' interactions with the dog and the researcher were audio and video recorded, and transcribed and annotated in ELAN, and subsequent auditory and acoustic analyses were conducted.

Regarding the analysis of English DDS, while intonational contours primarily followed those in ADS, DDS intonational contours exhibited different distributions, showing a high pitch accent density with 58.7% of the words carrying pitch accent, and greater pitch excursions. Furthermore, interesting stylised patterns were observed in the contour types of interjections, greetings, and questions. In ADS, interjections typically exhibit an exclamatory fall; however, several cases of interjections with a rising tune were found in DDS. Similarly, some greetings in DDS deviated from normal ADS greetings. Analysis revealed that in certain 'hello' tokens, both syllables were equally prominent, carrying two pitch accents within a single word, aligning with previous research by Zahner et al. (2015), who observed similar instances of multiple pitch accents within a word in German IDS. DDS exhibited versatile prosodic characteristics, manipulating features that were contingent upon linguistic characteristics, adaptive to the encounter's nature and dynamics, and sensitive to cultural factors and participants' familiarity with dogs.

In the tonal analysis of Taiwan Mandarin, participants were asked to play with the dog using toys that elicited fully voiced syllables like 'fish' [yǔ]. Pitch levels were measured and converted to pitch height to normalise intrinsic pitch differences among different speakers. Initial analyses indicate that tones in Taiwan Mandarin DDS exhibit significantly higher F0 mean, minimum, and maximum (Fig. 1), which is in line with CDS studies. Initial analysis of Mandarin DDS intonation will also be discussed alongside these findings on tone, and compared with findings for English DDS intonation.

By shedding light on the interplay between speech patterns and cultural influences in DDS in English and Taiwanese Mandarin, this study contributes to a deeper understanding of convergence and divergence in human-animal communication cross-culturally and cross-linguistically.

Proceedings of the Second International Conference on Tone and Intonation, pages 45–46
November 18–20, 2023.

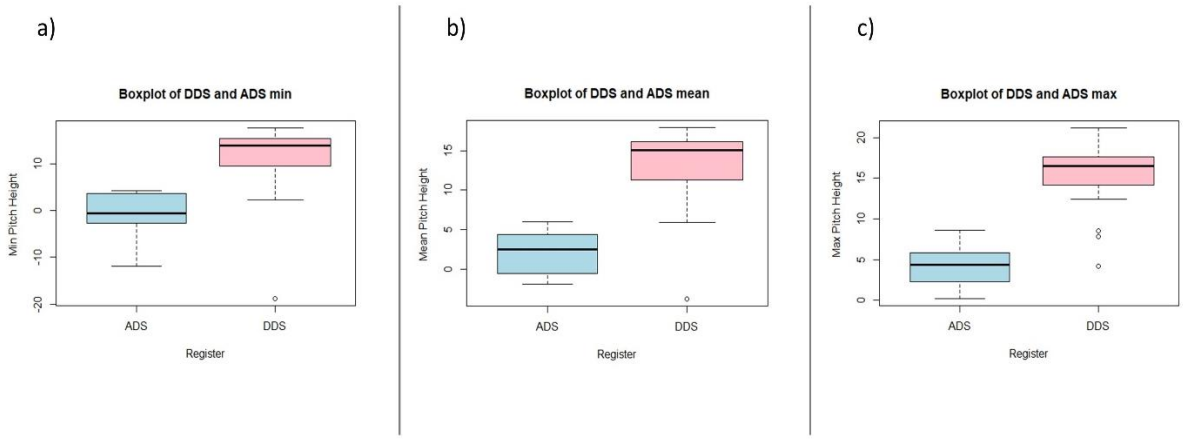


Figure 1: *Tones in Taiwan Mandarin DDS compared with ADS.*
a) Minimum, b) Mean, c) Maximum pitch height in semitones

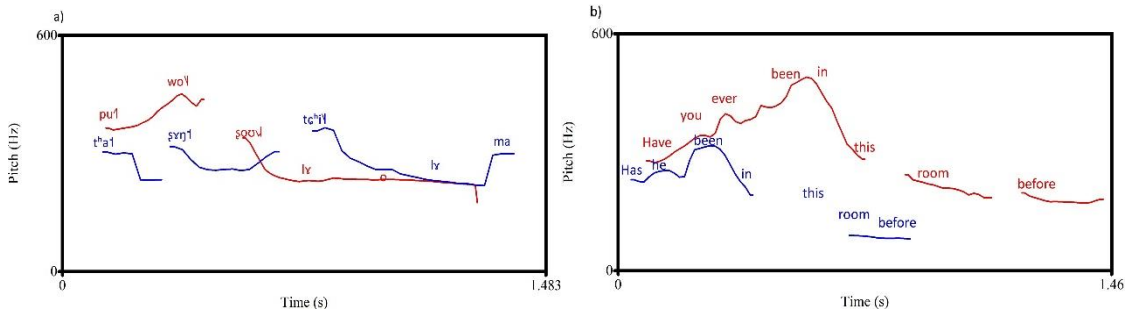


Figure 2: example f0 contours of DDS *(red line) and ADS (blue line) questions ('Is he angry?' vs 'No handshake anymore?') for a) Taiwanese Mandarin and b) British English. Notice the wider range of pitch excursions in DDS in both languages*

References

[1] Vilà, Carles, Savolainen, Peter, Maldonado, Jesús E., Amorim, Isabel R., Rice, John E., Honeycutt, Rodney L., . . . Wayne, Robert K. (1997). Multiple and ancient origins of the domestic dog. *Science (American Association for the Advancement of Science), 276*(5319), 1687-1689.

[2] Greenebaum, Jessica. (2004). It's a Dog's Life: Elevating Status from Pet to "Fur Baby" at Yappy Hour. *Society & Animals, 12*(2), 117-135.

[3] Levinson, B.M., 1982. The future of research into relationships between people and their animal companions. Int. J. Study Anim. Probl. 3 (4), 283–294.

[4] Hirsh-Pasek, K., & Treiman, R. (1982). Doggerel: motherese in a new context. *Journal of Child Language, 9*, 229– 237.Irons, S., & Alexander. (2016). Vocal fry in realistic speech: Acoustic characteristics and perceptions of vocal fry in spontaneously produced and read speech. Journal of the Acoustical Society of America, 140(4), 3397.

[5] Burnham, Denis, Kitamura, Christine, & Vollmer-Conna, Uté. (2002). What's New, Pussycat? On Talking to Babies and Animals. *Science (American Association for the Advancement of Science), 296*(5572), 1435.

[6] Zahner-Ritter, Katharina & Schönhuber (née Pohl), Muna & Braun, Bettina. (2015). Pitch accent distribution in German infant-directed speech. 10.21437/Interspeech.2015-10.

Tonal coarticulation in Angami level tones

Viyazonuo Terhiija, Zhonei i Gwirie & Priyankoo Sarmah
Indian Institute of Technology Guwahati
{viyazonuo,zhonei_i,priyankoo}@iitg.ac.in

Angami (Tenyidie) is a Tibeto-Burman language spoken in Nagaland, North-East of India with only level tones in its tonal inventory. In this study, the contextual effects of tones in disyllables are investigated. Angami native speakers living in the Kohima district of Nagaland provided speech data intended to capture tonal coarticulation. A total of 24 speakers consisting of 15 females and 9 males were recorded for this study. The mean speakers' age was 33.1 years (SD = 3.3) at the time of recording. For this study, we annotated the Angami tones with five-level categories as suggested in the MKS *Dieda, a* standard dictionary for Angami. There are 20 distinct meaningful disyllabic words with the syllable structure CVCV and one case of VCCV. The target disyllables were embedded in three environments: sentential, phrase and isolation, resulting in 2824 disyllabic tokens for the current study.

Cross-linguistically, numerous tonal languages exhibit the tone carryover effect and anticipatory effect. The F0 effects, however, differ depending on the language (Gandour,1994; Xu,1994). According to evidence from contour tonal languages such as Mandarin and Cantonese, the effects are bidirectional, with anticipatory effects dissimilating while carryover effects assimilate (Wong, 2006). Cantonese tones get higher when followed by a low onset tone showing the dissimilatory nature of anticipatory effects. In contrast, onset is higher when preceded by a high offset showing assimilatory carryover effects. Similar study is also attested in Mizo, a Tibeto-Burman language and Thai, where the effects were bidirectional, however unlike Mandarin and Cantonese, the anticipatory effect assimilates or dissimilates depending on the following tone (Sarmah, 2015; Gandour, 1994). In Mizo, anticipatory effects are more significant on contour tones where falling tones are lowered when followed by high, rising or low tones, which assimilate or dissimilate depending on the following. Mizo carryover effects are assimilatory where high or low onset (falling or high) is lowered by preceding low offset (low or falling). While there is a considerable number of studies on the tonal coarticulation or contextual tonal characteristics in contour tone languages, the contextual effects of tones in level tone languages still need to be studied. We could locate only the abstract of the study that reported contextual tones variations in Hausa, Bole and Yoruba (Yu, 2009). Hence, this study is essential as it studies tonal coarticulation in level tones.

To obtain tonal characteristics, we extracted F0 values from the vowel of the first syllable (V_1) and then continuing onto the onset consonant of the second syllable (C) and finally ending at the termination of the vowel in the second syllable (V_2). F0 values were extracted at every 2% of the total duration of the V_1CV_2, resulting in 51 points across the total duration where F0 were extracted. The extracted F0 values were used to visually represent the tonal contours and also to conduct exploratory statistical tests using Linear mixed effects (LME) modelling.

The overall F0 of the tones in Angami suggests there is an overlapping in tones T2 and T3, which are statistically insignificant. Hence, it confirms that Angami has four level tones, as opposed to the five tones proposed in the MKS dictionary. This finding is similar to the recent studies on tones in Angami. The results also showed that, in terms of anticipatory effects, only the high tone (T1) and mid-tone (T3) in the first syllable have dissimilatory effects from the second syllable (see Figure 1). However, no other tone in the first syllable showed any systematic effect of the following tones. Regarding carryover effects, the effects are both assimilatory and dissimilatory in nature. The results show that only the initial 25% of the tone contour of the second syllable is affected by the tone in the first syllable. However, both the anticipatory and carryover effects depend on the height of the tone as higher tones (T1, T2, T3) dissimilate while lower tones (T5) assimilate regardless of anticipatory or carryover effect. Statistical analyses conducted in the three middle points of the tone contours confirm the findings. The results also demonstrated that the tonal coarticulation patterns in level tone languages slightly differ from the often cited contour tone languages.

Proceedings of the Second International Conference on Tone and Intonation, pages 47–48

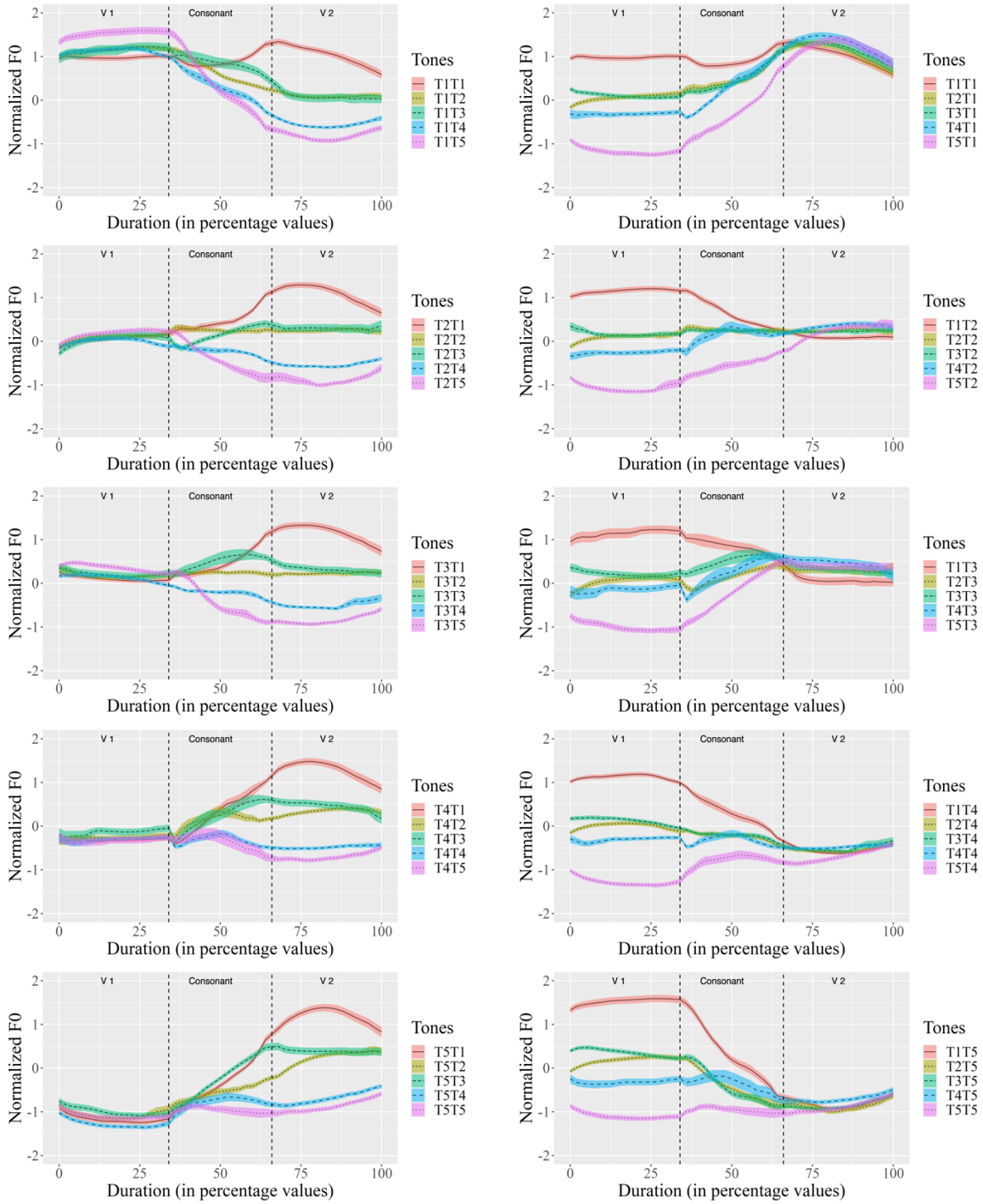


Figure 1: Normalized mean F0 in disyllables showing carryover and anticipatory effects. T1 represents the highest tone; T2, T3 & T4 are the intermediary tones while T5 is the lowest.

References

[1] J. Gandour, S. Potisuk, and S. Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics*, vol. 22, no. 4, 477–492.

[2] Y. Xu. 1994. Asymmetry in contextual tonal variation in Mandarin. *Advances in the study of Chinese language processing*, vol. 1, 383–396.

[3] Y. W. Wong. 2006. Contextual tonal variations and pitch targets in Cantonese, in *Proceedings of speech prosody*. Citeseer, 317–320.

[4] P. Sarmah, L. Dihingia, and W. Lalhminghlui.2015. Contextual variation of tones in Mizo, in *Sixteenth Annual Conference of the International Speech Communication Association*, 983–986.

[5] K. M. Yu. 2009. Contextual tonal variation in level tone languages. *The Journal of the Acoustical Society of America*, vol. 125, no. 4.

Speakers adaptively plan f0 trajectories under rate changes: Evidence from Thai contour tones

Francesco Burroni & James Kirby
Institute of Phonetics and Speech Processing, Ludwig Maximilian University of Munich
{francesco.burroni, jkirby}@phonetik.uni-muenchen.de

An open question in phonetics is how speakers of tone languages adjust the production of f0 contours when faced with changes in speech rate and word duration. This is particularly relevant for Thai contour tones, where previous research has yielded conflicting results. Some studies have suggested that Thai speakers "scale" contour tones to fit the duration of the syllables they are associated with [1]. Other studies have found that speakers truncate contour tones in connected speech, specifically the Falling (F) and Rising (R) tones, producing them without a complete fall or rise [2], [3]. Finally, another possibility is that contours are not truncated, instead they "spill over" onto the consonantal onset of following words, giving the impression of truncation when f0 tracking is limited to rhymes [4], [5]. We provide evidence and analyses supporting this last hypothesis. Using Generalized Additive Mixed Models (GAMMs), we examine entire f0 trajectories and demonstrate that Thai speakers adaptively plan and execute contour tones to accommodate their characteristic f0 shapes within the constraints imposed by speech rate changes, with interesting interactions between certain tonal combinations and speech rates.

Methodology. Two production experiments were conducted with 44 Bangkok Thai speakers. They produced disyllabic nonce combinations of syllables with sonorant onsets bearing **F**alling/**R**ising tones (20 speakers) or **M**id/**L**ow/**H**igh tones (24 speakers) in carrier sentences imitating the speed of a continuous rate cue. The disyllabic combinations were embedded in a carrier phrase consisting of Mid-toned syllables (i.e., M1 [M/L/F/H/R] [M/L/F/H/R] M2 M3) to minimize coarticulatory effects. Our focus is on the f0 trajectories of the three contour tones F/H/R, which serve as ideal targets to observe the impact of rate changes. We extracted f0 (following the methods of [5]) from the beginning of the preceding Mid-tone (M1) to the end of the consonant onset of the word following the target disyllable (M2). We analyzed the f0 trajectories of each tone separately. The GAMM models incorporated parametric terms for tonal context and duration, as well as smooths for time, duration, tonal context, and their interactions. Subject-specific factor smooths were also included. We chose GAMMs to generate predictions for each contour tone in different tonal contexts based on z-scored duration.

Results. Two key results emerge from the GAMM fits. First, Thai speakers overwhelmingly do not truncate contour tones. Second, reduction of the final portion of contour tones only emerges in specific combinations of durations and tonal context. Figure 1 shows f0 contours for the F/H/R tones at different illustrative durations and in different tonal contexts. Note that the results are confirmed in full model predictions of f0 contours over duration treated as a continuous variable. Starting from the F tone (top row), the final falling portion is never truncated at any duration/ in any context, except for the shorter duration in the F-F context. For the H tone (mid row), no truncation of the final falling portion is observed, except for the H-F context at normal and longer durations. Finally, for the R tone (bottom row), truncation of the final rise is only observed for the R-H and R-R contexts at shorter durations

Our analysis reveals two key findings. First, Thai speakers do not canonically employ truncation as a strategy to adapt to variations in speech rate. Instead, they demonstrate adaptive planning and execution of f0 contours, ensuring that the rate of f0 change corresponds to the duration of the associated segmental material, preserving characteristic f0 shapes. This entails the need for f0 tracking over larger windows beyond the immediate word. Second, specific combinations of tonal contexts and duration induce changes in contour tone shape that resemble truncation. These contexts share shorter durations and conflicting f0 direction, e.g., F-F sequences, where the final fall of the first tone conflicts with the initial rise of the second tone. Our findings have implications for models of f0 control. The dynamic scaling of f0 contours suggests a relationship between rate of f0 change and the units lexically associated with the tone, a relationship not incorporated in mainstream models of f0 control. The occurrence of truncation-like effects at faster speeds, with conflicting f0 movements, suggests that these effects may arise from competing demands on laryngeal articulation that cannot be met due to insufficient timing resulting in asymmetric blendings. We present preliminary modeling of these effects using an f0 control model incorporating aspects of the Fujisaki and Task-Dynamic model ([6]–[8]).

Proceedings of the Second International Conference on Tone and Intonation, pages 49–50
November 18–20, 2023.

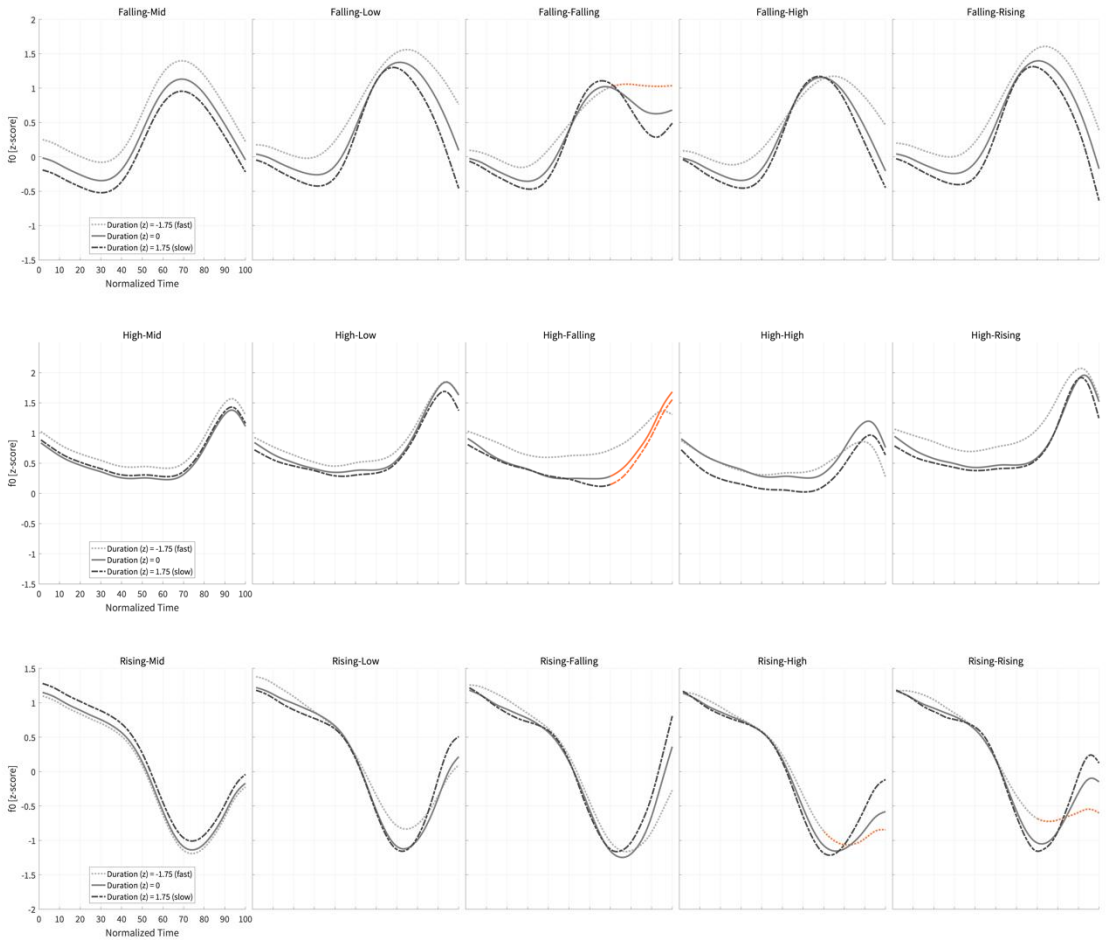


Figure 1: *F0 contours from GAMM model fits representing the mean realization of the Falling (top panel), High (mid panels) and Rising tone (bottom panel) followed by different tonal contexts and at different z-scored durations. Orange lines indicate truncation-like effects.*

References

[1] R. Roengpitya, "The variations, quantification, and generalizations of standard Thai tones," in *Experimental approaches to phonology*, Oxford University Press, 2007, pp. 270–301.

[2] B. Morén and E. Zsiga, "The lexical and post-lexical phonology of Thai tones," *Natural Language & Linguistic Theory*, vol. 24, no. 1, pp. 113–178, 2006.

[3] E. Zsiga and R. Nitisaroj, "Tone Features, Tone Perception, and Peak Alignment in Thai," *Language and Speech*, vol. 50, no. 3, pp. 343–383, 2007.

[4] P. Rose, "Mr. White goes to Market-Running Speech and Citation Tones in a Southern Thai Bidialectal," in *Proceedings of the 15th Australasian International Conference on Speech Science and Technology*, Christchurch, 2014.

[5] F. Burroni, "Dynamics of F0 Planning and Production: Contextual and Rate Effects on Thai Tone Gestures," Ph.D. Thesis, Cornell University, Ithaca, 2023.

[6] E. Saltzman and K. Munhall, "A dynamical approach to gestural patterning in speech production," *Ecological psychology*, vol. 1, no. 4, pp. 333–382, 1989.

[7] M. Gao, "Mandarin tones: An articulatory phonology account," Ph.D. Thesis, Yale University, 2008.

[8] H. Fujisaki, "In search of models in speech communication research," in *Ninth Annual Conference of the International Speech Communication Association*, 2008.

Challenging categorical perception of lexical tone and gradient perception of intonation
Evidence from Cantonese identification and discrimination studies

Yang Yang[1], Carlos Gussenhoven[2], Victoria Reshetnikova[3], Marco van de Ven[2]
[1]*Guangdong University of Foreign Studies,* [2]*Radboud University,* [3]*Utrecht University*
yangyanggw@gdufs.edu.cn, carlos.gussenhoven@ru.nl, v.reshetnikova@uu.nl, marco.vandeven@ru.nl

There appears to be a consensus that speakers of tone languages perceive tones categorically ([1], among others), but the picture is less clear for intonation. Recently, [2] conducted identification and discrimination experiments on Zhumadian Mandarin (ZM) using monosyllabic stimuli from 7-step acoustic continua between two tone contrasts (early and late falls, early and late rises, each contrast implemented on a declarative intonation and an interrogative intonation) and between intonation contrasts (statements and questions, implemented on each of the four tones). Results for native speakers and Indonesian controls showed that despite the subtle phonetic differences, the lexical tone contrasts in ZM in both pairs of tones are perceived categorically by the native speakers only, while the intonation contrast is gradient for the native group as well as the control group, for all four tones. This result showed an unambiguous dichotomy between phonological categories for tones and gradient variation for intonation in ZM. Because the tones in that study differed in shape and the intonations differed in pitch register, the question is whether these results can be replicated in tone languages that have a register difference between lexical tones and a pitch shape difference between intonations, so that the effects of the phonetic properties of the stimuli can be disentangled from the effects of their functional properties.

An Identification and Discrimination experiment on Cantonese tone and intonation

Accordingly, we conducted identification and discrimination studies like those in [2] on Cantonese, which language offers the crucial test case: it has a level tone contrast of T1 (high-level, 55) and T3 (mid-level, 33) as well as a contour tone contrast of T4 (low-falling, 21) and T5 (low-rising, 23), while using a final rise in questioned forms. Using the STRAIGHT morphing procedure [3], we created two declarative (S) and two interrogative (Q) continua for each lexical tone contrast (T1 to T3 and T4 to T5) as well as four S-to-Q continua, one for each lexical tone, from recordings by a female native speaker the three minimal quadruplets in Table 1. All eight continua consisted of seven stimuli, as depicted along the sides of the two squares in Figure 1. This procedure yielded 168 natural-sounding stimuli (7 (steps) × 4 (continua) × 3 (segmental syllables) × 2 (tone contrasts).

We presented the identification experiment and the discrimination experiment to 40 participants recruited from Guangdong University of Foreign Studies in Guangzhou as multiple-forced choice tasks in Praat. The experiments were self-paced during 30-minute slots, whereby the identification task preceded the discrimination task. We applied polynomial logistic regression analyses with the logit link function by means of the lme4 package [5] in R [6] for both the discrimination data and the identification data. We included the dependent variable *response* (T1/T3, T4/T5 for tone identification, D/Q for intonation identification, *Same/Different* for discrimination) and the independent variables syllable (*wai*, *yan*, *yau*), *step* (1 to 7), *type of contour* (statement/question), *type of contrast* (tone/intonation), and *step difference* (for discrimination: the step difference between the two stimuli presented). In addition, we included a random intercept for *participant*. The dome-shaped pattern of the discrimination curves for tone (Figure 2, left panel) was found for both the T1-to-T3 pitch range contrast and the T4-to-T5 pitch shape contrast, indicating that Cantonese listeners perceive tones categorically, i.e. with more precision around the category boundary, while the slide-shaped response curves for the intonation contrast in the right panel show that listeners are more sensitive to acoustic differences on the left of the continuum, suggesting a categorical effect of the final rise, which is easily detected once it is there, but is poorly discriminated across the remainder of the continuum. Strengthening that latter interpretation, Cantonese intonation identification even yields a more pronounced S-shaped curve than do the tone contrasts, as shown in Figure 3.

In conclusion, lexical tone contrasts are categorical, regardless of whether they are cued by pitch register or pitch shape differences, while intonation contrasts are gradient when differing in pitch register and categorical when differing in pitch shape.

Proceedings of the Second International Conference on Tone and Intonation, pages 51–52
November 18–20, 2023.

Table 1. Four tonal minimal quadruplets used in the experiments.

	[jau]	[jan]	[wai]
T1 (high-level)	‘to worry’ 忧	‘marriage’ 姻	‘authority’ 威
T3 (mid-level)	‘young’ 幼	‘to print’ 印	‘to comfort’ 慰
T4 (low-falling)	‘oil’ 油	‘people’ 人	‘violate’ 违
T5 (low-rising)	‘friend’ 友	‘to lead’ 引	‘great’ 伟

Figure 1. Acoustic continua between T1 (high-level) and T3 (mid-level) and T4 (low-falling) and T5 (low-rising) in statement and question intonations and between their respective intonations.

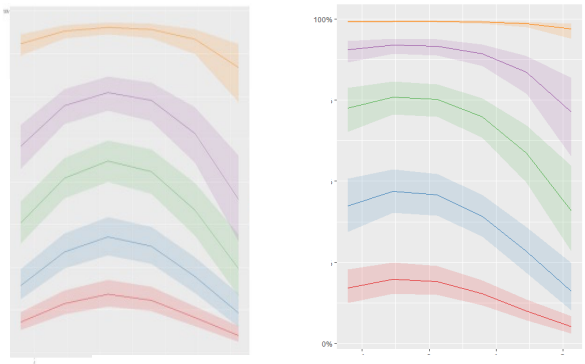


Figure 2. Predicted probabilities of ‘different’ responses for ‘tone’ (left) and ‘intonation’ (right) in the Discrimination experiment, separated by size of step difference, from 1 (bottom) to 5.

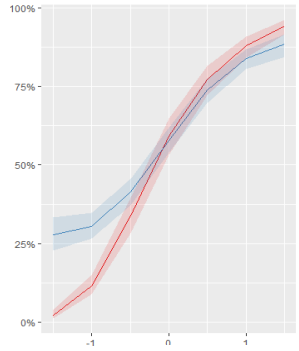


Figure 3. Predicted probabilities of the perception of tone (T1, T4; blue) and intonation (Q; red) in the Identification experiment.

References

[1] Shen, G. & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: an ERP study. *Bilingualism: Language and Cognition* 22, 253–265.
[2] Gussenhoven, C., & van de Ven, M. (2020). Categorical perception of lexical tone contrasts and gradient perception of the statement-question intonation contrast in Zhumadian Mandarin. *Language and Cognition*, 12(4), 614-648.
[3] Kawahara, H., Masuda-Katsuse, I. & Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequencybased F0 extraction: possible role of a repetitive structure in sounds. *Speech Communication* 27(3/4), 187–207.
[4] Boersma, P., & Weenink, D. (2008). *Praat: doing phonetics by computer*. Computer program.
[5] Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1).
[6] R Development Core Team, (2015). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

L2 learning and language attrition in intonation: analysis of Spanish L2 and Brazilian Portuguese L1 applying the Fujisaki model

Cristiane Silva[1], Hansjörg Mixdorff[2], Pablo Arantes[3]
[1]*Universidade Federal de Santa Catarina, Brazil,* [2]*Berliner Hochschule für Technik, Germany,* [3]*Universidade Federal de São Carlos, Brazil*
cristiane.conceicao@ufsc.br, hmixdorff@bht-berlin.de, pabloarantes@ufscar.br

Describing L2 speech prosody is a challenge, given the great variability observed in the production of bilingual speakers. The use of methods based on discrete labels poses a challenge: what tone inventory is to be used to label the production of bilinguals in the face of phenomena such as L1 to L2 transfer and language attrition (L2 to L1 influence)? [1, 2, 3, 4]. One alternative is to use quantitative methods that describe contour features that are not theory-dependent. In this study, we compare two quantitative techniques to describe the *f0* contours of wh-questions produced by Brazilian and Spanish monolingual speakers and by Brazilian bilinguals in Spanish L2 and Portuguese L1. The goal is to analyze the relationship between L2 learning and language attrition in L1. The first technique describes a given *f0* contour by means of its extreme points, ie, the local *f0* maxima and minima in the surface *f0* contour and *f0* velocity contour [5]. It is a simple procedure that does not rely on any strong underlying assumption. The second technique we tried is the Fujisaki model [6]. The goal of this study was to compare the results reported in [4] with the results obtained with the Fujisaki model to assess the advantages of a more sophisticated method that models the observed f_0 contour by superimposing global (phrase) and local (accent) components. A further advantage of this model is the possibility to resynthesize utterances with modified f_0 contour to test our hypotheses perceptually. For that, we explore the position and magnitude of phrase commands; the amplitude, duration, and position of accent commands relative to stressed syllables in final and non-final words. The previous study showed evidence of L2 learning, since most of the f_0 contours produced by Brazilian bilinguals that are Spanish-like (final rise, nuclear circumflex and double circumflex) are similar to the ones produced by Spanish monolinguals. The study also showed evidence of L1-L2 transfer, since the participants also produced f_0 contours typical of Portuguese L1 (global falling) in Spanish L2 and evidence of language attrition. The authors identified qualitative attrition, characterized by the production of Spanish-like contours in Portuguese L1 and quantitative attrition, since the bilinguals produced global falling contours in L1 with wider f_0 range than monolinguals. Figure 1 (Fujisaki model) shows an example of qualitative attrition identified in the previous study. We analyzed data from 25 speakers. Ten Brazilian and Spanish speakers (6 female and 4 male). The Spanish speakers never studied Portuguese as L2 and the Brazilians never studied Spanish as L2. Fifteen Brazilian speakers (10 female and 5 male) that also speak Spanish as L2. All started learning Spanish after the age of 18 and lived in Madrid at the time the data was collected. The results of the present study with Fujisaki model lead to a similar classification suggested by the accents commands for monolinguals and bilinguals. Furthermore, the results showed that there is a greater f_0 range in the production of bilinguals in L1 and L2 compared with the Brazilian monolinguals (Figure 2). The phrase commands of global falling contours produced by bilinguals have significantly higher mean magnitude in Portuguese L1 ($p = 0.002$) and a marginally significant higher mean value in Spanish L2 ($p = 0.072$) compared to the same contour type of Brazilian monolinguals. The analysis of the first accent command shows in Brazilian monolinguals a lower mean amplitude compared to bilinguals (both Portuguese L1 and Spanish L2 $p < 0.001$). The present findings suggest that the extended *f0* range observed in bilingual contours results from a combination of a stronger phrase accent and a stronger initial accent command aligned with the interrogative pronoun. There are also other differences that will be discussed in more detail in the presentation.

Proceedings of the Second International Conference on Tone and Intonation, pages 53–54
November 18–20, 2023.

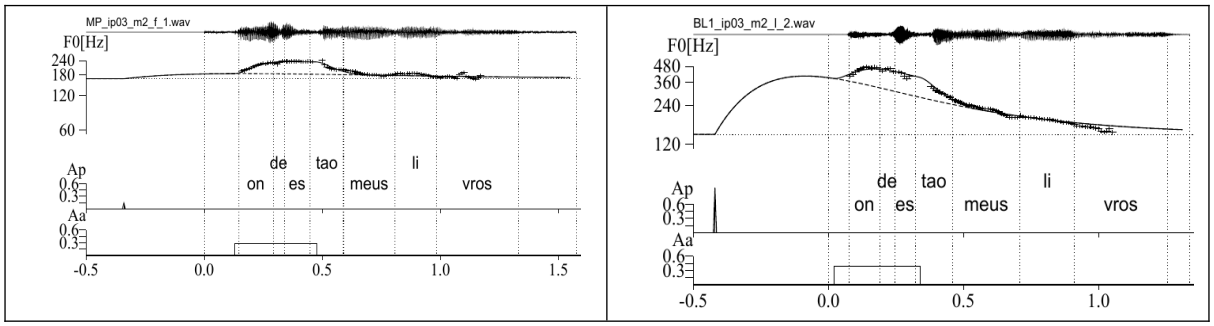


Figure 1: One *example of analysis in Portuguese L1 of the sentence "Onde estão meus livros?" Where are my books? With global falling contour of a female monolingual Brazilian speaker 2 (left) and of a female bilingual Brazilian speaker (right); Each panel displays from the top to the bottom: The speech wave form, the F0 contour (extracted +++, modelled ---) the underlying impulse-wise phrase commands and box-shaped accents commands of the Fujisaki model. Syllables boundaries are indicated by dotted vertical lines*

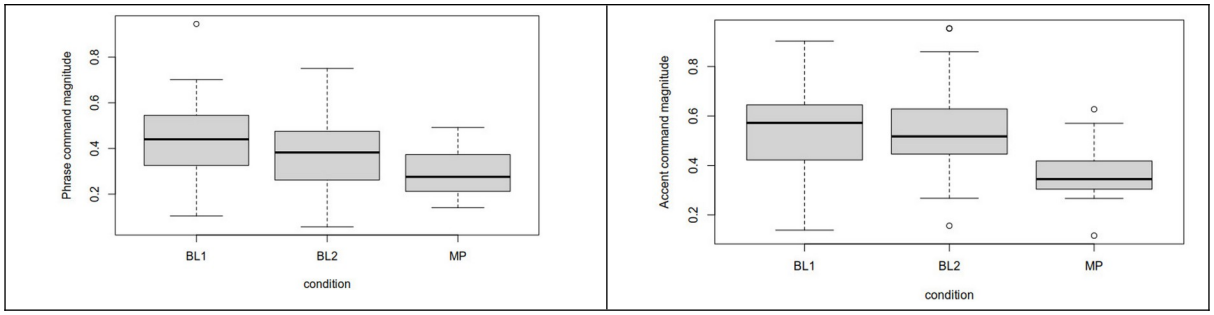


Figure 2: *Boxplots of phrase command amplitude Ap (left) and first accent command amplitude Aa (right) for MP (Portuguese L1 monolinguals), BL1 (Portuguese L1 bilinguals) and BL2 (Spanish L2 bilinguals)*

References

[1] Silva, C. & Arantes, P. 2021. Quantitative analysis of fundamental frequency in Spanish (L2) and Brazilian Portuguese (L1): evidence of learning and language attrition. *Journal of Speech Sciences* 10, 1-31.

[2] Silva, C. & Arantes, P. 2022a. A qualitative study on the variability in intonation learning and attrition in Brazilian Portuguese Bilingual speakers of Spanish L2. *Cadernos de Estudos Linguísticos* 64, 1-20.

[3] Silva, C. & Arantes, P. 2022b. Análise quantitativa da entoação de perguntas interrogativas pronominais: evidências de aprendizagem e atrito na produção de falantes de espanhol L2 brasileiros. In Silveira, A. & Arantes, P. (Ed.), Prosódia e biliguismo. Araraquara: Letraria, 257-286.

[4] Silva, C. & Arantes, P. in press. Prosody and L2 learning interface: the case of Spanish L2 and Brazilian Portuguese L1 intonation. In Oliveira Jr, M. (Ed.), Prosodic Interfaces. Berlin: Gruyter.

[5] Arantes, P. 2021. *parantes/f0_extrema*: Initial release. Zenodo.

[6] Fujisaki, H., & Hirose, K. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. Journal of the acoustical society of Japan. 5, 233-242.

Attitudinal Prosody in Hindi

Hansjörg Mixdorff[1], Archishman Ghosh[2], Prashant Khatri[3], Preeti Rao[3],
Alexsandro Meireles[4]
[1]*Berliner Hochschule für Technik, Germany* [2]*Kalinga Institute of Industrial Technology, Odisha, India,* [3]*IIT Bombay, Mumbai, India,* [4]*Federal University of Espirito Santo, Brasil*
hmixdorff@bht-berlin.de, 2004144@kiit.ac.in, prashant23@iitb.ac.in,
prao@ee.iitb.ac.in,alexsandro.meireles@ufes.br

This paper presents results from the prosodic analysis of short utterances of Hindi produced with varying attitudinal expressions. It is based on the framework developed by Rilliard et al. [1] eliciting 16 different kinds of social and/or propositional attitudes (see Table 1) which place the subjects in various social interactions with a partner of inferior, equal or superior status, respectively as well as positive, neutral or negative, valence. Although recordings also concern the visual channel, in this study we contrate on tonal features analyzed in the framework of the Fujisaki model [2] with respect to *F0*, as well as other prosodic features extracted using *PRAAT*[3].

Table 1: List of sixteen attitudes.

ADMI	admiration	OBVI	obviousness	DECL	neutral statement	SINC	sincerity
ARRO	arrogance	POLI	politeness	DOUB	doubt	SURP	surprise
AUTH	authority	QUES	neutral question	IRON	irony	UNCE	uncertainty
CONT	contempt	SEDU	seductiveness	IRRI	irritation	WOEG	walking-on-eggs

The corpus under investigation contains a total of 19 naïve presenters [4], students at IIT Roorkee, India, of which we selected the 12 most convincing subjects, 5 males and 7 females. The sound tracks of the video clips were saved as wav files at 16kHz/16bit. The focus of the current paper is the acoustic analysis of the target utterances in search for features that distinguish attitudes from one another. We examined macro-prosodic features such as fundamental frequency, speech rate and intensity, as well as voice quality features such as harmonics-to-noise ratio, jitter, and shimmer.

Attitudes such as arrogance, politeness or doubt were elicited through short dialogs which ended in the target sentences एक केला- ek kela (engl. *a banana*) or मैरी नाच रही थी - mairee naach rahee thee' (engl. *Mary was dancing*). Preceding the target dialog a test dialog was performed in order to prepare the speakers and help them immerse themselves in the context of the attitude. These dialogs were designed according to different social situations differing in social and linguistic aspects such as the type of speech act (propositional/social), hierarchical distance, social distance or valence of speech act (positive/negative).

Following the work presented in [5] we examined the prosodic features fundamental frequency, phone rate and intensity. All target utterances were force-aligned on the phone and word levels. *F0* contours were extracted at a step of 10 ms using the *PRAAT* default pitch extraction settings and subjected to manual inspection and correction. We performed Fujisaki model parameter extraction [6]. The Fujisaki model approximates natural *F0* contours by superimposing three components: A constant base frequency *Fb* (indicated by the dotted horizontal line), exponentially decaying phrase components which are the responses to the phrase commands and accent components which are the smoothed responses to the accent commands. *Fb* therefore indicates the *F0* floor of the whole utterance. *Ap* expresses the global slope of the *F0* pattern, the accent command amplitudes *Aa* reflect the magnitude of local *F0* gestures and onset and offset times of accent commands their alignment with the underlying segments. Since its components are superimposed in the log *F0* domain, the model performs a normalization of the raw *F0* contour. Figure 1 displays examples of the utterance "ek kela" uttered by female speaker S01, uttered with attitudes ADMI, DECL, DOUB and QUES. As can be seen from the various instances displayed by the same speaker, the attitude not only influences the range of *F0*, but also quite dramatically the shape of the *F0* contour.

Intensity contours were extracted in *PRAAT* with default settings, and max intensities in dB, were determined for each word. In addition to these features we used *PRAAT* to extract mean harmonics-to-noise ratio for vowels, as well as jitter in the target utterances, applying default settings. The following table shows results of statistical analysis comparing distributions of parameters as a function of the

Proceedings of the Second International Conference on Tone and Intonation, pages 55–56
November 18–20, 2023.

underlying attitude. Every cell indicates the parameters that significantly differ between two attitudes in terms of the Kruskal-Wallis Test of independent samples ($p < 0.01$).

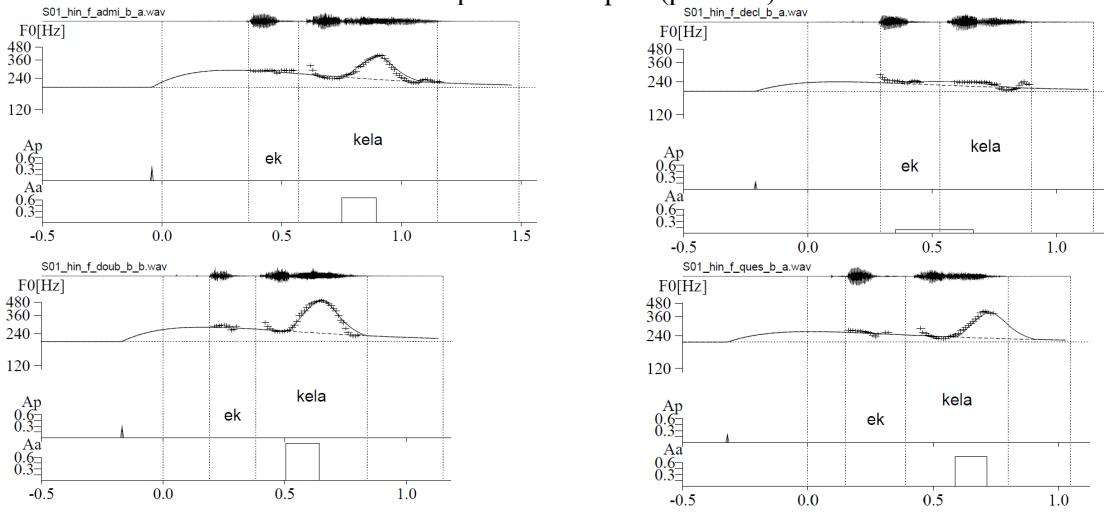


Figure 1: Results of Fujisaki model-based analysis of F0 contours, female speaker S01, Attitudes top: ADMI, DECL, bottom: DOUB, QUES. Each panel displays from the top to the bottom: The speech wave form, the *F0* contour (extracted +++, modelled ---) the underlying impulse-wise phrase commands and box-shaped accents commands of the Fujisaki model. Word boundaries are indicated by dotted vertical lines.

Due to space limits, we can only state a few observations, namely the obvious acoustic overlap between attitudes, eg similarity of ARRO, AUTH and CONT that we already found for German, and the specific properties of UNCE and WOEG being uttered more slowly and with a higher HNR than the other attitudes. We will discuss these in detail in our full paper.

	ADMI	ARRO	AUTH	CONT	DECL	DOUB	IRON	IRRI	OBVI	POLI	QUES	SEDU	SINC	SURP	UNCE
ARRO	DA														
AUTH	DH														
CONT	DAP														
DECL	DPI	AI	I	AI											
DOUB	DAP	AJ	AH	A	AI										
IRON	DP	A		A	I	A									
IRRI	I	DAPI	DPI	PI	DPI	DAPJI	PI								
OBVI	D			A	I	A		DI							
POLI	I	AI	I	A		AI	I	PI	I						
QUES	DAP	AJ	A	AJ	AI		A	DAPJI							
SEDU	API	DI	DI	I	DA	DAJI	AI	PI	DAI	A	DAJI				
SINC	PI	AI	I	A		AI	I	PI	I		A	A			
SURP	DAPI	AJI	AJI	DAJI	AI	I	AJI	DAPJ	AI	DAI	I	DAJI	DAI		
UNCE	I	DAIH	DIH	DA	DH	DAJI	DIH	IH	DH	D	DJ	A	DH	DAJIH	
WOEG	APJIH	DIH	DJIH	DIH	DAJH	DAJIH	DAPJIH	DIH	DAJIH	DAJH	DAJIH	H	DAJH	DAJIH	A

Figure 2: Acoustic features differing significantly between two attitudes. Abbreviations: D - phone duration, A - accent command amplitude Aa, P phrase command amplitude Ap, I - maximum intensity, J - local jitter, H - harmonic-to-noise ratio.

References

[1] Rilliard, A., Erickson, D., Shochi, T., de Moraes, J.A. 2013. Social face to face communication - American English attitudinal prosody. INTERSPEECH 2013. 1648-1652.

[2] Fujisaki, H., Hirose, K. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. Journal of the Acoustical Society of Japan 5, 233-241.

[3] Boersma, P., & Weenink, D. 2008. *Praat: doing phonetics by computer*. Computer program.

[4] Mixdorff, H., Nayan, N., Rilliard, A., Rao, P. and Ghosh, D. 2023. Developing a Corpus of Audio-visual Attitudinal Expressions in Hindi. Accepted for presentation at ICPhS 2023, Prague, Czech Republic.

[5] Mixdorff, H., Hönemann, A., Rilliard, A. 2015. Acoustic-prosodic Analysis of Attitudinal Expressions in German. Proceedings of Interspeech 2015, Dresden, Germany.

[6] Mixdorff. H. 2000. A new approach to the fully automatic extraction of Fujisaki model parameters. Proceedings of ICASSP 2000, vol. 3, 1281-1284, Istanbul, Turkey.

Encoding Tone Sandhi in Zhangzhou Southern Min: An Inter-disciplinary Exploration

Dr. Yishan Huang
The University of Sydney
yishan.huang@sydney.edu.au

Abstract:

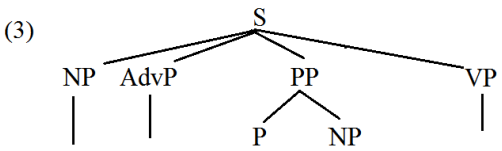
This study comprehensively explores the nature of tone sandhi in Zhangzhou, a Southern Min variety spoken in the southern part of Fujian province in southern China. This dialect presents a typical right-dominant tone sandhi system, whose sandhi domain is aligned with the boundary of syntactic phrases but is irrelevant to their internal structures. The sandhi (non-rightmost; non-phrase-final) tones are phonologically inert to the category of following tones, because, regardless of whether their subsequent tone is rising, level, or falling, they present a consistent tendency. The non-sandhi tones (rightmost; phrase-final) are highly sensitive to the phonetics of their occurring environments and present statistically significantly different variants, questioning the conventional default-principle on the specification of tone sandhi dominancy in Sinitic languages. Tonal neutralisations occur across linguistic contexts, creating difficulty in the determination of the directionality of tonal alternation, and the totality of tonal contrasts in this dialect. The nature of Zhangzhou tones is morphological. Each tone functions as a single morpheme that has two or more allomorphs (tonemes) that are independently stored in native speakers' mental grammar but are phonetically distant in real-world utterances. The relation between sandhi and citation tones are morphophonemic, while is allophonic between non-sandhi and citation tones, challenging those conventional proposals that assume a circular tonal alternation in Southern Min and posit various rule-based and/or OT-based explanations.

As indicated, investigating tone sandhi is not simply an issue to identify how tonal pitch is changed from one form to another. But rather, it turns out to be a series of sophisticated issues in relating to how tone sandhi operates as a system; how its operation may affect the sound pattern of the language being concerned; how its operation may be constrained by linguistic factors that go beyond phonology; how its inducing various tonal forms in connected speech are related to each other, as well as how its inducing various tonal forms are abstractly modelled in the mental grammar of native speakers. This study is scientifically grounded in acoustic utterances from 21 native speakers from the urban area of Zhangzhou city. The exploration substantially stretches and advances our knowledge of tone sandhi as an important language phenomenon in Southern Chinese dialects. It is hoped to serve as a model to investigate the nature of tone sandhi in languages where relevant. This study also enlightens the discussion on how human beings employ various linguistic levels (phonetics, phonology, semantics, morphology, and syntax) to encode a complex speech phenomenon as part of their cognitive activities, and how they decode the complexity in their language practices.

Keywords: tone sandhi, encoding, phonology, phonetics, syntax, morphology, Southern Min

Proceedings of the Second International Conference on Tone and Intonation, pages 57–58
November 18–20, 2023.

(3)



	ɠwɐ3	ɦjɐn5.tsɐj5	ti5	kɐw4.ɠjɔ8.ɦe5.tʰɔŋ3	kɐŋ1.tsɔk6
	1sg	now	prep.	education system	work
C:	[51]	[33][33]	[33]	[41][221][33][51]	[35][41]
S:	[66]#	[32][33]#	[32]	[63][32] [32][52]#	[33][41]#

"I am working in the education system."

Figure 1. Syntactic relevance of tone sandhi domain.

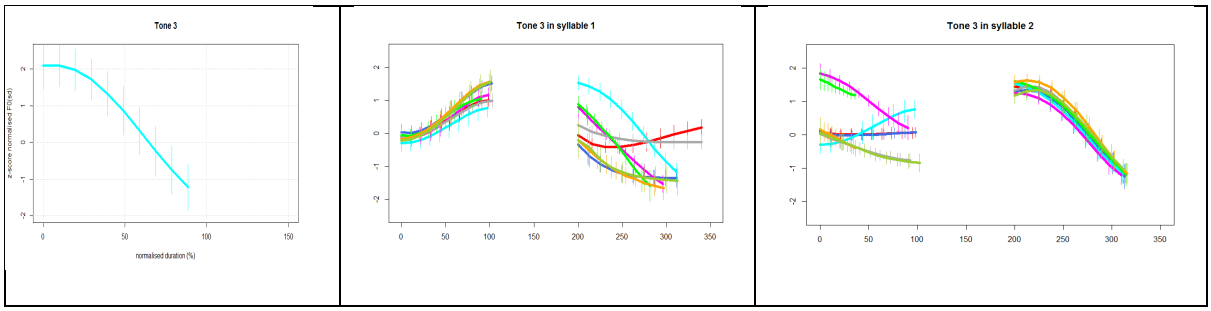


Figure 2: Right-dominant tone sandhi in Zhangzhou.

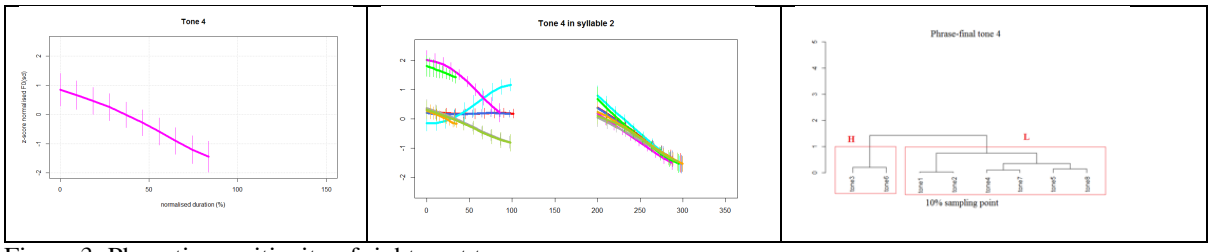


Figure 3: Phonetic sensitiveity of rightmost tones

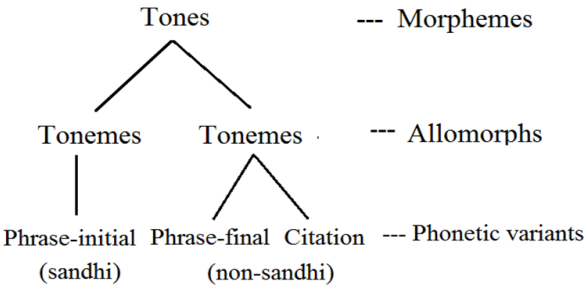


Figure 4: Morphological nature of Zhangzhou tones in general syntactic context.

F0 Change of Lexical Tones in Loud Speech

Hui Zhang, Weitong Liu, Jiajia Shi, Yuhan Ye
School of International Education, Shandong University,
hz.huizhang@sdu.edu.cn, lwt@sdu.edu.cn, 927840776@qq.com,
y195517311@163.com

Speakers modify their speech depending on communicative situations to maintain the transfer of speech signals. Increasing loudness to make speech more resistant to noise interference (Lombard speech) or signal reduction across long distance (loud speech) is a common phenomenon in daily communication. The acoustics of loud speech, as evidenced by empirical studies, are characterized by increased intensity, fundamental frequency (F0) and first formant (F1). From a production point of view, Gramming et al. (1988) proposed that the higher F0 in louder speech is caused by a higher rate of vocal fold vibration that accompanies increased subglottal air pressure (Alku, Vintturi & Vilkman, 2002). Disagreeing with Gramming et al. (1988), Titze (1994) considered higher F0 in loud speech a result of speakers' active manipulation of the vibration rate of vocal tract to increase loudness. Whichever interpretation it is, both proposals of the production mechanism of loud speech both predict concurrent change of F0 and intensity.

F0 is also found to be independently manipulated in a linguistic manner, for example in baby-directed speech and foreigner-directed speech for lexical tones. Specifically, in addition to the concurrent change of F0 and intensity, we could also observe an adjustment of F0 in loud speech that does not accompany intensity change, especially in tonal languages. The adjustment of F0 usually enriches the prosodic variation either at a sentence or a word level (F0 hyper-articulation).

Two research questions were examined in the present study. First, what is the relationship of F0 and intensity in loud speech? Second, is the hyper-articulation of F0 in loud speech, if any, phonemic or phonetic in nature?

Thirty undergraduates at Shan Dong University (15 male and 15 female) were recruited and compensated for their participation. The reading materials included 24 frequently-used characters from six syllables and four tones. The recording was executed in a sound-attenuated booth via Praat (sampling rate: 44100Hz). There were totally two recording sessions, normal speech and loud speech sessions. The normal speech session always preceded loud speech session to avoid the carryover effect of loud articulation.

F0 measurements were carried out using the ProsodyPro script developed by Xu (2013). Twenty normalized F0 values were extracted from each vowel. Manual correction was done when necessary (e.g. creaky voice). F0 was sent a Linear Mixed Effect Model as dependent variable with Tone, phonation mode, and Point as the fixed predictors, and vowels and speakers as random predictors.

Results showed that the mean F0 and intensity was higher in loud speech than normal speech (Figure 1), which is consistent with previous studies. However, the change of F0 and intensity was not constant across time. Moreover, their changes were not synchronized. Specifically, F0 increase was sharper when F0 was higher (Figure 2A), whereas intensity increase was more robust at the end of syllable (Figure 2B). This indicates that F0 manipulation is at least in part independent of intensity. Growth curve analysis showed that tonal space is expanded in loud speech since the slope distinction between Tone 1 and Tone 2 and Tone 1 and

Proceedings of the Second International Conference on Tone and Intonation, pages 59–61
November 18–20, 2023.

Tone 4, and the curvature between Tone 2 and Tone 3 are larger in loud speech than in normal speech (Figure 3), suggesting a phonemic nature of F0 hyperarticulation.

Table 1. List of reading materials.

Character	Tone	Syllable	Character	Tone	Syllable	Character	Tone	Syllable
巴	55	/pa/	科	55	/kʰɤ/	屋	55	/wu/
拔	35	/pa/	咳	35	/kʰɤ/	无	35	/wu/
把	214	/pa/	渴	214	/kʰɤ/	五	214	/wu/
爸	51	/pa/	客	51	/kʰɤ/	物	51	/wu/
低	55	/ti/	摸	55	/mo/	淤	55	/y/
敌	35	/ti/	磨	35	/mo/	鱼	35	/y/
底	214	/ti/	抹	214	/mo/	雨	214	/y/
地	51	/ti/	墨	51	/mo/	玉	51	/y/

Figure 1 (A) F0 as a function of normalized time point, gender, tone and phonation mode with 95% confidence intervals. (B) Intensity as a function of normalized time point, gender, tone and phonation mode with 95% confidence intervals.

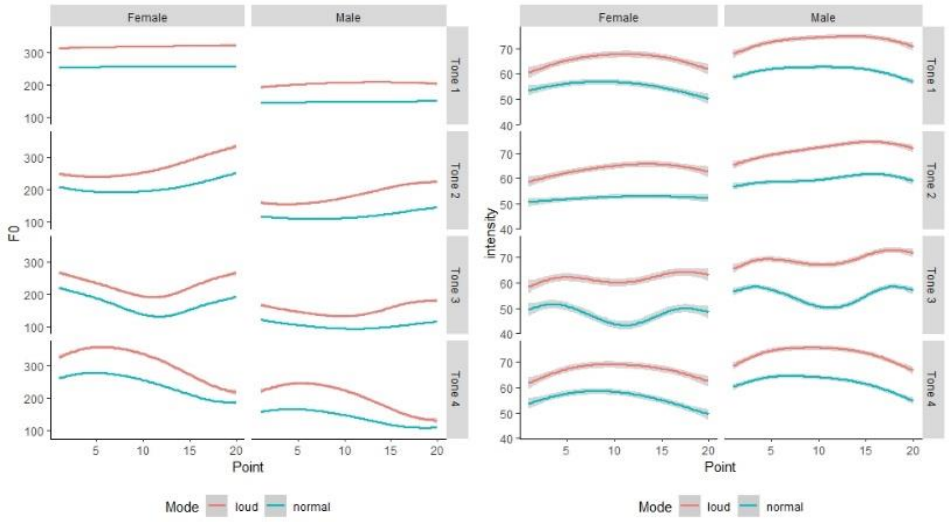


Figure 2 (A) F0 change across two phonation modes as a function of normalized time point, gender, and tone with error bars. (B) Intensity change across two phonation modes as a function of normalized time point, gender, and tone with error bars.

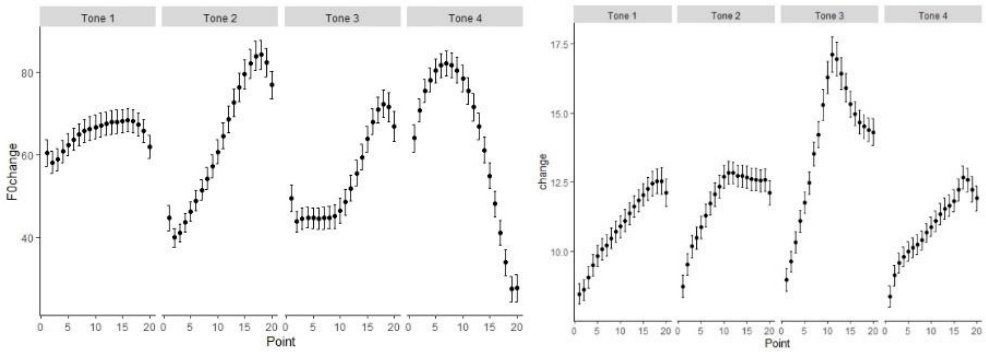
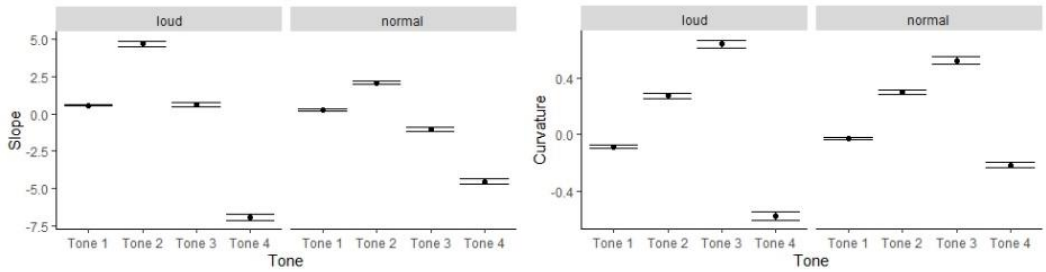


Figure 3 (A) Slope of Mandarin tones in normal and loud speech with error bars. (B) Curvature of Mandarin tones in normal and loud speech with error bars.



References

[1] Alku, P., Vintturi, J. & Vilkman, E. 2002. Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation. *Speech Communication*, 38, 321-334.

[2] Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France. 7-10.

[3] Koenig & Fuchs (2019). Vowel Formants in Normal and Loud Speech. *Journal of Speech, Language, and Hearing Research*, 62, 1278-1295.

[4] Zhao & Jurafsky (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37, 231–247.

Synthesising and Assessing Performative Speech Modes

Emily Lau[1], Brechtje Post[1], & Kate Knill[2]
[1]*Phonetics Laboratory,* [2]*Department of Engineering, University of Cambridge*
{ehynl2, bmbp2, kmk1001}@cam.ac.uk

Performative speech is a nuanced and complex phenomenon, so much so that it is difficult to pin down the exact aspects of a performance that make it appealing to the human ear. Nonetheless, this is a speech domain that has been the subject of much interest to researchers in both Linguistics and Artificial Intelligence, especially since it has been shown that performative speech has characteristics distinct from natural speech [1, 2]. While much of the research modelling expressive prosody directly maps acoustic parameters to different emotions [3, 4, 5], other work has examined emotional expression in context of the evolution of human communicative functions [6, 7]. Extending on the latter body of work, Xu et al [8] have proposed that affective speech is controlled by so-called Bio-informational Dimensions (BIDs) which control the vocal signal of the speaker in order to influence the behaviour of the receiver.

The two BIDs that are the most explored in literature and hence were investigated in this work were those of 1) size projection, the body size projected by the speaker to the listener, and 2) dynamicity, the vigorousness of the speaker's voice. Xu et al [9] showed that listeners are highly sensitive to F0 and vocal quality in judging body size and emotion, while Noble and Xu [10] similarly showed that listeners are particularly sensitive to the BIDs of size projection and dynamicity. Given these findings, BIDs are a promising theoretical framework in which to study performative speech. This work investigated how the BIDs, specifically size projection and dynamicity, impact the judgments of performative speech.

A listening test was run and designed to target the emotion "anger", for its recognizability [8]. The stimuli were utterances spoken by an SSBE-speaking male, which were re-synthesised using Praat script [8, 11] along the BIDs of size projection and dynamicity to varying degrees, in order to simulate performative expressions of anger. The listening task was divided into two sections - one containing resynthesized speech that already expressed performative anger (emotional base), and the other neutral speech (neutral base) that was resynthesized to simulate anger. 30 SSBE-speaking participants listened to pairs of these re-synthesised utterances, and were then asked to rate their differences in performative expression. Given the findings in Jurgens et al [1], it was predicted that listeners would find the stimuli with higher size projection and dynamicity more performatively angry.

As was predicted, MANOVAs for both the neutral and emotional base listening tests indicate that size projection has a significant and positive effect on listener expectations of performative anger. While dynamicity also had a significant effect, it was not as impactful as the effect of size projection. Contrary to the prediction that listeners would rate utterances with higher dynamicity as being more performatively angry, ratings consistently decreased when dynamicity increased.

In addition to using subjective measures, a series of objective measures were made to evaluate the re-synthesised utterances, namely Mel-cepstral distances, mean global variances, and mean global variance shift. The Mel-cepstral distances between the source and re-synthesised utterances were measured to assess the difference between re-synthesised utterances and the original audio. While it was expected that both size projection and dynamicity manipulations would create larger MCD values, the dynamicity manipulation had only a small effect, whilst the size projection manipulations had a noticeable impact in both directions of modulation. The mean global Mel-cepstral variances (GVs) of the utterances in each BID combination and the source utterances indicate changes in dynamicity are again shown to have little effect, despite predictions that manipulating both dimensions would create greater GVs. When considering the average GV shift between the source and resynthesised utterances, however, the size projection noticeably creates higher shifts in variance as it increases for neutral utterances. The findings from the objective measures appear to confirm the subjective findings.

These results have interesting implications about the impact of dynamicity and how it correlates with the expression of anger. While it is possible that the resynthesis process did not bring out the full effect of this manipulation, this calls into question listener expectations of performative anger and whether this was reflected in the stimuli. Further research should investigate these correlations and the nuances of different emotional subsets.

Proceedings of the Second International Conference on Tone and Intonation, pages 62–63
November 18–20, 2023.

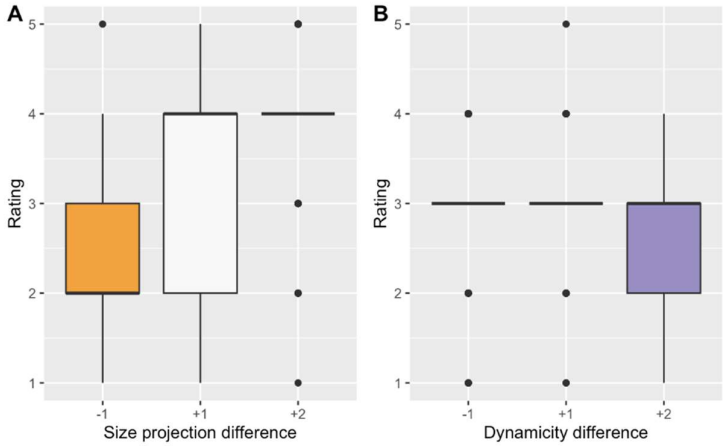


Figure 1: *Ratings of dramatic anger across all trials for the neutral synthesis base, where A indicates the ratings for each size projection difference, while B indicates the ratings for each dynamicity difference.*

Table 1: *Mel-cepstral distances between neutral source utterances and re-synthesised listening test utterances by BID combination.*

	0 DYN	**+1 DYN**	**+2 DYN**
0 SP	Control	5.886	6.126
+1 SP	6.716	6.722	6.802
+2 SP	8.699	8.551	8.536

References

[1] Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. Frontiers in psychology, 2, 180.
[2] Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of acting experience on emotion expression and recognition in voice: Non-actors provide better stimuli than expected. Journal of nonverbal behavior, 39, 195-214.
[3] Scherer, K. R. (1974). Voice quality analysis of American and German speakers. Journal of Psycholinguistic Research, 3(3), 281-298.
[4] Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance.
[5] Mozziconacci, S. (2002). Prosody and emotions. In Speech Prosody 2002, International Conference.
[6] Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. The American Naturalist, 111(981), 855-869.
[7] Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F_0 of voice. Phonetica, 41(1), 1-16.
[8] Xu, Y., Kelly, A., & Smillie, C. (2013). Emotional expressions as communicative signals. Prosody and iconicity, 33-60.
[9] Xu, Y., Lee, A., Wu, W. L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. PloS one, 8(4), e62397.
[10] Noble, L., & Xu, Y. (2011). Friendly Speech and Happy Speech-Are They the Same?. In ICPhS (pp. 1502-1505).
[11] Boersma, P., & Weenink, D. 2008. *Praat: doing phonetics by computer*. Computer program.

A Preliminary Investigation of the Phonetic Characteristics of Moklen Tones

Warunsiri Pornpottanamas[1], Pittayawat Pittayaporn[2] & Sireemas Maspong[3]
[1]*Department of English and Linguistics, Ramkhamhaeng University, Bangkok, Thailand,*
[2]*Department of Linguistics & Southeast Asian Linguistics Research Unit, Chulalongkorn University, Bangkok, Thailand,*
[3]*Institute for Phonetics and Speech Processing (IPS), LMU Munich, Germany*
warunsiri.p@rumail.ru.ac.th, pittayawat.p@chula.ac.th,
s.maspong@phonetik.uni-muenchen.de

Moklen, an endangered Austronesian language spoken on the Andaman coast of Southern Thailand, is a significant case for studying tonogenesis [1, 2, 3]. While previous research by Pittayaporn et al. [3] confirmed the presence of two lexical tones in Moklen, the nature of this tonal contrast remains unclear. Swastham [4] and Larish [5, 6] proposed that Moklen tones emerged through contact with Southern Thai, but it is uncertain whether Moklen tones developed from segmental sources following Haudricourt's model of tonogenesis, given that Moklen still maintains contrastive voicing in onsets [7]. Examining the acoustic characteristics of Moklen tones can provide insights into the hypothesis regarding their origins. In this preliminary study, we conducted an instrumental analysis on the phonetic properties of Moklen tones in different syllable types and onset voicing. Our results reveal that Moklen tones primarily differ in fundamental frequency (f0), accompanied by difference in phonation type at the beginning of the vowel.

Methodology: Four native Moklen speakers residing in Takua Pa District, Phang Nga Province, Thailand, participated in the study. They were asked to produce Moklen mono- and disyllables in isolation, with each word repeated three times. The 93 target words were carefully selected to have stressed final syllables with /a, aː/ vowels. These target words were systematically varied in terms of tones, onset voicing, vowel length, and coda classes.

Analysis: Acoustic measurements were obtained from the stressed syllables using PraatSauce [8]. Five acoustic measurements associated with tonation in various languages were selected, including f0, F1, F2, the difference between corrected first harmonics and the corrected spectral amplitude of F3 (H1*-A3*), and Cepstral Peak Prominence (CPP). Each measurement was z-scored by speakers and time-warped to a fixed length, facilitating meaningful comparisons. We employed separate linear mixed effect regressions with four dependent variables: (i) the mean values of f0, F1, F2, H1*-A3*, and CPP during the first quarter of vowel trajectories, (ii) the intercepts (indicating means of each trajectory), (iii) the linear coefficients (indicating slopes), and (iv) the quadratic coefficients (indicating curvatures). Polynomial curve fitting was applied to each acoustic trajectory to obtain coefficients (ii)-(iv). Independent variables included tone, onset voicing, vowel length, and coda class. Subject was included as a random intercept in the analyses.

Results: The results demonstrated significant differences in the first quarter values of f0 (Tone 1 > Tone 2), mean (Tone 1 > Tone 2), slope (with steeper Tone 2 slope > Tone 1), and curvature across Moklen tones. Notable differences were also observed in the beginning of H1*-A3* (Tone 1 < Tone 2) and CPP (Tone 1 > Tone 2), indicating a breathier voice quality for Tone 2. Additionally, we found significant effects of onset voicing on several acoustic measures, including the first quarter, mean, slope, and curvature of f0, the first quarter, mean, and slope of F1, the first quarter and mean of F2, the first quarter and mean of H1*-A3*, and the first quarter, mean, and slope of CPP. Moreover, vowel length exerted significant effects on the first quarter of f0, the first quarter, mean, and slope of F1, the first quarter and mean of F2, and the first quarter and curvature of CPP. Lastly, the coda class exhibited marginal effects on the first quarter of f0, the first quarter and slope of F1, the curvature of F2, the curvature of H1*-A3*, as well as the first quarter, mean, slope, and curvature of CPP.

Discussion and Conclusion: Our experimental findings suggest that f0 is the primary phonetic cue for tonal contrast in Moklen, accompanied by the difference in phonation type at the beginning of the vowel. Specifically, Tone 1 is characterized by higher pitch and a vowel with modal voice, while Tone 2 has a lower pitch and a breathier vowel. These characteristics bear similarities to register distinctions observed in Austroasiatic languages of Southeast Asia [9], suggesting a possible transphonologization of laryngeal properties into prosodic ones in Moklen. However, the exact segmental sources of Moklen tones still remain an open question.

Proceedings of the Second International Conference on Tone and Intonation, pages 64–65
November 18–20, 2023.

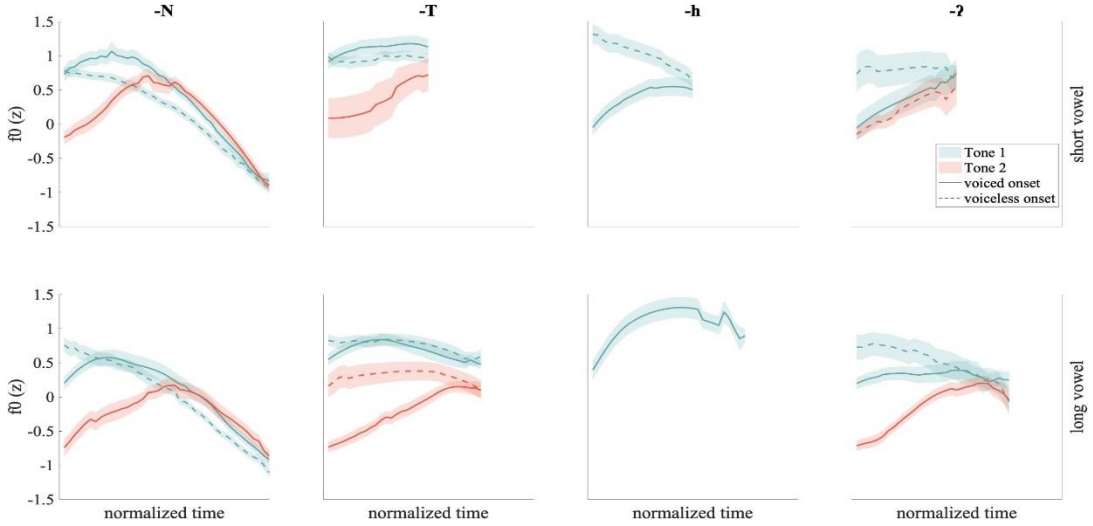


Figure 1: *F0 trajectories of Moklen tones over the vowel. Trajectories of f0 for CV(V)N syllables are over the entire rimes.*

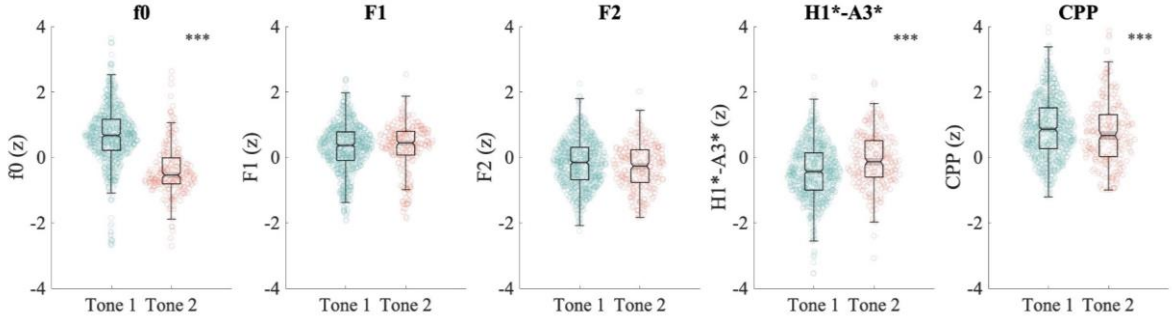


Figure 2: *Boxplot of mean values of f0, F1, F2, H1*-A3*, and CPP during the first quarter of vowel trajectories. Significant differences across tones are found for f0, H1*-A3*, and CPP.*

References

[1] Larish, M. D. 2005. Moken and Moklen. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar*. Routledge, 513-533.

[2] Premsrirat, S. 2006. Language situation in Thai society and ethnic diversity. *Journal of Language and Culture* 25(2), 5-17.

[3] Pittayaporn, P., Pornpottanamas, W., & Loss, D. (Eds.). Moklen-Thai-English dictionary: a pilot version. Bangkok: Academic Work Dissemination Project, Faculty of Arts, Chulalongkorn University, 2022.

[4] Swastham, P. 1982. A description of Moklen: a Malayo-Polynesian language in Thailand [Master's thesis, Mahidol University]. Mahidol University Library and Knowledge Center. http://mulinet11.li.mahidol.ac.th/e-thesis/scan/003600.pdf

[5] Larish, M. D. 1997. Moklen-Moken phonology: Mainland or insular Southeast Asian typology? In C. Odé & W. Stokhof (Eds.), *Proceedings of the Seventh International Conference on Austronesian Linguistics* (Rodopi), 125-149.

[6] Larish, M. D. 1999. *The position of Moken and Moklen within the Austronesian language family*. [Unpublished doctoral dissertation]. University of Hawai'i.

[7] Haudricourt, A. 1954. De l'origine des tons en viêtnamien. *Journal Asiatique* 242. 69-82.

[8] Kirby, J. 2021. PraatSauce: Praat-based tools for spectral analysis. Available at: https://github.com/kirbyj/praatsauce

[9] Brunelle, M., & Kirby, J. 2015. Re-assessing tonal diversity and geographical convergence in Mainland Southeast Asia. In N Enfield & B Comrie (Eds.), *Languages of Mainland Southeast Asia: The State of the Art* (Mouton de Gruyter), 82-110. DOI: 10.1515/9781501501685-004

No effects of f0 manipulation and phrase position in Korean word recognition

Constantijn Kaland[1], Matthew Gordon[2], Jiyoung Jang[2], Argyro Katsika[2]
[1]*Institute of Linguistics, University of Cologne*
[2]*Department of Linguistics, University of California, Santa Barbara*
ckaland@uni-koeln.de, mgordon, jiyoung, argyro {@ucsb.edu}

Studies have shown that the f0 shape of words facilitates their recognition (e.g., Laures & Wiesmer, 1999; Hillenbrand, 2003). This effect has been observed mainly for pitch accent languages, such as English, in which the f0 shape of an accent generally aligns with the word stress pattern (e.g. Cutler and Foss, 1977; Cutler, 2012). Less is known about the role of f0 in word recognition in languages without pitch accents and/or word stress. The main question is whether f0 facilitates word recognition at all in these languages, as this cue might solely operate on the phrase level, such as the intonational, intermediate or accentual phrase (IP/ip/AP). The current study extends earlier work that investigated the role of f0 and phrase position in Papuan Malay and American English (Kaland & Gordon, 2022). We investigate Korean, which mainly uses prosody to mark phrase edges and does not have word stress (an 'edge language' in Jun, 2014). Previous work has shown that Korean words often form an AP and that the AP-final rise facilitates listeners' word *segmentation* (Kim, 2004). It was also found that final lengthening is a helpful cue only to a limited extent (Kim & Cho, 2009). If the functions of Korean prosody are indeed restricted to phrase edges, the question remains whether f0 shape interacts with position in the phrase in this language and whether these factors facilitate word *recognition*.

A forced choice word identification task was carried out, in which listeners decided as fast as possible which of two written words (target and distractor) occurred in the auditorily presented carrier phrase (replication of Kaland & Gordon, 2022). Targets and distractors were disyllabic and had overlapping initial segments. Thus, there was a uniqueness point (UP) from which the target could be uniquely identified. Targets had an original f0 or a flattened (manipulated) f0, positioned either medially (1) or finally (2) in the utterance.

1)	네가	말한	그 단어	나비는,	나는	모르겠어
	ni-ga	mal.hankɯ	tan.ʌ	[T]-neun,	na-nɯn	mo.lɯ.ke.sʌ
	you-NOM	mention	the word	[T]-TOP,	I-TOP	don't know
	'The word [T] you mentioned, I don't know.'					

2)	나는	모르겠어,	네가	말한	그 단어	나비.
	na-nɯn	mo.lɯ.ke.sʌ,	ni-ga	mal.hankɯ	tan.ʌ	[T]
	I-TOP	don't know,	you-NOM	mention	the word	[T]
	'I don't know the word [T] (you mentioned).'					

The results show that Korean listeners were not affected by f0 shape or position (Figure 1, Table 1). Thus, participants took similar time to identify a target when it was presented with original or manipulated f0, and whether it occurred in medial or final position. This seems to indicate that f0 does not operate at the word level, which refines the results for the accentual phrase in word segmentation (Kim, 2004) and confirms the analysis of Korean as an edge language (Jun, 2014) and the analyses of Papuan Malay and American English as head/edge language (tentatively) and head language respectively (Kaland & Gordon, 2022). The results also seem to indicate that listeners did not benefit from final lengthening at the end of the carrier phrase, unlike listeners of head(/edge) languages (cf. Kaland & Gordon, 2022). It is not the case, however, that final lengthening did not occur on the Korean disyllabic target words (i.e., medial: 286 ms, final: 358 ms). The lack of phrase position effect could be explained when considering the status of the particle '-neun' (는). When this particle is analyzed as being part of the target word (which would be justified morphologically), the medial target words are approximately 240 ms longer (526 ms). In addition, Korean phrasing required a pause after the medial target. Thus, participants had at least as much time to process the medial targets than the final targets, which most likely counterbalanced any facilitative effect of final lengthening in the current stimulus set. In sum, the results lend further support for a prosodic analysis of Korean in which f0 operates solely on the phrase level (e.g., Kim, 2004).

Proceedings of the Second International Conference on Tone and Intonation, pages 66–67
November 18–20, 2023.

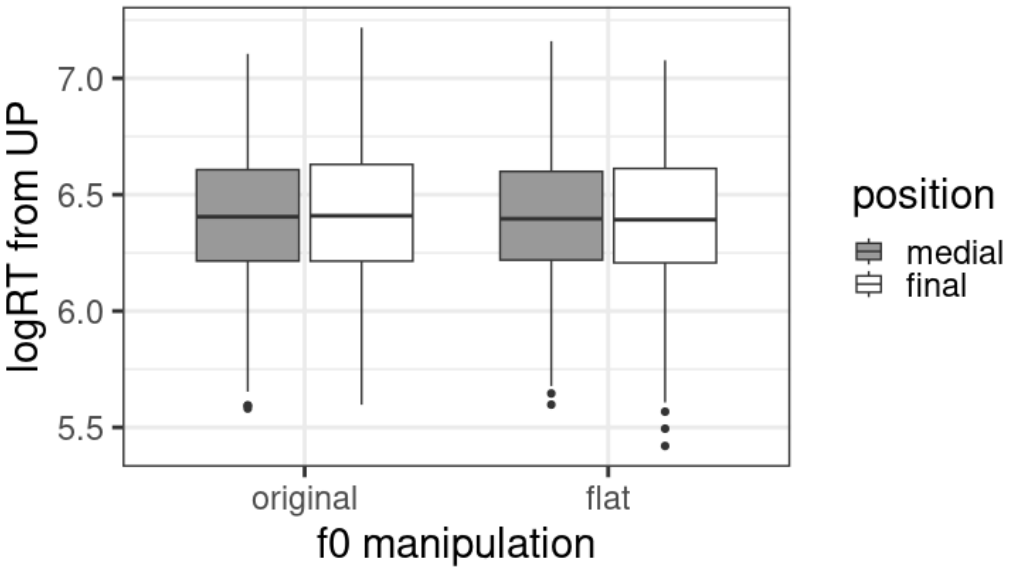


Figure 1. Log reaction times (logRT) to word identification from the uniqueness point (UP) in the target word, split for f0 manipulation (original, flat) and position (medial, final).

Table 1. Summary of the main fixed factors in the linear mixed effects model, factors not listed did not show a significant effect.
`logRT from UP ~ f0 manipulation * position + no. of segments before UP + stimulus order + (1|item) + (1|participant)`

Factor	*Estimate*	*SE*	*df*	*t*	*p*
(Intercept)	6.39	0.05	33.97	122.10	< 0.001
f0 manipulation	-0.01	0.01	1870.33	-1.09	n.s.
position	-0.01	0.04	39.73	-0.33	n.s.
f0 manipulation:position	0.01	0.02	1853.70	0.63	n.s.

References

[1] S. Kim, “The role of prosodic cues in word segmentation of Korean,” in *Interspeech 2004*, ISCA, Oct. 2004, pp. 3005–3008. doi: 10.21437/Interspeech.2004-754.

[2] J. S. Laures and G. Weismer, “The Effects of a Flattened Fundamental Frequency on Intelligibility at the Sentence Level,” *J Speech Lang Hear Res*, vol. 42, no. 5, pp. 1148–1156, Oct. 1999, doi: 10.1044/jslhr.4205.1148.

[3] A. Cutler and D. J. Foss, “On the Role of Sentence Stress in Sentence Processing,” *Lang Speech*, vol. 20, no. 1, pp. 1–10, Jan. 1977, doi: 10.1177/002383097702000101.

[4] C. Kaland and M. K. Gordon, “The role of f0 shape and phrasal position in Papuan Malay and American English word identification,” *Phonetica*, vol. 79, no. 3, pp. 219–245, Jun. 2022, doi: 10.1515/phon-2022-2022.

[5] J. M. Hillenbrand, “Some effects of intonation contour on sentence intelligibility,” *The Journal of the Acoustical Society of America*, vol. 114, no. 4, pp. 2338–2338, Oct. 2003, doi: 10.1121/1.4781079.

[6] S. Kim and T. Cho, “The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean,” *J. Acoust. Soc. Am.*, vol. 125, no. 5, p. 3373, 2009, doi: 10.1121/1.3097777.

[7] A. Cutler, *Native listening: language experience and the recognition of spoken words*. Cambridge, Mass: The MIT Press, 2012.

[8] S.-A. Jun, Ed., *Prosodic typology II: the phonology of intonation and phrasing*. in Oxford linguistics. Oxford: Oxford University Press, 2014.

Individual variability in the production of consonant-induced pitch perturbations at vowel onset in Thai

Alif Silpachai, Na Hu, & Amalia Arvaniti
Radboud University
alif.silpachai@ru.nl, na.hu@ru.nl, amalia.arvaniti@ru.nl

Previous research has suggested that consonant-induced pitch perturbations at vowel onset (CF_0) in a tone language can be produced differently by different speakers. For example, some speakers of Thai produce higher fundamental frequency (F_0) following a voiceless unaspirated stop than an aspirated stop, while some produce lower F_0 [1]. It is unclear whether individual differences in the productions of CF_0 in a tone language are structured (not random). It is possible that a speaker who produces high F_0 following /p/ also produces high F_0 following /t/ and /k/. This study examined the extent to which individual variation in the realization of CF_0 is structured. It was predicted that F_0 following heterorganic stops in the same voicing category (e.g., /p t k/) would be correlated with each other, unlike F_0 following homorganic stops (e.g., /b p p^h/).

Twelve native speakers of Thai (6 females and 6 males, *M* = 45.75, *SD* 13.97, years) spoken in Chonburi, Thailand produced monosyllabic words starting with /b/, /p/, /p^h/, /d/, /t/, /t^h/, /k/, or /k^h/, having /a/ or /aː/ as the vowel, bearing the falling (/51/), mid (/32/), or low (/21/) tone, and placed in a word either in isolation, a declarative statement, or an alternative question. The statement, adapted from the one in [2], could be translated as "I will say the word _____ for you," and the question, based on the one in [3], had the following structure: "Did {you, s/he} {say, read, write} _____ or Y again?" where Y was a word that differed from the target word either in its vowel quality or onset consonant. The mean of each speaker's CF_0 was measured in two ways. In one, we measured the mean of three F_0 points, 10, 20, and 30 ms, following the onset of voicing (vowel onset), and in another way, we measured F_0 changes over time, based on the differences between the last and first F_0 points following the onset of voicing. All measurements of F_0 were transformed to semitones relative to each speaker's mean F_0.

There were significant correlations between the mean F_0 following /p^h/ and /t^h/ and between /p^h/ and /k^h/ ($ps < .01$) (see Table 1). However, no correlations were found for voiced or unaspirated stops. There were also correlations between the mean F_0 change over time following heterorganic stops sharing a voicing category ($ps < .05$) (see Table 2). The results suggested that a speaker who produced high F_0 following /p^h/ tended to also produce high F_0 following /t^h/ and /k^h/. Additionally, those who produced large F_0 changes following one place of articulation (POA) (e.g., /b/) tended to produce large F_0 changes following another POA (e.g., /d/).

The results suggested that individual variability in certain suprasegmental cues associated with stops, particularly in F_0 changes over time, are highly structured. There was likely a constraint of uniformity that required that the talker-specific realization of members of a consonant class be uniform [4]. Such a constraint may aid the perception of talker-specific secondary cues of stops. For example, listeners may use the knowledge that the mean F_0 changes over time following /p/ and /t/ correlate with each other to predict the mean F_0 changes following /k/ produced by an unfamiliar talker (cf. [4]).

Proceedings of the Second International Conference on Tone and Intonation, pages 68–69
November 18–20, 2023.

Table 1

The *p* values of the correlation analyses of the relationship between F_0 following heterorganic consonants in the same voicing category. For brevity, only significant results are shown.

Correlation between	*p* value	*r*
/pʰ/ and /tʰ/	.003	.77
/pʰ/ and /kʰ/	.005	.75

Table 2

The *p* values of the correlation analyses of the relationship between mean F_0 changes over time following heterorganic consonants in the same voicing category. For brevity, only significant results are shown.

Correlation between	*p* value	*r*
/b/ and /d/	.001	.82
/p/ and /t/	.009	.72
/p/ and /k/	.007	.73
/t/ and /k/	.017	.67
/pʰ/ and /tʰ/	< .001	.93
/pʰ/ and /kʰ/	< .001	.86
/tʰ/ and /kʰ/	< .001	.96

References

[1] Erickson, D. (1976). *A physiological analysis of the tones of Thai* [Unpublished doctoral dissertation]. University of Connecticut.

[2] Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *The Journal of the Acoustical Society of America*, *120*(5), 2884–2895. https://doi.org/10.1121/1.2346131

[3] Kirby, J. P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics*, *71*, 326–354. https://doi.org/10.1016/j.wocn.2018.09.009

[4] 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. https://doi.org/10.1016/j.wocn.2017.01.001

The effect of post-tonic lengthening and F0 in Syrian Arabic prosody

Malek Al Hasan & Shakuntala Mahanta
Indian Institute of Technology Guwahati, India
malek@iitg.ac.in, smahanta@iitg.ac.in

The role of final boundary is well-known in the intonation literature [1], but whether post-tonic lengthening (PTL) influences intonation and prosodic structure in languages is not very well-known. There is some literature on prosodically conditioned PTL across languages [2] & [3], but the additional factor in Syrian Arabic (henceforth SyrA) is the pragmatic factor. In this paper, we investigate this in SyrA and show how it plays an important role in SyrA sentence-level prosody.

Experiment: Twenty utterances of statements and questions were recorded from six native speakers of SyrA. The target words consisted of two syllables representing stressed and unstressed conditions with stress falling on the penultimate syllable. The alignment of F0 peaks and valleys in the utterances along with the duration of the phrase-final segments were investigated. Durational measures were obtained from statements (SVO word order) and question utterances (wh & yes/no) to determine the acoustic features of vowel PTL [4] occurring in ips (non-utterance-final units) and in questions. In addition, similar utterances of statements were compared with the yes/no questions to determine the degree of post-tonic lengthening occurring in questions. Yes/no questions in SyrA have no morphological markings or syntactical means to differentiate them from statements. Since they have the same morpho-syntactic form, yes/no questions are distinguished from statements by intonation only.

[5] proposed an analysis of three levels of phrasing for SyrA intonation, the Accentual Phrase (AP), the Intermediate Phrase (ip), and the Intonational Phrase (IP). The AP in SyrA contains one or more prosodic words, and the IP can have more than one AP. The Intermediate Phrase (ip) is a domain of a syntactic structure SVO (the subject, the verb, and the direct object) in SVO (Subject – Verb (ditransitive) – Direct Object – Indirect Object) word order.

F0 was found to play a crucial role in conveying the prominence (pitch accents) and boundary information within the AP in SyrA. The AP in SyrA is marked on its left edge by a high tone H* realized on the stressed syllable and marked by a low tone La on the right edge of the AP. This suggests that the stressed syllable is the locus of the High tone and is associated with the rising tonal pattern of the AP. The default pattern of APs in SyrA was realized as [H* La], as shown in Fig. 1 below.

F0 and duration were considered to be important factors in cueing the Intermediate Phrase (ip) in SyrA. The ip in SyrA is also characterized by a rising contour but with a high tone aligned to its right edge. The ip boundary tone H- is distinct from the AP-initial tone H* in which the ip boundary tone H- has a higher pitch and boosted pitch range than the AP-initial tone H*. In Fig. 2 below, the peak on the last syllable of the word [ˈləʕbeh] ‘doll’ is higher than the preceding H* peak, breaking the declination slope among H* peaks. This higher peak is an ip boundary tone H- marking the direct object of the sentence. In addition to this, durational lengthening was reported in the ip (utterance-non-final item). The vowel duration of the last unstressed syllable [be] of the word [ˈləʕbe] is longer than the preceding stressed counterpart (Fig. 5).

The pitch contour of Wh-questions in SyrA is characterized by a *downstepped rise* on the last unstressed syllable of the phrase-final content word. The reason for this final *downstepped rise* in SyrA question tunes is due to phrase-final PTL in SyrA. Similarly, the utterance-final *up-step* in yes/no questions is accompanied by a phrase-final PTL. In SyrA questions, post-tonic short vowels are lengthened phrase-finally in word-final position. We show the durational differences between statements and questions in Fig. 6 below. A linear mixed-effects model was conducted using the *lmer* function from the *lme4* package in R to test the durational measurements in statements and questions. The results of the model showed a significant effect where the durations of the vowels in the target words were different. The results and the model will be presented later in the full paper.

PTL was found to play an essential discourse-level function in SyrA prosody at the phrasal level, which is found to be a typical feature of SyrA differentiating it from the other dialects of Arabic. The utterance-final high rise in SyrA questions is accompanied by a phrase-final PTL which is referred to by Kulk et al. [4] as "singing intonation." PTL in SyrA is both prosodically conditioned and at the discourse pragmatic level. Prosodic boundaries that have significant lengthening are the ip and the question IPs.

Duration has been shown to play a significant role in the dialects of Arabic such as Lebanese [6], Egyptian [7], and Jordanian [8]. However, SyrA is unique in the role that PTL plays at the level of

Proceedings of the Second International Conference on Tone and Intonation, pages 70–71
November 18–20, 2023.

intonation, not just in lengthening unstressed syllables but influencing the up-stepped F0 in yes/no questions and the downstepped rise in wh-questions.

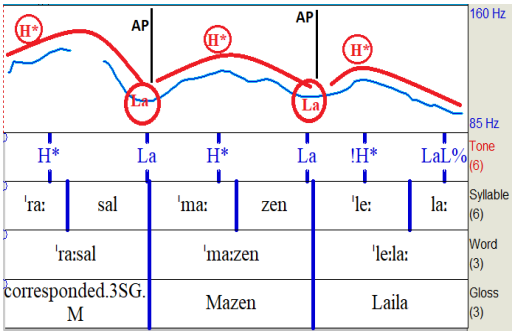


Figure 1: *AP in SyrA with VSO word order in the utterance [ˈra:sal ˈma:zen ˈle:la] "Mazen corresponded Laila."*

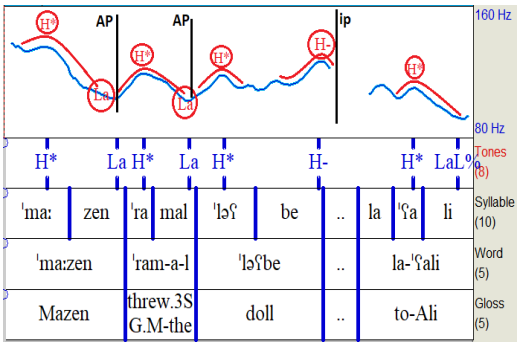


Figure 2: *The ip in SyrA with SVO word order. in the utterance [ˈma:zen ˈram-a l-ˈləʕbeh la-ˈʕa:del] "Mazen threw the doll to Adel"*

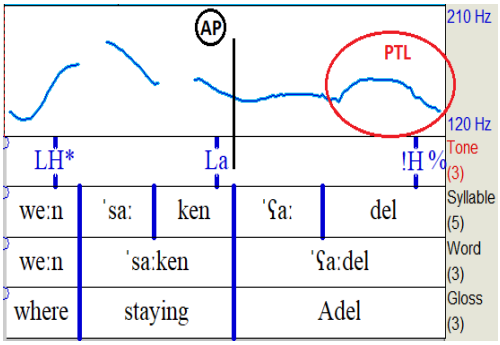


Figure 3: *Wh-Q in the utterance [we:n ˈsa:ken ˈʕa:del?] "Where is Adel staying?"*

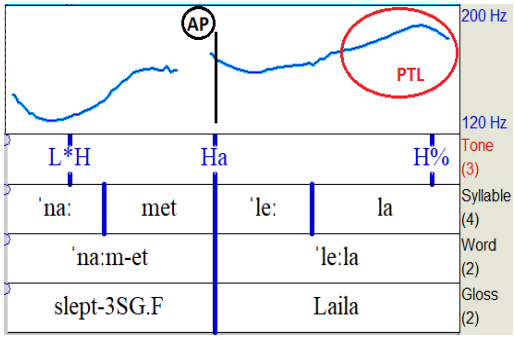


Figure 4 : *YNQ in the utterance [ˈna:met ˈle:la?] "Did Laila spleep?"*

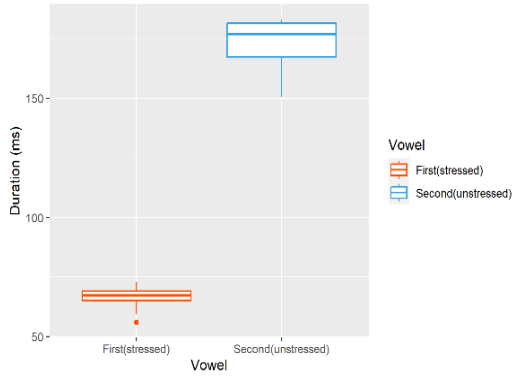


Figure 5: *Vowel duration of stressed vs. unstressed vowels representing post-tonic vowel lengthening in SyrA statements in the ip-final word of the utterance in Figure 2.*

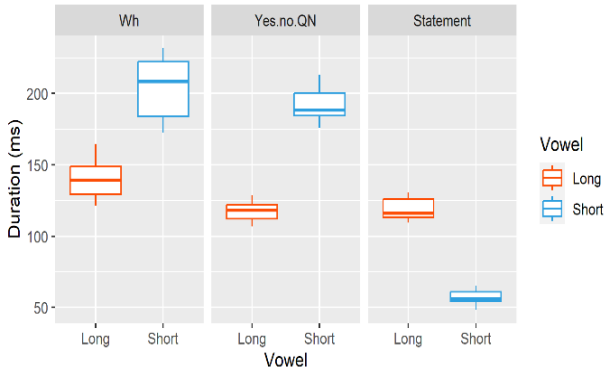


Figure 6: *post-tonic vowel lengthening in SyrA questions, and Yes/no utterances compared to similar statement utterances.*

References

[1] Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P., 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91:1707–17.

[2] Leal, E.D.G. and Santos, R.S., 2010. Post-tonic syllables and prosodic boundaries in Brazilian Portuguese. In *Speech Prosody 2010-Fifth International Conference*.

[3] Esposito, L., 2020. Linking gender, sexuality, and affect: The linguistic and social patterning of phrase-final posttonic lengthening. *Language Variation and Change*, 32(2), pp.191-216.

[4] Kulk, F., Odé C., Woidich M., 2003. The intonation of colloquial Damascene Arabic: a pilot study. In *IFA Proceedings*, vol 25, pp. 15-20.

[5] Al Hasan, M., Mahanta, S. 2022. The Intonational Phonology of Syrian Arabic: A Preliminary Analysis. *Proceedings of Speech Prosody 2022*, 190-194, doi: 10.21437/SpeechProsody.2022-39.

[6] Chahal, D. 2001. *Modeling the intonation of Lebanese Arabic using the autosegmental-metrical framework: a comparison with English*. Unpublished PhD thesis, University of Melbourne, 2001.

[7] Hellmuth, S. 2006. *Intonational pitch accent distribution in Egyptian Arabic*. Unpublished PhD thesis, SOAS.

[8] de Jong, K., Zawaydeh, B. 1999. Stress, Duration, and Intonation in Arabic Word-Level Prosody. *Journal of Phonetics*, vol. 27, pp. 3–22.

Vowels, Tones and Tonogenesis in Braj Bhasha

Neetesh Kumar Ojha[1] & Shakuntala Mahanta[2]
Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati, India
k.neetesh@iitg.ac.in[1] & smahanta@iitg.ac.in[2]

The aim of our study is (a) to first acoustically plot the vowel space of the understudied language Braj which is a great research gap, as vowels are the TBUs; and then (b) to examine the tone and tonogenesis in the language. Braj, or Brij, or Brajbhasha, (Skt. Vraja or Antarbēdī), was a prestigious language in northcentral India in medieval times [1], but its influence declined by the 18th century, replaced by Khadi-Boli Hindi. Currently, it is only limited to the Braj region [2][3].

We first plot 10 vowels from a corpus of paradigmatic words (33 tokens/vowel) using Lobanov normalization method of formants' analysis (figure 1). It was crucial as Braj vowels were never acoustically explored in any earlier study. Figure 1 shows 7 tense and 3 lax vowels, with a phonetic conditioning that the 3 lax vowels are always shorter in duration while the 7 tense vowels are always longer, such that, an average tense-vowel is approximately twice the duration compared to an average lax-vowel (confirmed by a t-test).

Then, for the first time in Braj, our study shows that Braj, despite being considered a variety in the Hindi continuum, patterns with the North West IA languages in attesting a low tone due to the F0 perturbation created by the loss of breathy voiced fricative [4][5][6]. Our study shows a tonogenetic basis created by the loss of [h] or [ɦ] at coda positions (also intervocalically) leading to lexical tonal contrasts. A small corpus of 9 pairs of monosyllabic Braj words (6 with monophthongs and 3 with diphthongs as nuclei) with 594 tokens (speakers×words×iterations=11×18×3) was used for our study. Table 1 has only few examples for lack of space in this abstract. The ANOVA conducted on the F0 information to test the existence of tonal contrast in our hypothesized tonal pairs from Braj speech showed a significant difference of F0 values (F-value 12.47, p-value 0.00958). The predictor variable "f0_neutral" had 1 degree of freedom and sum of squares of 172.19. The residuals had 7 and 96.64 respectively, and a mean square of 13.81. The results showed that the words labelled "T" (our hypothesized truth-value for the words from pairs with the hypothesized 'presence' of low-tone from field experience) really had lower F0 means (figure 2). The tonogenetic origins suggest that the loss of breathy glottal fricative at coda positions of NIA (New Indo-Aryan) cognates phonologically caused a low tone in Braj. Further analysis of the normalized contours (figure 3, figure 4) also showed a contour-contrast between monophthongs (phonetically falling contour) and diphthongs (phonetically falling-rising contours). The NIA cognates of the latter are generally found to have [h] or [ɦ] intervocalically.

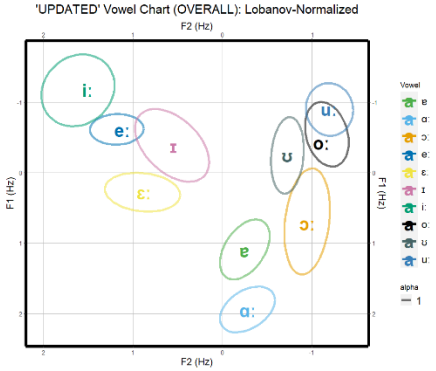


Figure 1: *Braj Bhasha Vowel Space*

Table 1: *Monosyllabic Tonal Pairs Classified as per Thong_ID, Mora_ID, Pair_ID and Low Tone*

S. No.	Thong_ID	Mora_ID	Pair_ID	Low Tone	Words	Glosses
1.	Monophthongs	Bimoraic	gɑ:- gɑ̀:	F	gɑ:	'sing'
2.				T	gɑ̀:	'give'
3.		Trimoraic	kɔ:r- kɔ̀:r	F	kɔ:r	'bite'
4.				T	kɔ̀:r	'mist'
5.	Diphthongs	Tetramoraic	kɑ:i:- kɑ̀:í:	F	kɑ:i:	'algae'
6.				T	kɑ̀:í:	'what was'

Proceedings of the Second International Conference on Tone and Intonation, pages 72–73
November 18–20, 2023.

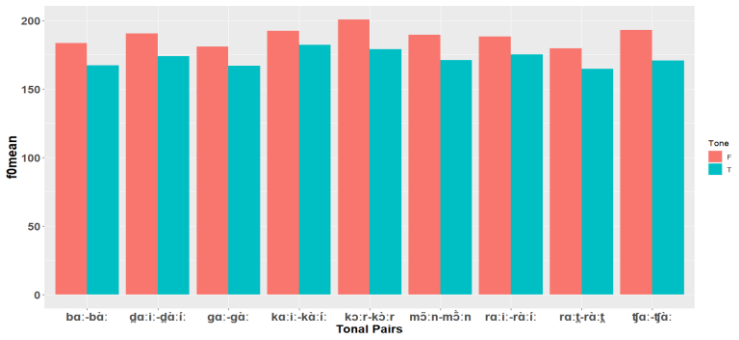


Figure 2: *Contrastive f0means in the Tonal Pairs of all Monosyllabic Words*

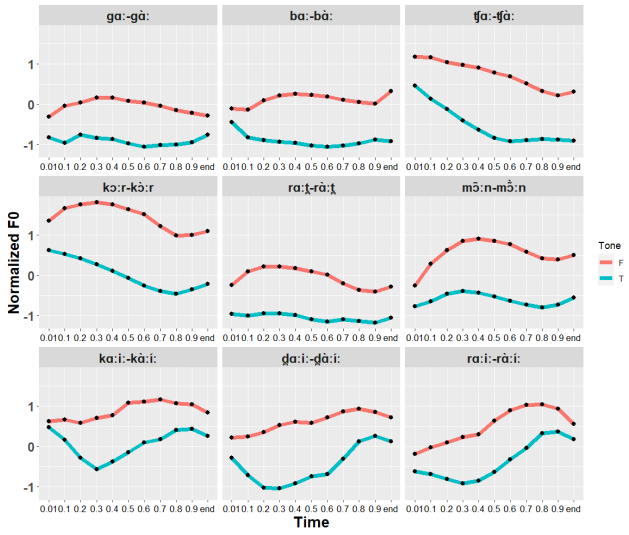


Figure 3: *Results of Normalized F0 Contour Analysis as grouped by individual Pair_ID*

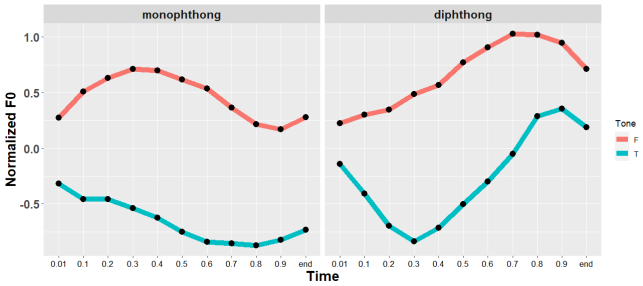


Figure 4: *Results of Normalized F0 Contour Analysis as grouped by Thong_ID*

References

[1] Grierson, G. A. (1916). Linguistic Survey of India, *Volume 9, Indo-Aryan Family. Central Group. Part 1. Specimans of Western hindi and Panjabi, 69-81*. Office of the Superintendent of Government Printing, Calcutta, India.

[2] Census of India, (2011). Paper 1 of 2018: LANGUAGE, *India, States and Union Territories (Table C-16)*. Office of the Registrar General, India, 2A, Mansingh Road, New Delhi - 110011.

[3] Chandan, S. K. (2014). Introducing Braj bhasha archive for the study of the history of Mughal India.

[4] Baart, J. (2014). Tone and stress in north-west Indo-Aryan. *Above and beyond the segments: Experimental linguistics and Phonetics, Amsterdam: John Benjamins*, 1-13.

[5] Kanwal, J., & Ritchart, A. (2015). An experimental investigation of tonogenesis in Punjabi. In ICPhS.

[6] Mishra, D., & Bali, K. (2011, August). A Comparative Phonological Study of the Dialects of Hindi. In *ICPhS* (Vol. 17, pp. 17-21).

Intonation of angry and happy Singapore English acted speech

Rae Jia Xin Koh[1,2] & Ying-Ying Tan[2]
[1] *Global Asia, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore* [2]*Department of Linguistics and Multilingual Studies, School of Humanities, Nanyang Technological University, Singapore*
kohj0058@e.ntu.edu.sg, yytan@ntu.edu.sg

Intonation conveys multiple levels of information such as organizing word, phrasal, and turn boundaries, invoking implicatures and pragmatic meaning, and central to the interest of this work – expressing or even revealing cues about the emotions that a speaker may be experiencing [1], [2]. Though it may seem to be the case that emotions can be "taken to be more directly expressed intonationally" [2, p.223], for example, anger and happiness presenting themselves in intonation with a larger pitch range and steep falling pitch contours [3], [4], it is still unclear how these same emotions modulate intonation for speakers of different cultures and even different language varieties. As [5, p.4] noted, "the same kind of melodic rise or fall can be the result of different grammatical features or properties, assigned in different ways to prosodic constituents." For Singapore English (SgE), this is even more complex. One needs to understand the ways in which intonation acts as a medium for emotion communication given Singaporeans' multicultural and multilingual background. This throws into question as to whether the same intonational patterns and structure derived from studies primarily focusing on English-speaking counterparts in the Western world could still be applicable to Singapore English. Singapore's population is made up of three main ethnic groups – the Chinese, Malays, and Indians. Studies have found that collectivism, which describes the three Asian communities, is correlated with less emotional expressivity norms as compared to Western countries that rank higher for individualism [6]. Furthermore, [7] reported ethnic differences among Singaporeans in the expression/ suppression of disgust. In tandem with these cultural differences, research on the prosody of SgE has also found that the variety is "radically different" from Southern British English [1, p.453]. Some observations of these intonational differences that have been made include the tendency for phrase-final syllable and word prominence that is not necessarily realized by a change in pitch but rather in terms of lengthening and increase in loudness, as well as the possibility for multiple prominent syllables in the same tone group even when it is non-contrastive [8]-[12].

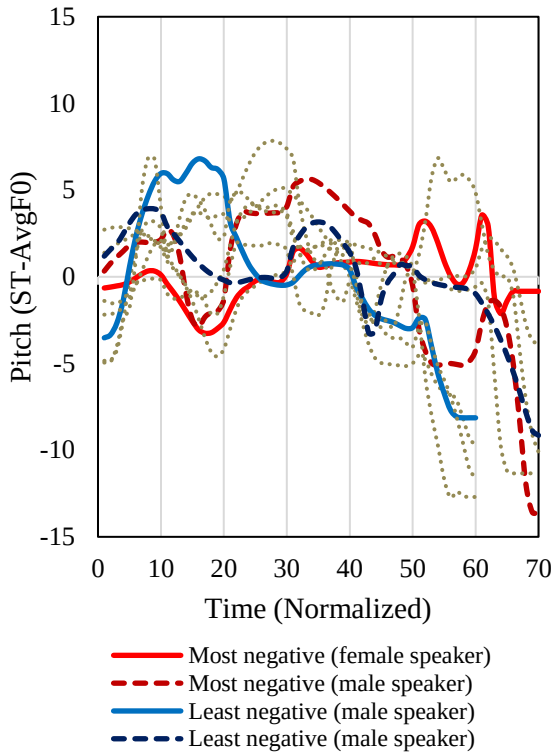
This work therefore is twofold. Firstly, it aims to present a first look at understanding the intonation produced by Singaporean actors in their portrayals of anger and happiness. Secondly, it aims to consider how SgE speakers of Chinese, Malay, and Indian ethnicity may perceive different emotions expressed from these intonations. An auditory perception test was conducted online using 26 stimuli (10 anger, 10 happiness, 6 neutral) acted by 12 Chinese Singaporeans from the VENEC corpus [13]. The corpus used two semantically neutral sentences of different word lengths as scripts. Respondents listened to each stimulus then rated the level of valence, arousal, and dominance they perceived based on a 9-point Likert Scale, as well as labelled the emotions in a Choose-All-That-Apply format.

256 eligible responses (167 Chinese, 48 Malay, 41 Indian) were analyzed for effects of ethnicity on the perception of anger and happiness in speech. The results so far showed that respondents, regardless of ethnicity, rated and labelled the anger and happiness portrayals similarly. Overall, anger portrayals were rated accurately for negative valence and respondents converged on what they perceived to be prototypical anger portrayal, with 60% rating it as most negative in terms of valence and 76% labelled it as "anger". Recordings that were rated not as negatively for valence were also the same across the ethnic groups. Similar rating patterns were observed for happiness portrayals across all the ethnic groups, with 40% of the respondents rating the same recording as most positive. Figures 1 and 2 show the pitch contours for anger and happiness portrayals respectively. F0 measurements were obtained in semitones using ProsodyPro [14] on Praat [15] based a 10 time-step for each annotated interval to normalize the duration of each recording. Obstruents were excluded from the measurements to prevent interferences with the pitch tracking [16]. The ST-AvgF0 method in [17] was used to normalize sex-related pitch variation in relation to each speaker's average pitch in each recording. Further analyses comparing the intonational qualities along with intensity and duration are currently in progress. The present results support the finding of a large pitch range in angry and happy speech. However, steep rising contours can also be observed for recordings that had congruent ratings. These results have

Proceedings of the Second International Conference on Tone and Intonation, pages 74–76
November 18–20, 2023.

implications for understanding how SgE intonation varies and interact with emotional prosody as well as the relevant cues that SgE speakers use to distinguish emotional meanings from the other levels of meaning that intonation encapsulates.

(Fig. 1) Anger portrayals



(Fig. 2) Happiness portrayals

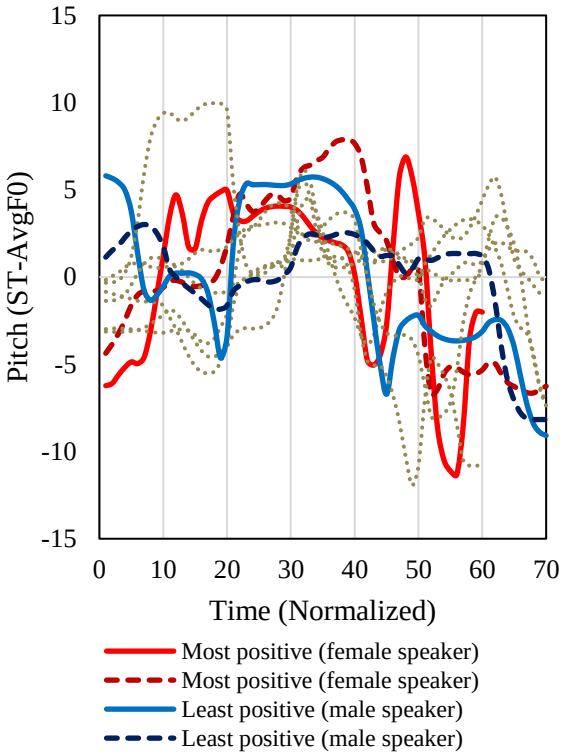


Figure 1: *Pitch contours for anger portrayals (Normalized F0 in semitones. Contours in red: recordings that scored most negative in valence; in blue: recordings that scored least negative in valence);* **Figure 2**: *Pitch contours for happiness portrayals (Normalized F0 in semitones. Contours in red: recordings that scored most positive in valence; in blue: recordings that scored least positive in valence)*

References

[1] Nolan, F. 2006. Intonation. In Aarts, B. & McMahon, A. (Eds.), *The Handbook of English Linguistics*. John Wiley & Sons, Ltd, 433-457.

[2] Warren, P., & Calhoun, S. 2021. Intonation. In Knight, R.-A & Setter, J. (Eds.), *The Cambridge Handbook of Phonetics*. Cambridge University Press, 209-236.

[3] Banse, R., & Scherer, K. R. 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology 70*(3), 614–636.

[4] Bänziger, T., & Scherer, K. R. 2005. The role of intonation in emotional expressions. *Speech Communication 46*(3–4), 252–267.

[5] Féry, C. 2017. *Intonation and prosodic structure*. Cambridge University Press.

[6] Matsumoto, D., Seung Hee Yoo, & Fontaine, J. 2008. Mapping Expressive Differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology 39*(1), 55–74.

[7] Hurley, C. M., Teo, W. J., Kwok, J., Seet, T., Peralta, E., & Chia, S. Y. 2016. Diversity from within: The impact of cultural variables on emotion expressivity in Singapore. *International Journal of Psychological Studies 8*(3), 50–64.

[8] Low, E.L., & Grabe, E. 1999. A contrastive study of prosody and lexical stress placement in Singapore English and British English. *Language and Speech 42*(1), 39–56.

[9] Lim L. 2004. *Singapore English: A grammatical description.* Amsterdam: John Benjamins Pubs.

[10] Goh, C.C.M. 2005. Discourse intonation variants in the speech of educated Singaporeans. In Deterding, D., Brown, A., & Low, E.L. (Eds.), *English in Singapore: Phonetic Research on a Corpus.* McGraw-Hill Education (Asia), 104-114.
[11] Deterding, D. 2007. *Singapore English.* Edinburgh: Edinburgh University Press.
[12] Low, E.L. 2010. Sounding local and going global: Current research and implications for pronunciation teaching. In Lim, L., Pakir, A. & Wee, L. (Eds.), *English in Singapore: Modernity and Management*. Hong Kong: Hong Kong University Press.
[13] Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. 2016. The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology 111*(5), 686–705.
[14] Xu, Y. 2013. ProsodyPro — A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody 2013* (Aix-en-Provence, France), 7-10.
[15] Boersma, P., & Weenink, D. 2008. *Praat: doing phonetics by computer*. Computer program.
[16] Beňuš, Š. 2021. *Investigating spoken English.* Cham: Palgrave Macmillan Cham.
[17] Zhang, J.W. 2018. A Comparison of Tone Normalization Methods for Language Variation Research. *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation* (Hong Kong), 823-831. Association for Computational Linguistics.

Prosodic Phrasing in Breton

Emily Elfner[1], Francesc Torres-Tamarit[2] & Mélanie Jouitteau[3]
[1]*York University,* [2]*Universitat Autònoma de Barcelona,* [3]*UMR 5478, IKER, CNRS, Université Bordeaux Montaigne, Université de Pau et des Pays de l'Adour*
eelfner@yorku.ca, francescjosep.torres@uab.cat,
melanie.jouitteau@iker.cnrs.fr

This paper provides a preliminary phonological analysis of prosodic phrasing in Breton (*brezhoneg*), an endangered Celtic language spoken in Brittany, France. Unlike other Celtic languages, which show a predominantly VSO word order, Breton has a canonical verb-second (V2) word order but also allows verb-initial word order in certain contexts [1]. In this study, we investigate prosodic phrasing in declarative sentences with a variety of verb-initial and verb-second word orders. In addition, we systematically vary the length of the subject and object (noun vs. noun-adjective), in order to account for possible effects of binarity on prosodic phrasing. In this preliminary report, we describe the patterns of prosodic phrasing and the distribution of an LH* pitch accent for two native speakers of Breton, HG and JS, and consider the relationship between syntactic structure and prosodic phrasing.

The two speakers, HG, and JS, are familiar to the third author who conducted all fieldwork. Both speakers are over 70 years old and are experienced in the exercise of elicitation and reading Breton. They each speak two Breton varieties: their native local Breton and a standardized variety. The two traditional varieties represented, East-Kerne for HG and Treger for JS, are fairly close dialects which are part of the set of dialects known as KLT (Kerne, Leon, Treger). Both speakers were asked to rewrite the sentences given to them in their local variety, then record the sentences as naturally as possible.

Each of the two speakers produced at least three repetitions of 14 possible sentences, including 8 tokens of the two transitive sentence types (Table 1) and 6 tokens of the four intransitive sentence types (Table 2). The sentences were elicited in a context designed to trigger a neutral (broad focus, all-new) information reading; any sentences with a clear narrow focus on one of the lexical items were not included in the present analysis (see [2] for further discussion of semantic focus for subjects). The F0 contours of the recordings were analysed qualitatively using Praat [3] and annotated by the first two authors, comparing the two speakers, assuming the autosegmental-metrical model of intonation [4, 5]. No quantitative analysis was performed at this stage.

The annotated speech of both speakers showed a regular pattern of rising pitch accents (LH*) on lexical words (i.e. verbs, nouns and adjectives). In terms of timing, the peak of the rising accent occurs on the stressed (penultimate) syllable (see Figure 1). For SVO transitive sentences (Type 1), we observe that pitch accents occur on each lexical word (S, V, and O) if S and O are short (i.e. a bare noun). If the noun is modified by an adjective, then only the rightmost element in the phrase receives a pitch accent, i.e. (N_{LH*}) vs. (NA_{LH*}). Auxiliary verbs (Type 2) were not marked with an LH* pitch accent. In the three sentence types with copula verbs (Types 3-5), the copula was consistently marked with an LH* pitch accent by both speakers only when it was found in sentence-initial position (Type 4), regardless of the length of the subject. In Types 3 and 5, where the copula occurred following either the subject or the verb, we found some variation between the two speakers: JS consistently left the copula unaccented in sentence-medial position, while HG variably accented the copula, but only if the subject was non-binary (i.e. an unmodified noun).

We propose that the LH* pitch accent marks the right edge of a phonological phrase, and is placed on the stressed syllable of the prosodic word (ω) closest to the right edge of this phrase. In most cases, each of the main components of the sentence (V, S, and O) each receive a pitch accent, suggesting that each ω is parsed as its own phonological phrase (φ) (see Tables 1 and 2). In Breton, lexical (prosodic) words are marked with a single primary stress, so when a syntactic phrase includes two lexical words, only the stressed syllable in the rightmost word will be marked with an LH* pitch accent. Auxiliary verbs and copulas differ from lexical verbs by not being parsed as ωs in neutral (broad focus) sentences, and thus do not contain a primary stressed syllable capable of bearing the LH* pitch accent. Thus, when the auxiliary or copula verb is in non-initial position in the sentence, it is preferentially encliticized onto the preceding lexical item, whether this is a verb or a subject (N or NA). In Type 4 sentences, where the copula occurs in sentence-initial position, it cannot be encliticized onto a preceding lexical word (since it is utterance-initial), and thus is pronounced as a full ω and is marked with an LH*

Proceedings of the Second International Conference on Tone and Intonation, pages 77–78
November 18–20, 2023.

pitch accent, indicating that it is also parsed as its own φ. Note that the copula takes the form *zo* in second position after a noun (Type 3) but *emañ* in Types 4 and 5.

Future work will further investigate boundary tones (both medial and final) and the interpolation/scaling between LH* pitch accents in order to provide a more complete picture of intonational marking in Breton.

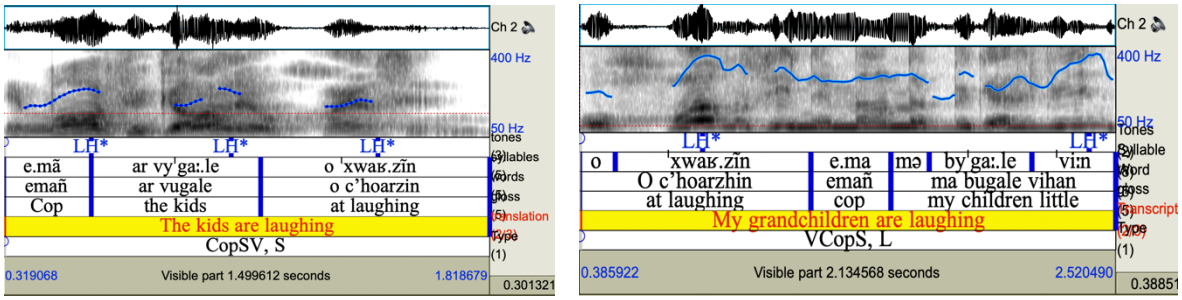


Figure 1: *F0 contours for JS (Type 4a) and HG (Type 5b)*

Table 1: *Transitive sentence types (S=short, L=long).*

Type 1	Sentence	Prosodic Phrasing	Type 2	Sentence	Prosodic Phrasing
SVO, SS	[N]V[N]	(N)(V)(N)	VAuxSO, SS	VAux[N][N]	(VAux)(N)(N)
SVO, LS	[NA]V[N]	(NA)(V)(N)	VAuxSO, LS	VAux[NA][N]	(VAux)(NA)(N)
SVO, SL	[N]V[NA]	(N)(V)(NA)	VAuxSO, SL	VAux[N][NA]	(VAux)(N)(NA)
SVO, LL	[NA]V[NA]	(NA)(V)(NA)	VAuxSO, LL	VAux[NA][NA]	(VAux)(NA)(NA)

Table 2: *Intransitive sentence types (S=short, L=long).*

	Type	Sentence	Prosodic Phrasing
Type 3	SCopV, S	[N]CopV	(NCop)(V) ~ (N)(Cop)(V)
	SCopV, L	[NA]CopV	(NACop)(V)
Type 4	CopSV, S	Cop[N]V	(Cop)(N)(V)
	CopSV, L	Cop[NA]V	(Cop)(NA)(V)
Type 5	VCopS, S	VCop[N]	(VCop)(N) ~ (V)(Cop)(N)
	VCopS, L	VCop[NA]	(VCop)(NA)

References

[1] Jouitteau, M. and F. Torres-Tamarit. 2023. The syntax of Modern Breton. In J. Eska, et al. (eds.) *The Handbook of Celtic Languages*. Palgrave.

[2] Jouitteau, M. 2007. The Brythonic reconciliation: From V1 to generalized V2. In J. van Craenenbroeck and J. Rooryck (eds.) *The Linguistics Variation Yearbook 7*. The Netherlands, 163-200.

[3] Boersma, P. and D. Weenink. 2022. Praat: doing phonetics by computer (Version 6.2.17 [Computer program]. http://www.praat.org/. Date accessed: August 24, 2022.

[4] Pierrehumbert, J. 1980. The Phonology and Phonetics of English Intonation. Doctoral dissertation, MIT.

[5] Ladd, D.R. 2008 [1996]. *Intonational Phonology*. Cambridge: Cambridge University Press.

Phonetic corelates of syllable prominence in Mundari

Luke Horo, Pamir Gogoi & Gregory D.S. Anderson
Living Tongues Institute for Endangered Languages
luke.horo@livingtongues.org, pamir.gogoi11@gmail.com,
gdsa@livingtongues.org

The objective of this study is to examine phonetic prominence in Mundari disyllables and polysyllables with the help of acoustic analysis. Mundari is an Austroasiatic language spoken by approximately two million people in India. Previous studies on Mundari have conflicting views on the topic. According to Cook (1965: 100), Langendoen (1963: 14-15), and Sinha (1975: 39), Mundari is a stress language, while Osada (1992: 36) considers it to be a pitch accent language. Moreover, Cook (1965) argues that if the final syllable is closed it is accented, otherwise it is the initial syllable in disyllabic words, proposing a quantity sensitive trochaic system. Likewise, Hoffmann (2001: 59) claims that in disyllabic words the accent is on the first syllable, with (lexical) exceptions. On the other hand, Sinha (1975) claims Mundari stresses the second syllable in disyllabic words if it is of the shape of C1V1C2V2 or C1V1C2V2C3 but in words of the shape C1V1C2C3V2, stress falls on the initial syllable, suggesting a quantity sensitive iambic system. Also, according to Sinha, if the word is trisyllabic, stress falls on the 2nd syllable regardless of the shape. Similarly, Osada (2008: 104) 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. Additionally, Osada (1992: 34) states that "in Mundari a phonological word maximally consists of three syllables". However, these previous studies are impressionistic, and they do not provide experimental data to verify the claims. Hence, this study describes our initial findings from an ongoing study of intonation in Mundari. Here we have analyzed three acoustic cues of phonetic prominence, namely vowel duration, vowel intensity and fundamental frequency, using Mundari disyllabic forms of any function and polysyllabic nouns and verbs that are inflected for a variety of case, possession, number, tense, aspect etc. categories. The study is based on Mundari speech data recorded in the field from female and male speakers as they produced the target forms in (i) isolation (ii) a carrier phrase (iii) an out of focus frame and (iv) an exclusive focal frame. Preliminary findings suggest that longer vowel duration is found in utterance final position and therefore, is not a reliable cue for identifying word prominence in Mundari. Likewise, vowel intensity does not exhibit a consistent pattern to indicate prominence in Mundari disyllables nor in polysyllables. Intriguingly, fundamental frequency measurement reveals a gender-based prominence system in Mundari. While f0 is never observed to be high in the initial syllable, f0 peaks are found differently realized in male and female speakers of Mundari. In case of female speakers, f0 peak is observed in the second syllable and in case of male speakers, f0 peak is observed in the final syllable.

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, T. 1963. *Mundari Phonology*. Unpublished manuscript. Cambridge,
[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.

Setting the "tone" first and integrating into the syllable later:

An EEG study of lexical tonal encoding in Mandarin word production

Xiaocong Chen[1], Caicai Zhang[1]
[1]*Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University*
xiaocong.chen@polyu.edu.hk; caicai.zhang@polyu.edu.hk

Lexical tones, which are pitch information used to distinguish lexical meanings, are an important phonological property in tonal languages. However, it remains hotly debated how lexical tones are planned in the phonological encoding stage of word production. Whereas some researchers proposed that lexical tones are independently retrieved and encoded in ways similar to segments in production [1, 2], other researchers proposed that lexical tones are encoded like metrical stress in non-tone languages [3, 4, 5]. Moreover, the relative encoding timing of lexical tones is also unclear. Some researchers suggested that lexical tones are encoded at the early stage together with syllabic information [5, 6], whereas other researchers indicated that lexical tones may be encoded at a later stage [7, 8]. To address these issues, we employed a phonologically-primed picture naming task and utilized the high temporal resolution of electroencephalography (EEG) to investigate the encoding process of lexical tones in overt Mandarin Chinese word production.

Forty native Mandarin speakers (21 males) from North China were asked to produce the disyllabic names of 104 target pictures and 52 filler pictures while their oral responses and EEG signals were recorded. Each picture was preceded by a monosyllabic visual prime, presented together with its auditory form. We manipulated the tonal relatedness and syllabic relatedness of the primes in relation to the initial morpheme of the target picture names (e.g., 鹦鹉, *ying1-wu3*, 'parrot'), resulting in four prime conditions: (1) homophone prime (e.g., 英, *ying1*, sharing the syllable and tone); (2) syllable-overlap prime (e.g., 营, *ying2*, only sharing the syllable); (3) tone-overlap prime (e.g., 纲, *gang1*, only sharing the tone); (4) unrelated prime (e.g., 悖, *bei4*, phonologically unrelated). The behavioral results (Fig.1A) revealed that there was a significant interaction between tonal and syllabic relatedness. The syllable-related primes (homophone and syllable-overlap prime) yielded significantly shorter naming latencies than the other syllable-unrelated primes. Moreover, the homophone prime yielded significantly shorter onset latencies than the syllable-overlap prime conditions, but the tone-overlap prime exhibited significantly longer onset latencies than the unrelated prime. This suggested that additional tonal overlap facilitated the production only when the syllabic information could be prepared but hampered the production when the syllabic information was not readily prepared. Regarding the EEG data, the analysis of the stimulus-locked ERP (i.e., ERP time-locked to the picture onset; Fig.1B) only revealed there was an early tonal relatedness main effect in the 250~350 ms time window (with more negativity in left frontocentral regions for tonal related primes than tonally unrelated primes), and a later syllabic relatedness main effect in the 350~500 ms time window (with larger negativity in left frontocentral regions but more positivtiy in right frontal region and bilateral posterior regions for syllable-related primes than syllable-unrelated primes). In contrast, analysis of the response-locked ERP (i.e., ERP time-locked to the speech acoustic onset; Fig.1C) revealed that an early syllable-relatedness effect in the -450~-250 ms time window (with more negativity in frontocentral regions but more positivtiy in posterior regions for syllable-related primes than syllable-unrelated primes), but a significant interaction between tonal and syllabic relatedness in the -250~-100 ms time window. Further analysis showed that the homophone prime elicited larger negativity in left frontocentral regions than the syllable-overlap prime, but no difference was observed between the tone-overlap prime and the unrelated prime. These results indicate that lexical tones can be independently retrieved in a eariler time window than the syllable at the phonological encoding stage, but need to be planned and integrated with the syllable at the later phonetic encoding stage.

Proceedings of the Second International Conference on Tone and Intonation, pages 80–82
November 18–20, 2023.

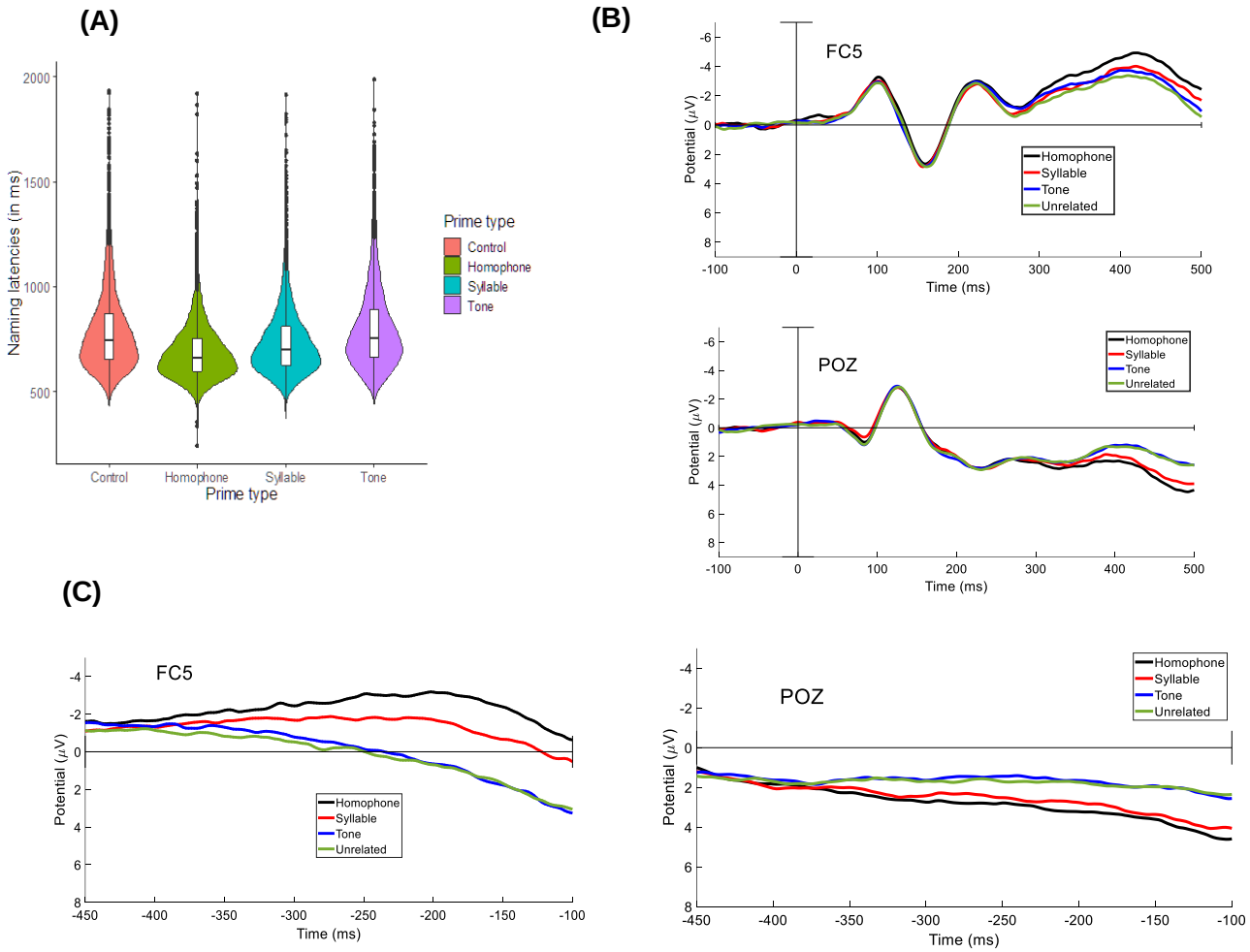


Fig. 1 (A) Mean RTs (in ms) for different prime types; (B) Stimulus-locked average ERPs in representative electrodes (FC5 and POZ) and (C) Response-locked average ERPs in representative electrodes (FC5 and POZ).

References:

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

[2] Alderete, J., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition*, 191, 103952.

[3] Chen, J. Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics. 20*(2), 289–301.

[4] 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*(4), 751–781.

[5] Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research*, *57*, 22–37.

[6] Zhou, X., & Zhuang, J. (2000). Lexical tone in the speech production of Chinese words. *6th* ICSLP.

[7] Zhang, Q., & Damian, M. F. (2009). The time course of segment and tone encoding in Chinese spoken production: An event-related potential study. *Neuroscience*, *163*(1), 252–265.

[8] Zhang, Q., & Zhu, X. (2011). The temporal and spatial features of segmental and suprasegmental encoding during implicit picture naming: An event-related potential study. *Neuropsychologia*, *49*(14), 3813–3825.

Tone in Binumarien Loans: Incorporating Tok Pisin Words in a Kainantu (Papuan) Tonal System

Renger van Dasselaar
PhD candidate in Linguistics, Philology and Phonetics, University of Oxford
renger.vandasselaar@stx.ox.ac.uk

This abstract explores the adoption of Tok Pisin loanwords (of English origin) into the Binumarien language. Binumarien is characterized by its tonal system and the absence of lexical stress. In contrast, Tok Pisin is a non-tonal language presumed to have word stress. This study investigates how Binumarien assigns tone to Tok Pisin loanwords, emphasising the role of syllable structure rather than stress, which is remarkable considering the minor influence of syllable structure on tone in other Binumarien words.

Binumarien is spoken by approximately 1200 people in the Eastern Highland Province, Papua New Guinea. Binumarien features a two-level system, comprising L (low) and H (high) tones. Tones are assigned at the mora level, where moras can be classified as L, H, or toneless. Toneless segments are realized as either L or H, depending on the surrounding context. Binumarien does not seem to exhibit lexical stress (van Dasselaar, 2019). While certain syllable structures in Binumarien words appear to align with specific tonal patterns, my ongoing analysis suggests no one-on-one correspondence. Moreover, there appears to be no link between metricality and tone.

Previous research by Wurm (1985) on Eastern Highlands Tok Pisin suggests that Tok Pisin exhibits lexical stress, primarily marked by higher pitch. Stress typically falls on the first syllable, but there are exceptions. Wurm also notes that the distinction between stressed and unstressed syllables in Tok Pisin is less prominent compared to English. He also observes that considerable variation exists depending on region and familiarity with English. Moreover, Faraclas (1989), who examined Tok Pisin stress in East Sepik, found variations based on the speaker's gender but surprisingly not on the substrate language.

The Tok Pisin loanwords are from a recently collected corpus and originate from English. The analysis compares the stress and syllable structure of these loanwords in Tok Pisin with their tonal assignment in Binumarien. The processing of data was not yet fully completed when this abstract was written, but several noteworthy observations emerge:

1. Tone assignment is primarily based on syllable structure rather than the original metrical structure in Tok Pisin.
2. The original Tok Pisin metrical structure has minimal influence on tonal assignment in Binumarien. This contrasts with findings in other languages, for example for English loans into Mandarin (Glewwe, 2021).
3. Most loanwords exhibit a floating L tone.
4. In contrast to other Binumarien words, /st/ onsets receive tone, while consonants in other Binumarien words never carry tone.

For instance, CVCV words like "kara" (car) are assigned the HH(L) tonal pattern. Similarly, CVVCV words such as "koofi" (coffee) receive the LLH pattern, unless the final vowel is not present in Tok Pisin, as observed in "beeta" (bed), where the tone pattern is HHH(L). CVCVV words like "gitaa" (guitar) exhibit the LHL tonal pattern. The onset in e.g. *stoora* ‘story’ is assigned tone: LHHH, where L is assigned to /st/.

This study sheds light on the process of tone incorporation in Binumarien, particularly in the context of adopting Tok Pisin loanwords. The findings indicate that syllable structure plays a significant role in tone assignment, overriding the influence of stress and metrical structure. These results contribute to our understanding of the interaction between tonal and

Proceedings of the Second International Conference on Tone and Intonation, pages 83–84
November 18–20, 2023.

non-tonal languages, highlighting the characteristics of Binumarien's tonal system in accommodating loanwords from Tok Pisin.

Faraclas, M. 1989. ‘Prosody and creolization in Tok Pisin.’ *Journal of Pidgin and Creole Languages* 4: 132-139.

Glewwe, E. 2021. ‘The phonological determinants of tone in English loanwords in Mandarin.’ *Phonology* 38 (2): 203-239.

van Dasselaar, R. 2019. Topics in the Grammar of Binumarien: Tone and Switch Reference in a Kainantu Language of Papua New Guinea. Thesis.

Wurm, S. A. 1985). ‘Phonology: Intonation in Tok Pisin.’ In S. A. Wurm & P. Mühlhäusler, *Handbook of Tok Pisin (New Guinea Pidgin),* 309-334. (Pacific Linguistics C70.) Canberra: Australian National University.

Exploring intonational patterns of poetic speech: Insights from a large corpus of German poetry

Nadja Schauffler[1], Nora Ketschik[1], Kerstin Jung[1], André Blessing[1], Julia Koch[1], Toni Bernhart[1], Anna Kinder[2], Jonas Kuhn[1], Sandra Richter[1,2], Rebecca Sturm[2], Gabriel Viehhauser[1], Thang Vu[1]
[1]*University of Stuttgart*, [2]*German Literature Archive Marbach*
nadja.schauffler@ims.uni-stuttgart.de

This study investigates the intonational patterns of poetic speech in a large corpus of German poetry recitations. Previous studies have identified a number of intonational features that are typical for verse recitation in English, and modeled them in what they called "a formula for poetic intonation" [1,2]. Byers identified features such as a slow speech rate, short intonation units, more pauses, intonation units of relatively equal length, low average pitch, a narrow pitch range, simple falling melodies, and simple falling nuclear tones; Barney, building upon this formula, differentiated between general performance features and specific poetic characteristics and added echoes of pitch patterns to the latter (which was also reported by [3]). While these studies were limited to a small number of speakers and poems, advancements in computational methods now enable us to analyze larger corpora.

In our study, we investigate features of poetic intonation in a corpus comprising recitations of German poems collected within the project »textklang«.[1] We are interested in how prosodic features associated with poetry, such as speech rate and pitch range, may vary over time and differ depending on aspects such as length of poem or authorship. This investigation serves as the basis for refining our poetic speech synthesis models [4] to incorporate specific styles, speakers, authors, or epochs (allowing us to synthesize a recitation to sound, for example, like a Goethe poem read by a female speaker in the 1950s). Additionally, these models will be used in perception studies to explore the aesthetic effects and functions of the investigated features.

Our analyses encompass 1148 recitations of 682 German poems from the Romantic period, spoken by 120 different speakers (33 female) between 1951 and 2020. The corpus was automatically annotated at the level of both the written material (poem) and speech data (recitation) [5]. The annotations follow the GRAIN pipeline [6] combined with manual revisions for text-speech alignment. Pauses, pitch and duration values were extracted by means of the synthesis system Festival (University of Stuttgart version [7]). Intonation events (pitch accents and boundary tones) were automatically annotated as described in [8]. This extensive corpus makes it possible to revisit the concept of poetic intonation from a macro-analytic perspective.

For the study at hand, we calculated articulation rate as the number of syllables per second (excluding pauses), and recitation rate as the number of syllables per second over the length of the respective poem (including pauses). Pitch range was calculated over individual stanzas as the difference between the highest and lowest pitch in the middle of the respective syllable.

First results reveal significant variation within the investigated features that is highly dependent on the speaker. Nevertheless, we also identify speaker independent effects indicating diachronic developments in recitation patterns. Recitation rate appears to change over time and is significantly lower in the 1980s and 1990s (Fig. 1). Also pitch range differs in certain time periods, particularly between the 1960s (decrease compared to the 1950s) and the 1980s (increase, Fig. 2). Interestingly, these two trends are also found in the data of an individual speaker in our corpus (Gert Westphal) from whom we have multiple recitations spanning from the 1960s to the 1990s. In terms of formal aspects, we find that longer poems have a higher recitation rate (Fig. 3). Comparing a subset of the six most frequent authors in our corpus, we observe significant differences in recitation rates, with some authors (e.g. Schiller) being read at significantly higher rates than others (e.g. Goethe, Fig. 4).

These preliminary results show how a data-driven approach can provide insights into specific intonation features over time, based on the recited material and the speaker. Identifying such trends brings us closer to defining and investigating different recitation styles and their perception. In our talk, we will share additional findings on intonation features from an expanded analysis, including aspects like pause frequency. Furthermore, we will take a closer look at factors such as the speaker's gender and, based on initial manual annotations, how the poem was recited in terms of emphasis and metrics.

[1] The interdisciplinary project »textklang« develops a mixed-methods approach for the systematic investigation of the relationship between written text and its sonic realization, cf. https://textklang.org/.

Proceedings of the Second International Conference on Tone and Intonation, pages 85–86
November 18–20, 2023.

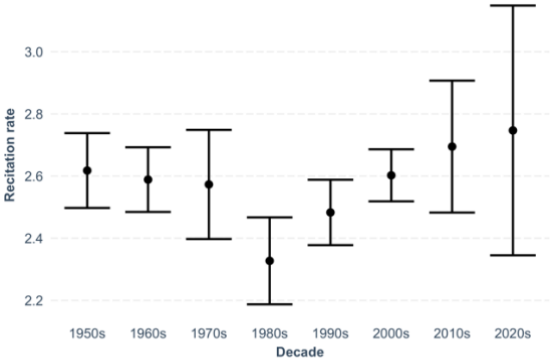


Figure 1: Effect of Decade on recitation rate (syllables per second over recitation, including pauses).

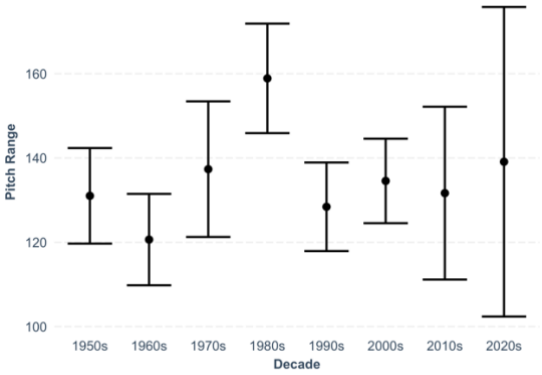


Figure 2: Effect of Decade on Pitch Range (over individual stanzas).

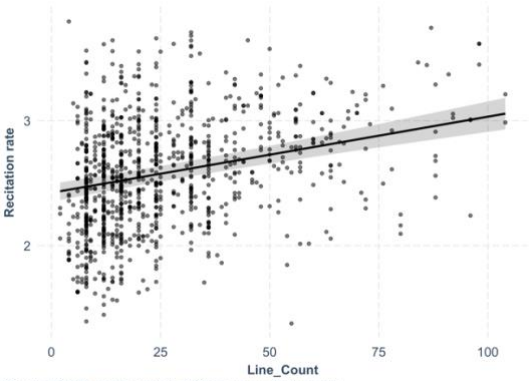


Figure 3: Effect of number of lines on recitation rate.

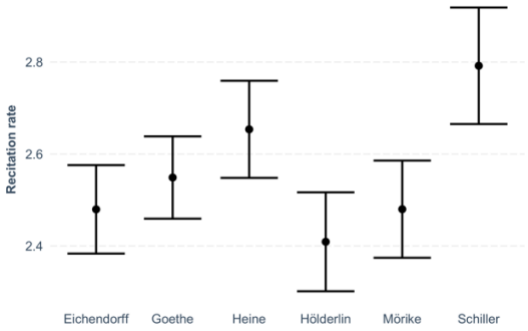


Figure 4: Effect of authorship on recitation rate in a subset with the six most frequent authors.

References

[1] Barney, T. 1999. Readers as Text Processors and Performers: A New Formula for Poetic Intonation. In *Discourse Processes 28 (2)*, 155-167. DOI: 10.1080/01638539909545078.

[2] Byers, P. 1979. A Formula for Poetic Intonation. In *Poetic 8*, 367-380.

[3] Menninghaus W., Wagner V., Knoop C. A., Scharinger, M. 2018. Poetic speech melody: A crucial link between music and language. PLoS ONE 13(11): e0205980. https://doi.org/10.1371/journal.pone.0205980

[4] Koch, J., F. Lux, N. Schauffler, T. Bernhart, F. Dieterle, J. Kuhn, S. Richter, G. Viehhauser, & T. Vu. 2002. PoeticTTS - Controllable Poetry Reading for Literary Studies. In *Proceedings of Interspeech 2022*. DOI: 10.48550/arXiv.2207.05549.

[5] Schauffler, N., T. Bernhart, A. Blessing, G. Eschenbach, M. Gärtner, K. Jung, A. Kinder, J. Koch, S. Richter, G. Viehhauser, T. Vu, L. Wesemann, & J. Kuhn. 2022. »textklang« - Towards a Multi-Modal Exploration Platform for German Poetry. In *Proceedings of the 13th edition of the Language Resources and Evaluation Conference (LREC)*, 5345-5355.

[6] Schweitzer, K., K. Eckart, M. Gärtner, A. Falenska, A. Riester, I. Roesiger, A. Schweitzer, S. Stehwien, & J. Kuhn. 2018. German radio interviews: The GRAIN release of the SFB732 silver standard collection. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*, 2887-2895.

[7] Festival. 2010. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. IMS German Festival home page. www.ims.uni-stuttgart.de/phonetik/synthesis.

[8] Schweitzer, A. 2010. *Production and Perception of Prosodic Events – Evidence from Corpus-based Experiments*. Doctoral dissertation, Universität Stuttgart.

Clustering lexical tones with intonation variation

Katrina Kechun Li[1], Francis Nolan[1] & Brechtje Post[1]
[1]*University of Cambridge*
kl502@cam.ac.uk, fjn1@cam.ac.uk, bmbp2@cam.ac.uk

Intonation and tones are interwoven in tone languages, both conveyed through the fundamental frequency (f0). A prevailing assumption in the previous literature is that the preservation of tonal categories is prioritised over any intonation manipulation that conveys pragmatic functions. As a result, these studies tend to adopt a 'top-down' approach, examining how intonation modifies the f0 contour for each tonal category (see [1] and references therein). In this paper, we present a 'bottom-up' analysis, using a contour clustering technique, to investigate how tonal contours are grouped into categories based solely on f0, without prior knowledge of tonal categories.

Our study utilises data from 14 Cantonese speakers who read sentences under different prosodic focus conditions. The sentences adhere to a Subject-Verb-Object structure, with the subject comprising a name prefix /a/ followed by a syllable representing personal names. These name syllables, our analysis targets, encompass the full inventory of Cantonese tones (including checked tones T7-T9), and are either focused or unfocused. We extracted f0 data from 10 equidistant points of the rhyme part of the syllable using Praat, and identified tracking errors using the method proposed in [2]. F0 was converted to semitones with each participant's average f0 as their base. We employed a hierarchical agglomerative clustering technique implemented in the R application [3] with the complete linkage criterion. The optimal number of clusters was determined by minimising the within-cluster variance while maximising the between-cluster variance based on Euclidean distances.

After removing outlier contours, we obtained 863 unlabelled f0 contours (93.5% of the data), which were subsequently grouped into four clusters (Figure 1). By mapping predefined tonal categories to these clusters (Figure 2), we observed that the clustering primarily reflected tonal register. The majority of two high tones (high-level T1 and high-checked T7) were grouped together as Cluster 3, while most of the two rising tones (high-rising T2 and low-rising T5) were placed in Cluster 4. The low register tones were collapsed into Cluster 2, including low-falling T4, low-level T6 and low-checked T9. Cluster 1 appeared to represent the mid-register tones, encompassing a substantial portion of mid-level T3 and mid-checked T8, although many T3 and T8 contours also fell into the low register cluster.

Our findings concur with a 'top-down' analysis using GAMM modelling, which shows that focus marginally influences f0 contour in Cantonese [4]. However, focus does seem to improve the consistency of the clustering. For instance, when focused, T2 was classified into Cluster 4 in 77% of the cases, in contrast to 58% when unfocused. This supports the view that focus induces hyper-articulation of the tonal target and thereby enhancing tonal contrasts [5]. Interestingly, increasing the number of clusters does not further contribute to the distinction of the intonation or tonal categories within each cluster. This suggests that specific tone pairs can be challenging to discern due to their greater surface similarities in continuous speech, which is also evidenced by the ongoing tone mergers of these pairs (T2/T5, T4/T6, T3/T6) [6].

In summary, this study supports the view that lexical tonal contrast is largely maintained across different intonation conditions. Nonetheless, the emergent clusters are unlikely to fully distinguish all lexical tones when relying solely on f0, due to the influence of both tonal similarities and intonation variations. Our next step involves applying the clustering technique to two other languages in our dataset, Chengdu and Changsha. In these languages, intonation might lead to more pronounced variations in tonal targets. Our study will therefore reveal how a bottom-up approach using contour clustering can provide further insight into the interaction of tone and intonation across tone languages.

Proceedings of the Second International Conference on Tone and Intonation, pages 87–88

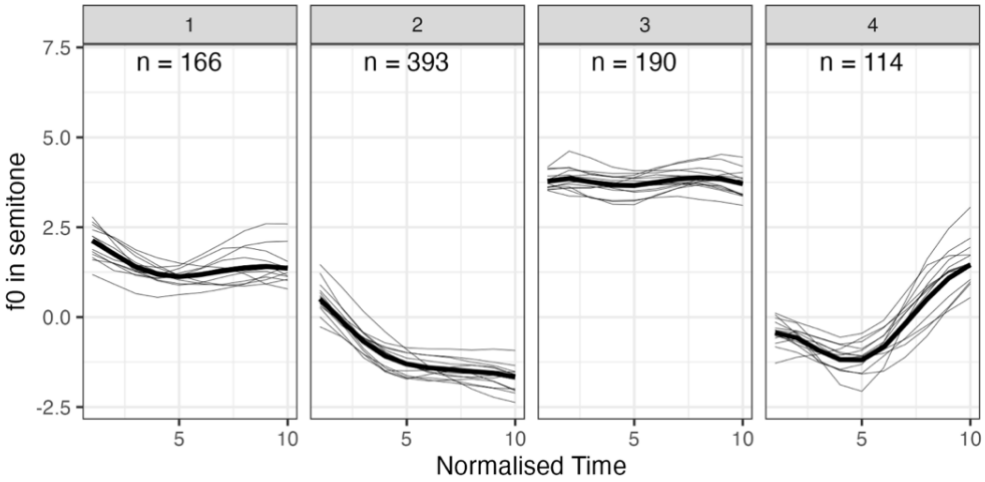


Figure 1: Cluster means (dark lines) and the contributing f0 contour (light lines) of each cluster.

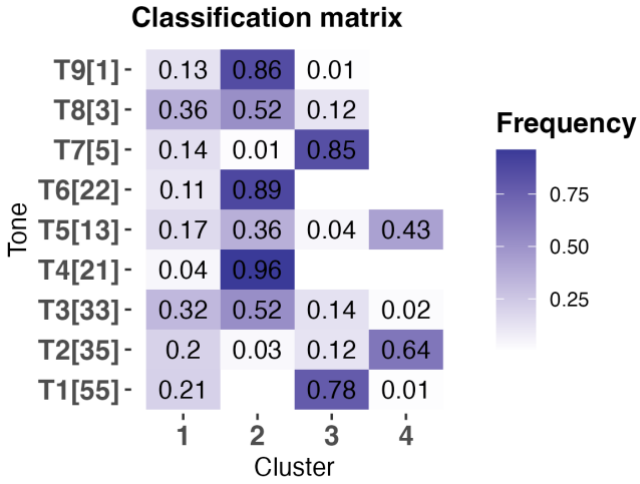


Figure 2: The proportion of each tone that contributes to each cluster (rows indicates lexical tones represented using Chao number, 1-lowest, 5-highest). Coloration indicates the frequency of tones in each cluster, with numeric label.

References

[1] Y. Chen, 'Tone and Intonation', in *The Cambridge Handbook of Chinese Linguistics*, C.-R. Huang, I.-H. Chen, Y.-H. Lin, and Y.-Y. Hsu, Eds., in Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, 2022, pp. 336–360. doi: 10.1017/9781108329019.019.

[2] J. Steffman and J. Cole, 'An automated method for detecting F measurement jumps based on sample-to-sample differences', *JASA Express Lett.*, vol. 2, no. 11, p. 115201, Nov. 2022, doi: 10.1121/10.0015045.

[3] C. Kaland, 'Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours', *J. Int. Phon. Assoc.*, pp. 1–30, Apr. 2021, doi: 10.1017/S0025100321000049.

[4] K. K. Li, F. Nolan, and B. Post, 'Variations of focus prominence in three tone languages', *Proceedings of the 20th International Congress of Phonetic Sciences*, pp.1420-24, Aug. 2023.

[5] Y. Chen, 'Post-focus F0 compression—Now you see it, now you don't', *J. Phon.*, vol. 38, no. 4, pp. 517–525, Oct. 2010, doi: 10.1016/j.wocn.2010.06.004.

[6] R. S. Y. Fung and C. K. C. Lee, 'Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception', *J. Acoust. Soc. Am.*, vol. 146, no. 5, pp. EL424–EL430, Nov. 2019, doi: 10.1121/1.5133661.

Variable Pitch Accent and Prosodic Phrasing in Japanese Adjectival Complex DPs

Le Xuan Chan[1], Rina Furusawa[2], Seunghun J. Lee[2,3]
[1]*National University of Singapore,* [2]*International Christian University,* [3]*IIT Guwahati*
lxlinguistics@gmail.com, furusawaling@gmail.com, seunghun@icu.ac.jp

Background The variable pitch accent of adjectives in Tokyo Japanese has been well documented in literature [1-3], whereby there is an increasing trend of accented realizations (accentuation) of unaccented adjectives becoming more acceptable. In particular, such accentuation is more acceptable at the phrase-final position than when directly modifying a following noun. These patterns suggest that pitch accent is sensitive to prosodic phrasing. Though much has been discussed about how prosodic phrasing triggers the *phonetic realization* of pitch accents, i.e. pitch reset and downstep [4,5], less has been discussed about how prosodic phrasing modifies the underlying pitch accent of a lexical item itself, or whether a certain pitch accent combination would result in a particular prosodic phrasing. In this study, we investigate whether the variable nature of pitch accent in adjectives allows for an interaction between pitch accent and prosodic phrasing.

Data & Analysis To test this, complex DPs made up of two adjectival modifiers (Adj1/Adj2) and a head noun (N) were elicited, as shown in (1) embedded in a carrier sentence.

(1) *Gakkoo-de,* ***omoi marui iruka*** *dake hakkiri mieta*
'At school, I clearly saw **heavy round dolphins** only'

This structure was suitable for testing as it provides speakers freedom to produce the DP as one entire phonological phrase, as in (Adj1 Adj2 N), or to place a left boundary between Adj1 and Adj2, as in (Adj1 (Adj2 N)), where Adj2 marks the beginning of a separate phonological phrase. To look at accentuation, three accent combinations for Adj1-Adj2 were analyzed: unaccented-unaccented /UU/, unaccented-accented /UA/, and accented-unaccented /AU/. 16 sentences were constructed for each accent combination, yielding 48 items. A total of 576 tokens were elicited from 6 native speakers of Tokyo Japanese in their early twenties.

Each token was first coded for its surface pitch accent (i.e. /UU/ → [AU]) of the two adjectives. Prosodic phrasing was then determined by coding for the presence (1 or 0) of a left boundary at Adj2, which is signaled by a pitch reset for an accented Adj2, or an initial F0 rise for an unaccented Adj2 [4,5]. Examples are shown in Figures 1-4. Statistical analysis was performed using a binary logistic regression model in R with surface pitch accent and underlying pitch accent as fixed effects, while speaker, item, and repetition were included as random effects.

Results Overall, 366 tokens were phrased as (Adj1 Adj2 N) without an Adj2 boundary, and 210 tokens were phrased as (Adj1 (Adj2 N)) with an Adj2 boundary. The results showed a significant effect of surface pitch accent on prosodic phrasing. [AU] realizations correspond to an Adj2 boundary ($z=6.34$, $p<.001$), while [AA] ($z=-2.25$, $p<.02$) and [UA] ($z=-2.32$, $p<.02$) realizations correspond to the *absence* of an Adj2 boundary. The results for [UU] realizations were more mixed ($z=.16$, $p=.87$), with 60% tokens phrased without an Adj2 boundary, and the remaining 40% phrased *with* an Adj2 boundary. This is summarized in Table 1. No effect was found for underlying accent.

Discussion & Conclusion These findings indicate that [AU] structure facilitates (Adj1 (Adj2 N)) type phrasing. These findings pattern with previous findings where phrase-final adjectives are more likely to be accentuated than noun-modifying adjectives. In a (Adj1 (Adj2 N)) type phrasing, Adj1 is more likely to be accentuated as it precedes a left boundary, while Adj2 is more likely to remain unaccented as it directly modifies the head noun. [AA], [UA], [UU] realizations, on the other hand, correspond to (Adj1 Adj2 N) type phrasing. Together, these results show that pitch accent and prosodic phrasing may not be mutually exclusive.

Proceedings of the Second International Conference on Tone and Intonation, pages 89–90

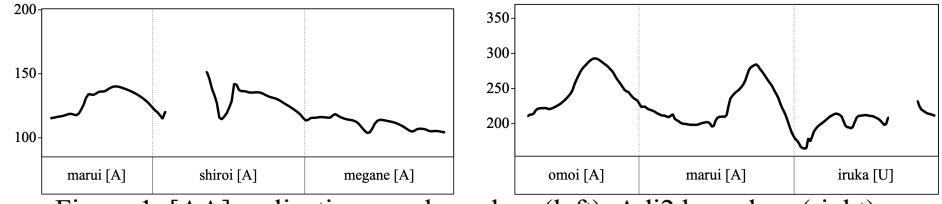


Figure 1: [AA] realizations, no boundary (left), Adj2 boundary (right)

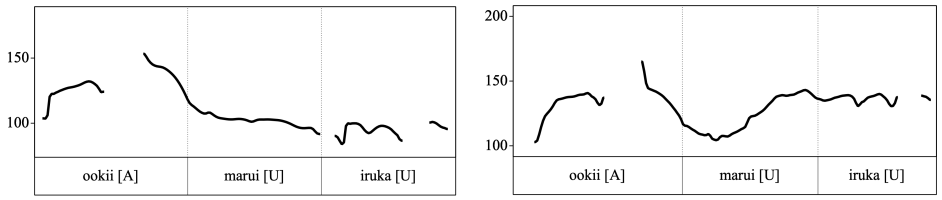


Figure 2: [AU] realizations, no boundary (left), Adj2 boundary (right)

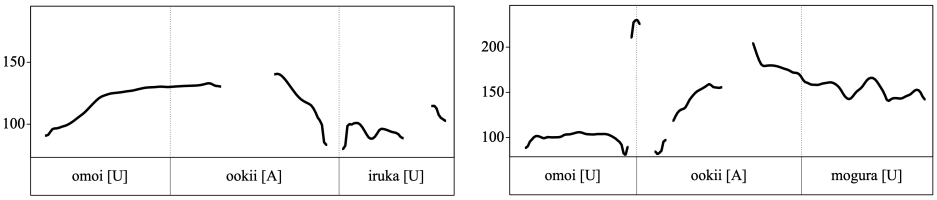


Figure 3: [UA] realizations, no boundary (left), Adj2 boundary (right)

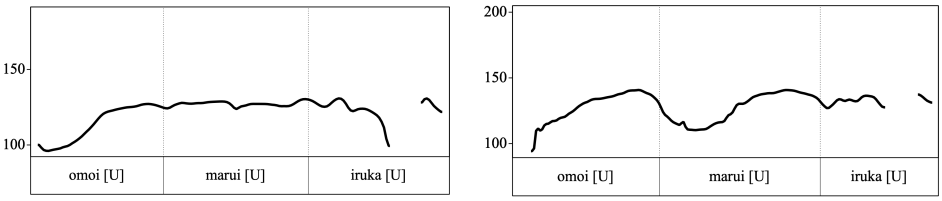


Figure 4: [UU] realizations, no boundary (left), Adj2 boundary (right)

Table 1: Number of tokens by surface pitch accent and prosodic phrasing

Surface Accent	(Adj1 Adj2 N)	(Adj1 (Adj2 N))	Std. Error	z	p
aa	126	20	0.92	-2.25	<.02
au	42	141	0.51	6.34	<.001
ua	118	9	0.55	-2.32	<.02
uu	80	40	0.59	0.16	0.87

References

[1] Johansson, T. 2020. Accentual change in the lexical form of Japanese i-adjectives: Comparison over time with attention to rentaikei and shuushikei.

[2] Kawahara, S. 2015. 11 The phonology of Japanese accent. In *Handbook of Japanese phonetics and phonology*, 445-492. De Gruyter Mouton.

[3] Kobayashi, M. 2003. An Examination of Adjective Accent Change in Tokyo Japanese and Factors Influencing the Change. *Journal of the Phonetic Society of Japan* 7(2), 101–113.

[4] Ishihara, S. 2016. Japanese downstep revisited. *Natural Language Linguistic Theory* 34(4), 1389–1443.

[5] Kubozono, H. 2021. *Ippan gengogaku kara mita Nihongo no purosodī: Kagoshima hōgen o chūshin ni (Japanese prosody from general linguistic perspectives)*. Kuroshio Shuppan.

Untangling the Word-Tone System: The Basic Tonal and Prosodic Patterns in Choca-ngaca

Xiyao Wang

The University of Sydney, Australia

xwan5208@uni.sydeny.edu.au

Choca-ngaca is a Tibetan language, under the Bodic subfamily of Tibeto-Burman languages [1, 2]. The language has a word-tone system that has been widely recognized in many Tibetan [3, 4], Tamangic [3, 5, 6, 7], Magaric [3] and East Bodish languages [8]. However, even within the word-tone system, patterns run differently across languages, such as the four types discussed in [4]. But due to relatively smaller speaking population, word-tone languages are often understudied, compared to the syllable-tone languages like Sinitic, which can pose biases to the construction of tone typology. Choca-ngaca is such an underdocumented language, spoken by only about 20,000 speakers in the east of Bhutan, with only one descriptive work [1], not to mention any work with a focus on its tone and prosody.

The current study will thus make a descriptive contribution to this underdocumented language by demonstrating a basic profile of its tone and prosody, and also to a more complete typological picture of the tone systems by examining how tone works in a word-tone language with only two tonal contrasts. The results showed that tone in Choca-ngaca falls on the initial syllable of a word, and the high or low category leads the prosodic pattern of the entire word or phrase in the disyllabic context and beyond. Consonantal laryngeal categories also come into interaction with tone to distinguish morphemes.

The data comes from Tongshan dialect in Trashi Yangtse district, as spoken by one female native speaker. Words and phrases were recorded both in isolation and carrier phrase. Due to time limitation, only isolated tokens were checked for acoustic results. The high and low-level tones, as the mere two tonal contrasts of Choca-ngaca, were both found to be able to fall on the vowels following sonorant consonants. On the contrary, voiced vs. voiceless obstruent consonants are found to be followed by different types of tone. The voiced initials are generally identified with low tone, and the voiceless with high tone.

For disyllabic words, the whole prosodic pattern is determined by the initial syllable. As displayed in Figure 1. (1) and 1. (2), the high-toned disyllabic mono- and di-morphemes (e.g., *kíli*, 'elbow') have a higher register in overall than the low-toned ones (e.g., *kùto*, 'head'). And the initial syllables of the high-toned words all carry a distinctively higher pitch than that of the low-toned words. However, the pitch contrast between the second syllables of high- and low-toned words is neutralized – very similar registers were attained on the second syllables of all disyllabic words.

Trisyllabic words perform very similar to the disyllables. Figure 2. (1) shows that the words with high tones (e.g., *khó-rang-ya*, 'himself also') have a higher register than those with low tones (e.g., *mò-rang-ya*, 'herself also'). Higher pitch is overwhelmingly applied to the initial syllables of the former than that of the latter. But again, a neutralization effect is imposed on the following syllables in all trisyllabic words. Pulling the prosodic performance of di- and tri-syllabic words together, the overall pattern is supposed to depend on the prosody of the first syllable.

Similarly, in multisyllabic noun phrases, the overall prosodic pattern is primarily determined by the tone type of the initial words, as illustrated in Figure 2. (2). The phrases with a 'H+H' (e.g., *chó-i nyúgu*) and 'H+L' tone sequence (e.g., *chó-i kùto*) start at a higher pitch than those with a 'L+H' (*ngà-i nyúgu*) and 'L+L' sequence (e.g., *ngà-i kùto*). The former all has a high-toned word at the onset (e.g., *chó-i*), and the latter with a low-toned word (e.g., *ngà-i*).

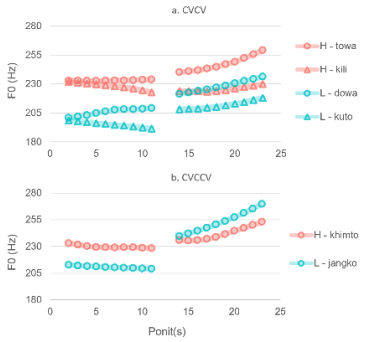
Moreover, the tone type of the second words, especially their initial syllable, also affects the prosodic pattern. Whether the words are high- or low-toned, they retain the same tonal pattern in a phrasal context as they are alone. This is revealed by the higher pitch on the third syllable of the 'H+H' and 'L+H' sequences than that of the 'H+L' and 'L+L' sequences, which is also the initial syllable of the second constituent of the whole noun phrase (e.g., *nyúgu* vs. *kùto* in *chó-i/ngà-i nyúgu* and *chó-i/ngà-i kùto*). The neutralization effect has also fallen on the second syllable of the second constituent, i.e., the last syllable of the phrases.

Proceedings of the Second International Conference on Tone and Intonation, pages 91–92
November 18–20, 2023.

However, the 'neutralization' phenomenon for word-tone system need to be further checked by comparing the acoustic performance of isolated vs. unisolated tokens, given the possible interpretation as 'boundary tone'.

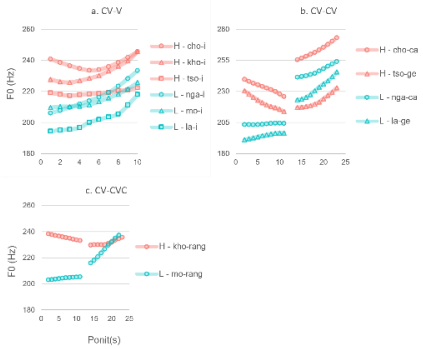
Figure 1: *Disyllabic tone patterns in Choca-ngaca.*

(1) Disyllabic monomorphemes



Notes:
a. /tówa/: 'stomach'; /kíli/: 'elbow'; /dòwa/: 'stone'; /kùto/: 'head'
b. /kʰímto/: 'roof'; /jàŋko/: 'chest'

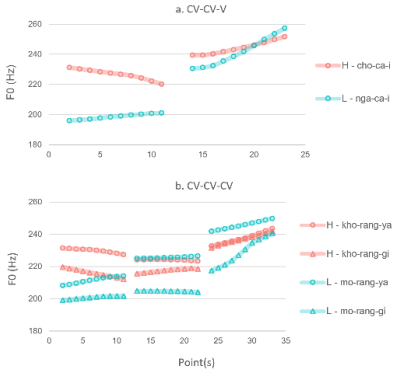
(2) Disyllabic dimorphemes



Notes:
a. /tɕʰó-i/: 'your'; /kʰó-i/: 'his'; /tsó-i/: 'lake's'; /ŋà-i/: 'my'; /mò-i/: 'her'; /là-i/: 'mountain pass's'
b. /tɕʰó-tɕa/: 'you (pl.)'; /tsó-ge/: 'in lake'; /ŋà-tɕa/: 'we'; /là-ge/: 'in mountain pass'
c. /kʰó-raŋ/: 'himself'; /mò-raŋ/: 'herself'

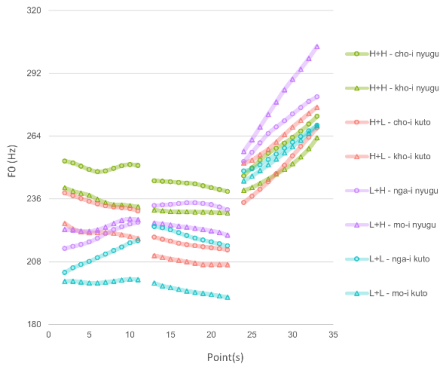
Figure 2: *Multiyllabic tone patterns in Choca-ngaca.*

(1) Trisyllables



a. /tɕʰó-tɕa-i/: 'yours (pl.)'; /ŋà-tɕa-i/: 'ours'
b. /khó-raŋ-ja/: 'himself also'; /khó-raŋ-gi/: 'it is him only'; /mò-raŋ-ja/: 'herself also'; /mò-raŋ-gi/: 'it is her only'

(2) Multisyllabic noun phrases



H+H: /tɕʰó-i ɲúgu/, 'your pen'; /kʰó-i ɲúgu/, 'his pen'
H+L: /tɕʰó-i kùto/, 'your head'; /kʰó-i kùto/, 'his head'
L+H: /ŋà-i ɲúgu/, 'my pen'; /mò-i ɲúgu/, 'her pen'
L+L: /ŋà-i kùto/, 'my head'; /mò-i kùto/, 'my head'

References

[1] Tournadre, N. L., & Rigzin, K. 2015. Outline of Choca-ngacakha. *Himalayan Linguistics*, *14*(2), 49–87.

[2] Hyslop, G. 2014. Waves Across the Himalayas: On the Typological Characteristics and History of the Bodic Subfamily of Tibeto-Burman. *Language and Linguistics Compass*, *8*(6), 243–270.

[3] Hildebrandt, K. A. 2007. Tone in Bodish languages: Typological and sociolinguistic contributions. In M. Miestamo & B. Wälchli (Eds.), *Trends in Linguistics. Studies and Monographs*. Mouton de Gruyter.

[4] Watters, S. 2002. The Sounds and Tones of Five Tibetan Languages of the Himalayan Region. *Linguistics of the Tibeto-Burman Area*, *25*(1), 1–65.

[5] Mazaudon, M. 1973. Comparison of six Himalayan dialects of Tibeto-Burmese. *Pakha Sanjham*, *6*, 78–91.

[6] Mazaudon, M. 2005. On tone in Tamang and neighbouring languages: Synchrony and diachrony. *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena*, 79–96.

[7] Mazaudon, M. 2014. Studying emergent tone-systems in Nepal: Pitch, phonation and word-tone in Tamang. *Language Documentation*, *8*, 26.

[8] Hyslop, G. 2021. Between Stress and Tone: Acoustic Evidence of Word Prominence in Kurtöp. *Language Documentation*, *15*, 551–575.

PraaPer: simple, rich and intuitive representations for tone and intonation

Francesco Cangemi & Aviad Albert
University of Cologne, Germany
fcangemi@uni-koeln.de, a.albert@uni-koeln.de

The visual representation of speech plays an important role in research on tone and intonation. It is used in many phases of the research process, from the exploration of data to the dissemination of findings. Currently, the *de facto* standard to visualize the phonetic aspects of tone and intonation is a run chart of the fundamental frequency, often extracted with the software *Praat* [1]. This chart can be stacked or overlayed with other phonetic representations, such as waveforms or spectrograms. The display is usually complemented by a tier showing speech segmentation at the word or syllable level, and further annotated with orthographic labels (see Figure 1).

This standard visualization has two important advantages. First, run charts of f0 capture the most salient aspect of tone and intonation, namely pitch height, and therefore have been used for more than a century. As a consequence, they are now extremely familiar to trained scholars. Second, these displays are relatively easy to extract in Praat, which is a popular choice among linguists, including those who do not specialize in acoustic phonetics. This allows many researchers in phonology, pragmatics and language documentation to join research on tone and intonation, with positive consequences for the field.

However, the standard representation also presents two important drawbacks. First, it is not an intuitive representation. The reader has to integrate information across the different tiers (e.g. waveform, pitch track, segmentation). In many cases, the reader must also filter erroneous, dubious or irrelevant information (e.g. uncorrected octave jumps, phone boundaries or micro-prosodic perturbations, respectively). Second, it is not a rich representation, since it only includes f0 height and overlooks information regarding pitch strength, i.e. the amplitude of the periodic portion of the signal [2]. Stretches of f0 coinciding with more periodic (vocalic) sounds are perceptually more salient [3], but their visual representation in the standard approach is not different from that of f0 stretches on less periodic voiced consonants.

To address these limitations, we introduce *PraaPer*, a new workflow for the visualization of pitch. This workflow is derived from the existing *ProPer* toolbox [3,4], which uses a set of complex *R* scripts to achieve similar goals. The new workflow is based on a simple Praat plug-in which outputs rich and intuitive phonetic representations in an easy-to-use Praat-native procedure (see Figure 2).

When using the Basic version of the plug-in, the script asks the user to mark the location of syllabic nuclei and to provide an orthographic transcription. The use of syllabic nuclei instead of syllabic boundaries has the advantages of being more perceptually relevant, less dependent on theoretical or practical stipulations, and easier to implement. Then the script extracts f0 candidates, and prompts the user to correct or filter inappropriate f0 choices, thus removing this burden from the readers' shoulders. In this stage, the user can efficiently stylize the chosen f0 by playing the resynthesis alongside the original sound. The script then smooths the f0 contour and scales it in a two-octave range around the speaker's median [5]. The script extracts pitch periodicity and plots the contour as a *periogram* [2], i.e. in a (time, frequency) plane, using greyscale for pitch strength. Finally, the script adds orthographic labels along the f0 curve, the speaker's median f0 along the central dotted line, and the file duration in the bottom right corner. Figure 3 offers a schematic representation of the workflow.

The PraaPer plug-in also offers an Advanced version, which allows the user to tweak several extraction and visualization parameters. This Advanced version can be seen as a stepping stone towards the ProPer suite [4], which allows to quantify prosodic aspects of speech such as prominence, speech rate and f0 contours [6,7]. After the conference, the PraaPer plug-ins and accompanying video-tutorials will be made available in a public repository.

Proceedings of the Second International Conference on Tone and Intonation, pages 93–94
November 18–20, 2023.

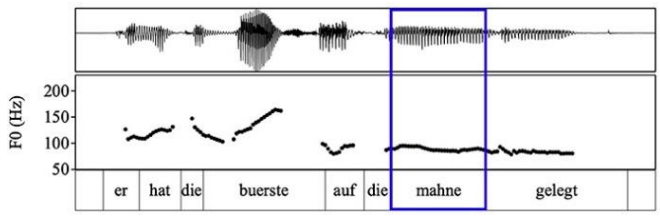


Figure 1. *Standard visualization, adapted from [7].*

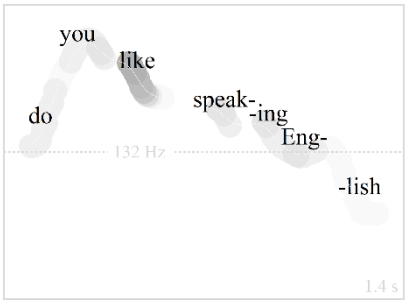


Figure 2. *Proposed PraaPer visualization.*

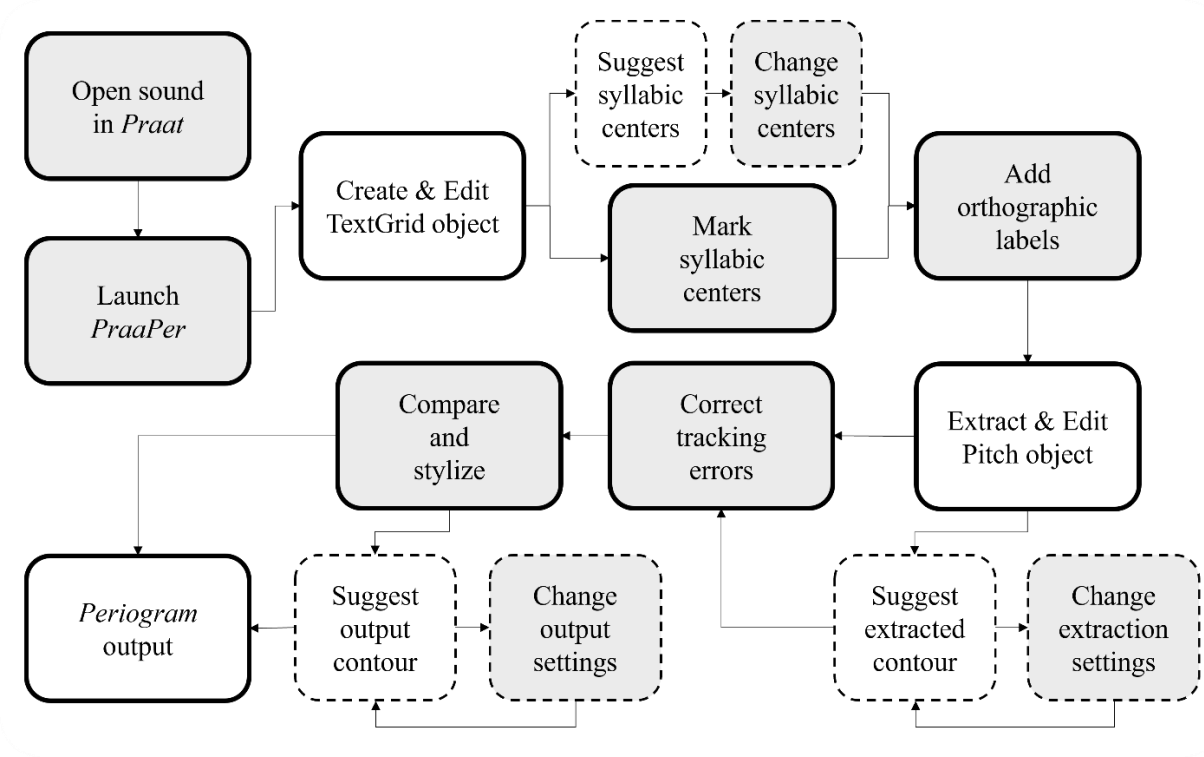


Figure 3. *PraaPer workflow for the extraction of periograms.*
Boxes with light grey fill indicate steps which require manual input from the user.
Boxes with thin dashed outline indicate steps which are only available in Advanced mode.

References

[1] Boersma, P., & Weenink, D. 2008. *Praat: doing phonetics by computer*. Computer program.

[2] Albert, A., Cangemi, F., & Grice, M. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Proceedings of Speech Prosody* (Poznan, Poland), 1-5.

[3] Albert, A. 2023. *A model of sonority based on pitch intelligibility*. Berlin: Language Science Press.

[4] Albert, A., Cangemi, F., Ellison, M., & Grice, M. 2023. *ProPer: PROsodic analysis with PERiodic energy*. https://osf.io/28ea5

[5] Hirst, D., & De Looze, C. 2021. Fundamental frequency and Pitch. In Knight, R. & Setter, J. (Eds.), *The Cambridge Handbook of Phonetics*. Cambridge: CUP, 336-361.

[6] Cangemi, F., Albert, A., & Grice, M. 2019. Modelling intonation: Beyond segments and tonal targets. *Proceedings of International Congress of Phonetic Sciences* (Melbourne, Australia), 1-5.

[7] Roessig, S., Winter, B. & Mücke, D. 2022. Tracing the phonetic space of prosodic focus marking. *Frontiers in Artificial Intelligence*, 5:842546.

Aspiration and tones in Guangxi Cantonese

Bei Yang & Wenting Xian
Sun Yat-sen University
yangb76@mail.sysu.edu.cn, 205433230@qq.com

The tone split by aspiration refers to the differentiation of tones caused by both the tone pitch and the feature of aspiration, i.e., the air flow of the initial consonant (Wang, 2016). It is also known as "air flow induced tone split". In such a phenomenon, it is controversial which feature, i.e., tone pitch or aspiration, is the main feature to distinguish the two tone categories (e.g. Zhu & Xu, 2009, Shi, *et al*. 2020). Simian Cantonese is employed to explore this question.

Simian Cantonese in Pingnan County, Guangxi Province, China, has 8 non-entering tones and 4 entering tones. The differences between *Quanqing* tone, which co-occurs with an unaspirated initial consonant in a syllable, and *Ciqing* tone, which co-occurs with an aspirated initial consonant, exist in pitch as well as aspiration. Among the 12 citation tones, three types, *Yinping*, *Yinqu* and *Shangyinru*, were divided into *Quanqing* and *Ciqing* tones respectively. The current study conducts the perceptual experiments on *Quanqing Yinping* tone and *Ciqing Yinping* tone.

Methods

By swapping the fundamental frequencies of *Quanqing Yinping* tone and *Ciqing Yinping* tone, two new stimuli were synthesized. These 2 new tones, *Quanqing* [24] tone and *Ciqing* [44] tone, and the originally natural tones, *Quanqing* [44] tone and *Ciqing* [24] tone, formed 4 tones for the perceptual experiments.

Thirty-eight native speakers of Cantonese (Male:22, Female:16) who came from Simian Town, Pingnan County, Guangxi Province participated in this study.

Two experiments were conducted. In the identification experiment, we selected 10 minimal pairs of disyllabic words. In each pair, all the segments and the tone carried by the first syllable were the same, and the only difference between the two words was the tone carried by the second syllable, one was the original *Quanqing* [44] tone while the other is the *Ciqing* [24] tone. Then we synthesized 10 *Quanqing* [24] tone and 10 *Ciqing* [44] tone based on the natural tones mentioned above. The participants listened to the 40 words three times randomly, so totally there were 120 trials. On each trial, participants heard one word, and their task was to report the meaning of the word.

In the discrimination experiment, we matched a natural tone with a synthesized tone to form a pair (Pisoni, 1973). Finally, there were four groups, and each group contained 10 pairs of words (see Table 5). The participants listened to the 40 pairs of words three times randomly, so totally there were 120 trials. On each trial, participants heard two words, and their task was to report whether they were the same or different.

Results

Table 1-4 shows the results of the identification experiment, and Table 5 shows the results of the discrimination experiment. These results indicate that either the tonal value 24 or the aspirated feature of *Ciqing Yinping* tone can be used to distinguish *Quanyin* tone from *Ciyin* tone, while both features, i.e., the pitch value 44 and unaspiration, of *Quanqing Yinping* tone are required to distinguish *Quanyin* tone from *Ciyin* tone. Generally, the influence by the pitch value is greater than that by the aspirated feature on perception.

Discussion

The results reveal the trend of the development of the tone split by aspiration. If *Quanqing* and *Ciqing* still maintain two tone categories, then, among the four features, aspirated, non-aspirated, low-pitched and high-pitched, the aspirated initials should be merged into unaspirated initials. The results in Table 5 demonstrates this: the two conditions in group1 are equivalent to the situation of aspiration merge, and the percentage of the discrimination (92.37%) is the highest among the four groups. The evidence from Ho (1989) and Cao (2014) in the previous studies also supports this viewpoint.

The article further discusses the reasons why *Ciqing Yinping* tone can still be distinguished from *Quanqing Yinping* tone after the aspirated initials are merged into the unaspirated initials.

Proceedings of the Second International Conference on Tone and Intonation, pages 95–96
November 18–20, 2023.

Table 1 *Results of the identification experiment for natural Quanqing [44] tone*

	Quanqing [44]	**Ciqing [24]**	**Yangping [21]**	**Do not know**	**Sum**
Number	1112	23	4	1	1140
Percentage	97.54%	2.02%	0.35%	0.09%	100%

Table 2 *Results of the identification experiment for natural Ciqing [24] tone*

	Quanqing [44]	**Ciqing [24]**	**Yangping [21]**	**Do not know**	**Sum**
Number	21	1117	2	0	1140
Percentage	1.84%	97.98%	0.18%	0%	100%

Table 3 *Results of the identification experiment for synthesized Quanqing [24] tone*

	Quanqing [44]	**Ciqing [24]**	**Yangping [21]**	**Do not know**	**Sum**
Number	56	1072	10	2	1140
Percentage	4.91%	94.03%	0.88%	0.18%	100%

Table 4 *Results of the identification experiment for synthesized Ciqing [44] tone*

	Quanqing [44]	**Ciqing [24]**	**Yangping [21]**	**Do not know**	**Sum**
Number	257	873	10	0	1140
Percentage	22.54%	76.58%	0.88%	0%	100%

Table 5 *Conditions and results of the discrimination experiment*

Group	**Condition1 same feature**	**Condition2 different features**	**Perceived differently**
1	Quanqing (unaspirated)	[44] and [24]	92.37%
2	Ciqing (unaspirated)	[44] and [24]	54.65%
3	Quanqing ([44])	Aspirated & unaspirated	79.47%
4	Ciqing ([24])	Aspirated & unaspirated	21.75%

References

[1] Cao, Z. 2014. The phenomenon of deaspirated of the initial consonants in Tongdao Kam language -- Comparison with Gan dialect. *Minzu Yuwen* 3, 37-44.

[2] Ho, D. 1989. The tone-split by aspiration and other related problems. *Bulletin of IHP* 60(4), 765-778.

[3] Shi, M., Y. Chen, and M. Mous. 2020. Tonal split and laryngeal contrast of onset consonant in Lili Wu Chinese. *The Journal of the Acoustical Society of America* 147(4), 2901-2916.

[4] Pisoni, D. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics* 13(2), 253-260.

[5] Wang, L. 2016. *Tone split in Chinese dialects*. Beijing: Yuwen Chubanshe (Language and Culture Press).

[6] Zhu, X., & Xu, Y. 2009. Slack voice as a cause of tone split: A phonetic investigation into the pitch contour of the syllables with a voiceless aspirated onset in the Wujiang (Songling) Wu dialect. *Zhongguo Yuwen* 4, 324-332.

Study on Chinese Tone Acquisition of Learners From Different Language Backgrounds

LI Jinhao
University International College, Macau University of Science and Technology
3220003145@student.must.edu.mo

As a tonal language, Chinese uses different tones to express different meanings. The long-term Chinese teaching experience shows that Chinese tones have always been a difficult problem for teachers to teach and learners to acquire. These difficulties cause many learners to have foreign accents (also called "Yangqiang yangdiao"[洋腔洋調]) when speaking Chinese(Lin, 1996; Shi & Wen, 2012). Therefore, the acquisition of Chinese tones by learners deserves attention.

From the perspective of cross-lingual influence, foreign accent is the product of the mutual influence between different languages (Kang, Wu, Huang &Li, 2017), that is, the influence of the L1 phonetic system on the Chinese as L2 phonetic system. This effect can be called "L1 transfer" (Lado, 1957). According to "L1 transfer", the similar or same phenomes between L1 and L2 will promote the acquisition of L2(positive transfer); the difference between L1 and L2 will hinder the acquisition of L2 (negative transfer). However, further studies have found that differences between L1 and L2 are not necessarily the cause of learners' errors (Dulay & Burt, 1973; Flick, 1980). People have gradually discovered that the similarities between L1 and L2 will also produce negative transfer (Ellis, 1999). The Speech Learning Model (SLM) (Flege, 1995) unusually believes that the similarity between L1 and L2 is likely to cause learners to fail to acquire similar phonemes successfully. This is because learners will miscategorize L2 phonemes into L1 phonemes under the cognitive mechanism of equivalence classification.

We can make two hypotheses with opposite conclusions: (1) According to the "L1 transfer", the hypothesis that tonal language speakers (TLSs) produce better Chinese tones than non-tonal language speakers (NTLSs) is proposed; (2) According to the SLM, the hypothesis that non-tonal language speakers produce better Chinese tones than tonal language speakers is proposed. Regarding the acquisition of Chinese tones, the divergence of result between different theories need to be investigated and tested.

In this study, 12 Bangladeshi learners (BdLs) as NTLSs and 12 Thai learners (ThLs) as TLSs were invited. In all BdLs, 6 of them had passed the HSK1/2 (Elementary level, Chinese learning time is between 6 and 10 months), and the other 6 has passed the HSK 3/4 (Intermedia level, Chinese learning time is between 26 and 30 months). In all ThLs, 6 of them had passed the HSK1/2 (Elementary level, Chinese learning time is between 6 and 10 months), and the other 6 has passed the HSK 3/4 (Intermedia level, Chinese learning time is between 28 and 29 months). Production of 8 Chinese native speakers (CNSs) was used as a reference to investigate the influence of different L1 backgrounds on the production of Chinese tone. Use Praat to process audios and extract their F0 data. Normalization was performed according to LZ-Score (Zhu, 2004). Tone plots were made according to Zhu (2010). Growth Curve Analysis was also used to observe differences in production. In addition, 4 CNSs were invited to score the tones that randomly selected.

The research results show that: 1) ThLs performed significantly better than BdLs in T1, T3 and T4; there was no significant difference in T2, but ThLs performed better than BdLs in audibly. 2) The tone duration pattern of BdLs and ThLs is not much different from that of CNSs, but the T4 duration of ThLs is longer. 3) ThLs already have the awareness of tone sandhi in T3; BdLs are more sensitive to duration in T4. In summary, we support the hypothesis of "L1 transfer" (Lado, 1957) that TLSs produce better Chinese tones than NTLSs, tonal experience of L1 facilitates the acquisition of Chinese tones .

Key words: Chinese Learners, Across-languages, Chinese Tones

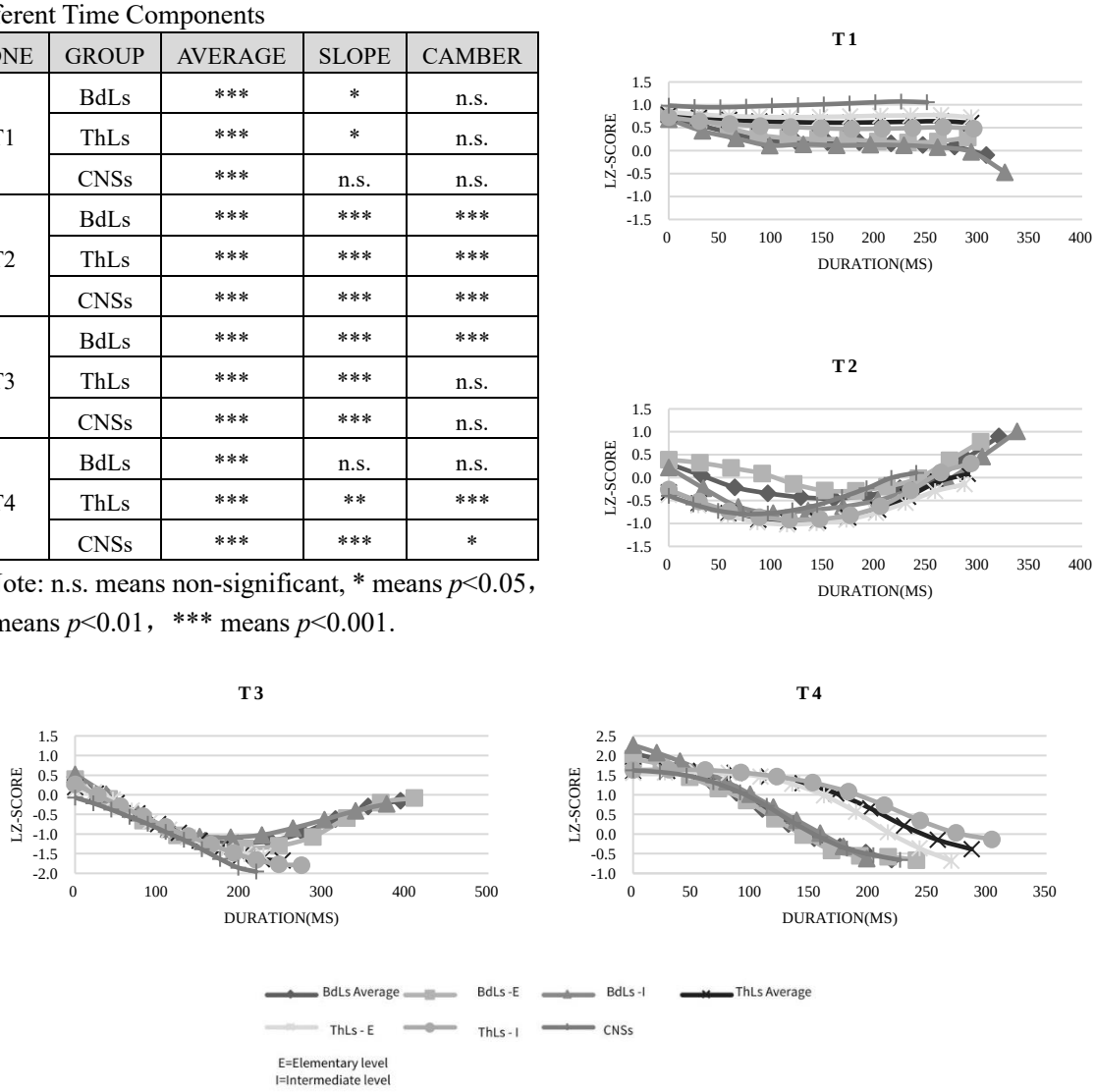
Proceedings of the Second International Conference on Tone and Intonation, pages 97–98
November 18–20, 2023.

Table 1 Significance Results of Effect Estimates on Different Time Components

TONE	GROUP	AVERAGE	SLOPE	CAMBER
T1	BdLs	***	*	n.s.
	ThLs	***	*	n.s.
	CNSs	***	n.s.	n.s.
T2	BdLs	***	***	***
	ThLs	***	***	***
	CNSs	***	***	***
T3	BdLs	***	***	***
	ThLs	***	***	n.s.
	CNSs	***	***	n.s.
T4	BdLs	***	n.s.	n.s.
	ThLs	***	**	***
	CNSs	***	***	*

Note: n.s. means non-significant, * means $p<0.05$, ** means $p<0.01$, *** means $p<0.001$.

Graph 1 Tone Curves of T1, T2, T3 and T4



References

[1] Kang, Y., Wu, C., Huang, L., & Li L. 2017. Effects of Tone Language Experience on Second Language Tone Acquisition. *Modern Linguistics* 5(3), 234-240.

[2] Li, L., Li Y.., Kang, Y., & Wang, L. 2020. Effects of Tonal Language Experience on Perception of Mandarin Tones by Chinese-as-a-Second-Language Speakers. *Journal of South China normal University (Social Science Edition* 1, 83-91.

[3] Li, Q., Shi, M., & Chen, Y. 2020. A New Statistical Approach in the Studies of Tones: Two Case Studies on Chinese Dialect Using Growth Curve Analysis. *Studies of the Chinese Language* 5, 591-608+640.

[4]Liao, Y., & Zhang, W. 2019. On the Effects of L1 Background on the Perception of Mandarin Tones. *Chinese Language Learning* 1, 75-86.

[5] 林燾. 1996. 語音研究和對外漢語教學. *世界漢語教學* 3, 20-23.

[6] 劉增慧. 2021. *白漢雙語兒童漢語韻律焦點發展研究*. 北京: 中國社會科學出版社.

[7] Shi, L. & Wen, B. 2012. Exploration on Foreign Accent: A Research of the Acquisition of Chinese Intonation by American Students. *Nankai Linguistics* 1, 42-49.

[8] Zhu, X. 2004. F_0 Normalization: How to Deal with Between – speaker Tonal Variations? *Linguistic Sciences* 2, 3-19.

[9] 朱曉農. 2010. *語音學*. 北京: 商務印書館.

The acquisition of L2 Mandarin T3 sandhi and neutral tone by Japanese speakers

Tong Shu & Peggy Pik Ki Mok
The Chinese University of Hong Kong
tongshu@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

Many studies have been conducted on the acquisition of isolated Mandarin lexical tones (T1-T4) by L2 learners [1]. However, a comprehensive understanding of how they acquire the more complex prosodic patterns of Mandarin tones in various contexts (e.g., T3 sandhi) remain limited. Previous studies on Mandarin T3 sandhi have found that the Half T3 sandhi is easier for L2 learners to acquire than the Full T3 sandhi due to clearer phonetic motivation [2, 3]. However, this claim of phonetic motivation can be further tested with more contextual tone variations in Mandarin and with learners from prosodically distinct L1s (see [4] for a study on English speakers). In this preliminary study, we further compared the acquisition of the two types of T3 sandhi and neutral tone by Japanese learners of two proficiency levels. The aim is to investigate the effect of phonetic motivation and L2 proficiency on the acquisition of L2 contextual tones.

In Mandarin, T3 undergoes two different sandhi processes in specific contexts. The low-dipping T3 (214) becomes T2 (35) when followed by another T3 (Full T3 sandhi), and changes to a Half-T3 (21) when followed by T1/2/4 (Half T3 sandhi) [5]. Both types of T3 sandhi are obligatorily applied in disyllabic words across different morphological structures. In addition to the four lexical tones, Mandarin also has a neutral tone (T0). The neutral tone is different from lexical tones in that it cannot appear independently and resembles unstressed syllables with shorter duration, vowel reduction, and underspecified pitch contours [6]. The neutral tone is obligatory in morphemes that do not have citation tones (e.g., the suffix -de) but is non-obligatory in morphemes with citation tones [7].

The neutral tone can be considered more phonetically motivated than the two types of T3 sandhi, as it is a reduction phenomenon, and its surface pitch contour can be automatically derived from the preceding tone via the carryover effect [8]. In contrast, both types of sandhi T3s are full tone syllables which require more articulatory efforts. Among the two types of T3 sandhi, the Half T3 sandhi is believed to have clearer phonetic motivation than the Full T3 sandhi [5] because it is natural to simplify a complex pitch contour in speech (Half T3 sandhi), while the change of T3 into T2 in Full T3 sandhi is more arbitrary. Therefore, based on phonetic motivation, we predicted the relative ease of acquisition for the three aforementioned contextual tones to be: neutral tone > Half T3 sandhi > Full T3 sandhi.

Six Japanese learners of Mandarin with intermediate proficiency (HSK3~4) and 10 with advanced proficiency (HSK 5~6) participated in a reading experiment. The production stimuli included 60 disyllabic words for all T3 sandhi contexts (T3T1, T3T2, T3T3, T3T4) and 52 words for all preceding tone combinations with the neutral tone (T1T0, T2T0, T3T0, T4T0). Two native Mandarin speakers made auditory judgments on their production. Several generalized linear mixed-effect models were run to analyze the accuracy rate of different contextual tones by different groups of speakers.

Figure 1 shows the accuracy rate of T3 sandhi and the neutral tone for intermediate learners. The results showed that some tonal combinations are more difficult than others within each contextual tone category, i.e., T3T2 in Half T3 sandhi and T3T0 in the neutral tone. Excluding these exceptions, pairwise comparison results showed that the accuracy rate was: obligatory T0 > Half T3 > Full T3; Non-obligatory T0 > Full T3, which supported our hypothesis that the neutral tone was easier than the T3 sandhi. Figure 2 shows the accuracy of T3 sandhi and the neutral tone by intermediate and advanced learners. The advanced learners showed significantly higher accuracy rates than their intermediate counterparts for most tonal combinations for both the T3 sandhi and the neutral tone. However, similar to the intermediate group, the T3T2 and T3T0 sequences were still not as good as the other combinations. This may be due to the T2-T3 confusion that L2 learners of Mandarin have difficulties distinguishing T2 and T3 both in perception [1] and production [9, 10]. Therefore, when two confusable tones occur in the same sequence (T3T2), it may be more difficult for L2 learners to apply the correct sandhi pattern. Also, the most common error pattern of T3T0 was to produce it as T2T0, which suggests the influence of T2-T3 confusion.

To conclude, our study generally supports that phonological patterns that are more phonetically motivated are easier for L2 learners to acquire. Additionally, not all tonal combinations within the same contextual tone category are equally learnable, as this is also influenced by the acquisition of individual tones.

Proceedings of the Second International Conference on Tone and Intonation, pages 99–100

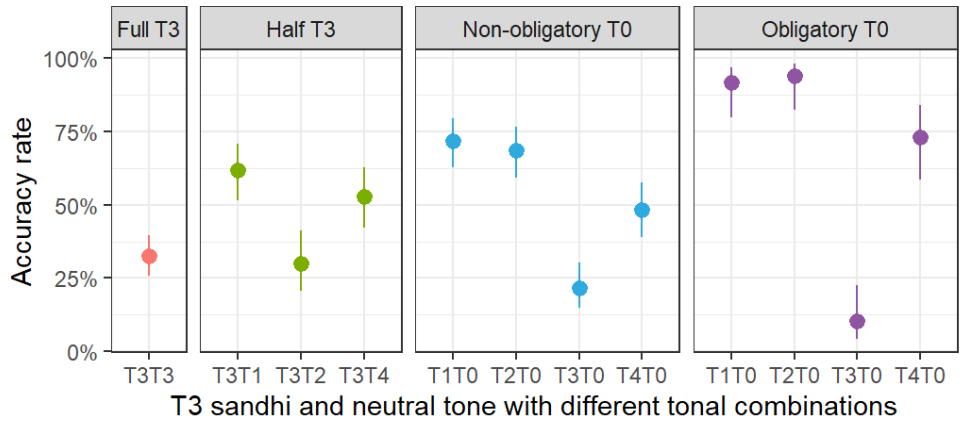


Figure 1. *Mean accuracy rate of T3 sandhi and neutral tone in intermediate learners*.

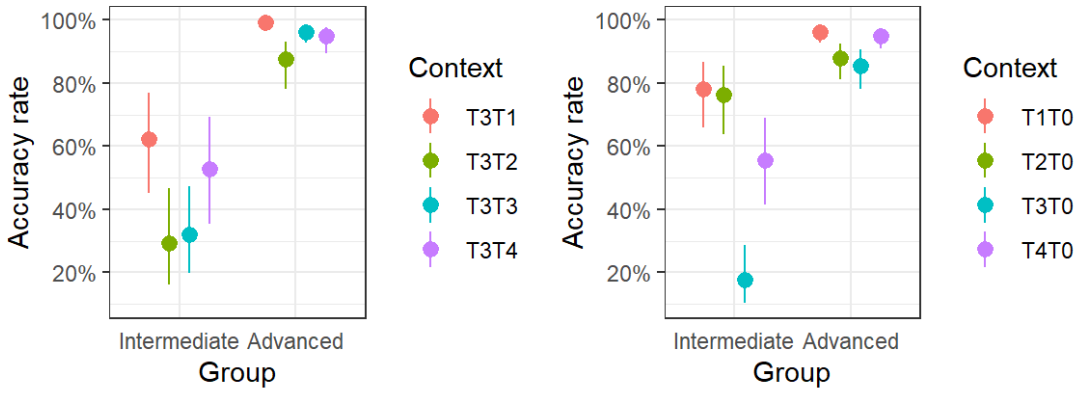


Figure 2. *Mean accuracy rate of T3 sandhi (left) and neutral tone (right) of the intermediate and advanced groups*

References

[1] E. Pelzl, "What makes second language perception of Mandarin tones hard?," *Chinese as a Second Lang. J. Chinese Lang. Teach. Assoc. USA*, vol. 54, no. 1, pp. 51–78, 2019.

[2] C. Yang, "Acquisition of Mandarin Tone 3 sandhi Interaction of phonology, phonetics, and pedagogy," in *The Acquisition of L2 Mandarin Prosody : From experimental studies to pedagogical practice*, John Benjamins, 2016.

[3] Z. Qin, "The Second-Language Productivity of Two Mandarin Tone Sandhi Patterns," *Speech Commun.*, vol. 138, no. January, pp. 98–109, 2022.

[4] W. Jin, "Acquisition of tone sandhis by English speaking learners of Chinese," *J. Natl. Counc. Less Commonly Taught Lang.*, vol. 25, no. 1, pp. 67–107, 2019.

[5] J. Zhang and Y. Lai, "Testing the role of phonetic knowledge in Mandarin tone sandhi," *Phonology*, vol. 27, no. 1, pp. 153–201, 2010.

[6] W.-S. Lee and E. Zee, "Prosodic characteristics of the neutral tone in Beijing Mandarin," *J. Chinese Linguist.*, vol. 36, no. 1, pp. 1–29, 2008.

[7] C. B. Chang and Y. Yao, "Production of neutral tone in mandarin by heritage, native, and second language speakers," in *the 19th International Congress of Phonetic Sciences*, 2019, pp. 2291–2295.

[8] Y. Chen and Y. Xu, "Production of weak elements in speech - Evidence from F0 patterns of neutral tone in standard chinese," *Phonetica*, vol. 63, no. 1, pp. 47–75, 2006.

[9] J.-Y. Tu, Y. Hsiung, M.-D. Wu, and Y.-T. Sung, "Error patterns of Mandarin disyllabic tones by Japanese learners," in *Interspeech 2014*, 2014, pp. 2558–2562.

[10] J.-Y. Tu, Y. Hsiung, J.-H. Cha, M.-D. Wu, and Y.-T. Sung, "Tone production of Mandarin disyllabic words by Korean learners," in *Proceedings of the International Conference on Speech Prosody*, 2016, pp. 375–379.

Carryover Tonal Variations for Speech Recognition in Standard Chinese

Hana Nurul Hasanah [1,3], Qing Yang [1,2], Yiya Chen [1,2]
[1]*Leiden University Centre for Linguistics,* [2]*Leiden Institute for Brain and Cognition,* [3]*Chinese Studies Universitas Indonesia*
h.n.hasanah@hum.leidenuniv.nl, q.yang@ hum.leidenuniv.nl, yiya.chen@hum.leidenuniv.nl

We know that lexical tones co-articulate in connected speech; when excised out of tonal context with robust carryover pitch variations and played in isolation, listeners' tonal identification rate may drop below the chance level [1], suggesting that tonal co-articulatory cues play no role in offline processing of tones in isolation. In real-time processing, however, studies have shown that listeners are sensitive to detailed within-category pitch variations for tone recognition in isolation [2-4]. Using the Visual World Paradigm (VWP, [5]) we complement the literature by investigating how exactly listeners utilise tonal co-articulatory pitch variations for online speech recognition.

Thirty-two native speakers of Standard Chinese (SC) listened to bi-syllabic nonce words with tonal co-articulatory cues and performed a forced-choice task selecting the corresponding nonce word out of two printed in Chinese characters. Nonce words were used to present required phonological overlaps. To complete the task successfully, participants must attend to the segmental and tonal information of the second syllable (S2). For each target nonce word (e.g., 瓦咪 *wa*3*mi*1), the S2 contained a high-level tone (T1), which surfaces with a rising f0 contour when preceded by a low tone (T3) in the first syllable (S1) but stays level when preceded by another high tone (T1) [6]. The S2 of each target nonce word was contrasted with a competitor (i.e., the other printed word) which differed in tone only (CRITICAL trial, e.g., 瓦弥 *wa*3*mi*2) or in segments and tone (BASELINE trial, e.g., 瓦册 *wa*3*ce*4). Auditorily, the S1 was manipulated such that segments and tone information were present as original (UNMASKED) or masked by pink noise. With S1 masked, we manipulated the visual presentation of the first syllable, with one compatible with the original and therefore having the right tonal coarticulatory information on S2 (MASKED APPROPRIATE) and the other with a character that differed in tone from the original and therefore misleading tonal coarticulatory information (MASKED INAPPROPRIATE) (see Table 1).

Listeners' eye fixations on visually presented characters were recorded and analysed from 200 ms post auditory stimulus onset (given the known oculomotor delay) until 1100 ms (roughly when maximum fixation was reached). Analysis of the point of divergence (POD) in listeners' eye fixations was performed to determine when listeners began to differentiate the target from the competitor [7]. Results showed that the POD was significantly earlier when a target nonce word was presented alongside a competitor that differed in segment and tone (UNMASKED BASELINE), compared to the tone-only condition (UNMASKED CRITICAL) (Fig. 1A vs. Fig. 1B). For the UNMASKED CRITICAL, the POD was observed about 54 ms after the onset of S2, suggesting that listeners attended to the pitch differences between the target and competitor tones from early on to recognise the correct characters in the absence of segment information, even though the pitch differences were very subtle in the initial portion of S2 due to tonal coarticulation. In the MASKED APPROPRIATE condition (Fig. 1C), the POD fell around 108 ms post the S2 onset. The POD was further delayed (157 ms), when the tonal coarticulatory cue was inappropriate (MASKED INAPPROPRIATE condition; Fig. 1D). This trend of POD difference (49 ms) in the latter two conditions hints that listeners were somewhat sensitive to the inappropriate tonal co-articulatory cue in the MASKED INAPPROPRIATE condition, which caused the delay. The POD difference was verified in our further analysis of the proportions of fixations (POF) using Generalised additive mixed-modelling (GAMM; *mgcv* package version 1.8-41) [8]. POF in the MASKED INAPPROPRIATE condition was significantly lower than the MASKED APPROPRIATE condition ($p<0.01$) for an interval of ±160 ms shortly after the onset of S2.

This study revealed that during speech recognition in context, native SC listeners incrementally process pitch information as an auditory stimulus unfolds over time. Despite the low ceiling fixation rate, nonce words provide a direct test of listeners' perception of co-articulatory cues. An inappropriate tonal co-articulatory cue hinders the tonal recognition process, suggesting the utilisation of tonal coarticulatory information in online speech processing. Our results lend evidence to the possibility that details of pitch variation for carryover tonal co-articulation are stored as part of lexical tone representations and facilitate tonal processing even when the preceding tonal context is absent.

Proceedings of the Second International Conference on Tone and Intonation, pages 101–102

Table 1: *Sample of visual stimuli presentation across conditions (transcription in Pinyin with tone numbers below Chinese characters was not presented during the experiment).*

Trial	**Unmasked**		**Masked appropriate**		**Masked inappropriate**	
	Target	*Competitor*	*Target*	*Competitor*	*Target*	*Competitor*
Critical	瓦咪 *wa3mi1*	瓦弥 *wa3mi2*	瓦咪 *wa3mi1*	瓦弥 *wa3mi2*	洼咪 *wa1mi1*	洼弥 *wa1mi2*
Baseline		瓦册 *wa3ce4*		瓦册 *wa3ce4*		洼册 *wa1ce4*

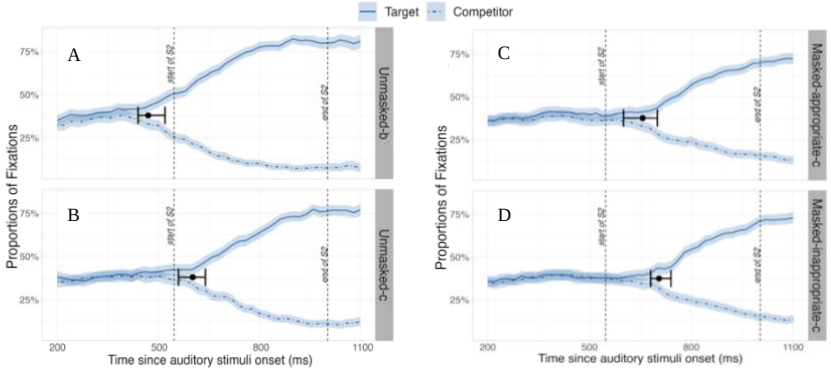


Figure 1: *Means of POD and 95% CI over the fixation curves.* A*: UNMASKED BASELINE*, B*: UNMASKED CRITICAL*, C*: MASKED APPROPRIATE CRITICAL*, D*: MASKED INAPPROPRIATE CRITICAL. The first and second vertical dotted lines, respectively, indicate the second syllable onset (547 ms) and ending (1003 ms).*

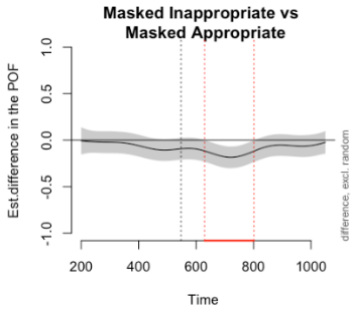


Figure 2: *GAMM difference curves contrasting MASKED INAPPROPRIATE - MASKED APPROPRIATE in CRITICAL trials. The vertical black dotted line indicates the start of the second syllable (TIME = 547). The vertical red dotted lines mark the interval of significant difference between two fixation curves.*

References

[1] Xu, Y. 1994. Production and perception of coarticulated tones. *J. Acoust. Soc. Am.* 95(4), 2240-2253.

[2] Qin, Z., Tremblay, A., & Zhang, J. 2019. Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *Journal of Phonetics*, 73, 144-157.

[3] Shen, J., Deutsch, D., & Rayner, K. 2013. On-line perception of Mandarin Tones 2 and 3: Evidence from eye-movements. *J. Acoust. Soc. Am.* 133(5), 3016-3029.

[4] Yang, Q., & Chen, Y. 2022. Phonological competition in Mandarin spoken word recognition. *Language, Cognition and Neuroscience* 37:7, 820-843.

[5] Magnuson, J.S. 2019. Fixations in the visual world paradigm: When, where, why? *J. Cult Cogn Sci* 3, 113-139.

[6] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.

[7] Stone, K., Lago, S., & Schad, D.J. 2020. Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition* 24, 833-841.

[8] Wood, S.N. 2020. *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.* https://cran-r-project.org/web/packages=mgcv

Production of Mandarin Third Tone Sandhi by the Young Generation in Malaysia

Xin Ren[1] & Poh Shin Chiew[2]
[1, 2]*Universiti Malaya*
tva180025@siswa.um.edu.my, chiewpohshin@um.edu.my

Today Mandarin is a regional-global language, no longer the property of a certain region. Accordingly, 'the Global Chinese (GC) concept' has burst upon the linguistic field, indicating a conceptual shift and pluricentric approach to investigating Mandarin varieties. Although Mandarin is known for its intricate tonal patterns, prior research on Mandarin tone sandhi has predominantly focused on *Putonghua*, the standard Mandarin used in mainland China, and relatively neglected have been the Mandarin varieties outside mainland China. Malaysia plays an important role in GC due to its spread and maintenance of Mandarin. As Malaysia is a multi-ethnic and multilingual society, third tone (T3) sandhi in Malaysian Mandarin (MalM) may have considerable variations. Several studies showed that albeit T3 sandhi remains in MalM, it exhibits variations in realizations ([1, 4]); Khoo, however, (2014) suggested that T3 sandhi does not occur in MalM. Hence, it is worthwhile to explore T3 sandhi in disyllabic words in MalM as disyllabic tone sandhi serves as a basic unit of intonation ([7]).

Seventy-one female Chinese Malaysians were recruited, aged from 20 to 25 (Mean = 22.0), who represented the young generation of Mandarin speakers in Malaysia. Given that the variations are probably attributed to the influence of the home-domain language, the Chinese dialects [3, 4, 6], the participants were divided into two groups based on the language used most frequently at home. While 36 participants chiefly used at least one Chinese southern dialect at home (CD_D), the rest dominantly spoke Mandarin at home (M_D). Other than Mandarin and Chinese southern dialects, the participants occasionally used Malay and English. Seven disyllabic targets were taken from a disyllabic wordlist which was designed in terms of Middle Chinese. Recordings were made with smartphones to overcome the COVID-19 pandemic restrictions in 2020. Manual checking was conducted to validate the recordings. The total number of recorded tokens were 994 (7 targets x 2 repetitions x 71 participants). However, only 992 tokens were examined by visual inspection and auditory perception as two tokens were excluded due to background noise. Among the valid tokens, 565 tokens were further acoustically measured while the rest were excluded from the acoustic examination due to the failed pitch tracking caused by creaks (Figure 1). Fundamental frequency (F_0) was the main correlate used in the acoustic measurement of T3 sandhi, and the acoustic results were presented in Chao's notation with five distinctive levels, ranging from 1, the pitch floor, to 5, the pitch ceiling.

The findings suggest that the T3 sandhi rule was implemented consistently in the majority of tokens in both groups (Table 1). This indicates that MalM entails a similar phonological process of *Putonghua*, where T3 sandhi confines to the co-occurrences of two T3s, such that the non-final T3 is realized similarly to T2 when precedes another T3. This is also reported in Singapore Mandarin and Taiwan Mandarin ([2, 8]). Despite the similar phonological process, the variations seem to take place at the phonetic level. First, our results (Table 2) show that the non-final T3 in MalM is mainly pronounced as [334]. This differs from *Putonghua* where the non-final T3 is mainly produced as [35]. Second, the final T3 is prominently produced as the mid-falling contour [31] in MalM despite being before a pause. However, in *Putonghua*, T3 as a low-falling tone [21] is common everywhere except before a pause. Third, in the dialect group, non-neutralization of T3 sandhi also occurs in a subset of tokens despite the low frequency of occurrences. Besides, despite the consistency in both groups, more variants happen in CD_D than in M_D, which indicates a greater possibility of instability of CD_D.

While the phonological process of T3 sandhi in this research is rather consistent in both fluent speakers of Mandarin and those who use it as the dominant language at home. This lends sufficient support to the fact that T3 sandhi remains a systematic feature in MalM. Accordingly, the current study agrees with Huang (2016) and Chiew (2021), contra Khoo (2014). However, the phonologically stable tonal patterns can exhibit dialectal variation in realizations as compared to *Putonghua*. For example, as aforementioned, in MalM the non-final T3 is mainly realized as [334] in T3 sandhi, and the final T3 is mainly uttered as the mid-falling contour [31]. Considering these variations, we may surmise that the development of MalM is similar but not exactly the same as *Putonghua*. In addition, when comparing these two language groups, it is suggested that the influence of the home-domain language occurs at the phonetic implementation level.

Proceedings of the Second International Conference on Tone and Intonation, pages 103–104
November 18–20, 2023.

Table 1: *Distributions of tonal realization of third tone sandhi in the Chinese dialect group (a) and in the Mandarin group (b) in Malaysian Mandarin*
Notes: While numbers in parentheses indicate the total number of tokens, percentages preceded the parentheses are the frequency of occurrences.

(a)

Non-final tone	Final tone			
	Falling	Dipping	Level-rising	Level
Level-rising	89.6% (468)	3.6% (18)	2.4% (12)	
Falling	0.2% (1)	/	/	0.4% (2)
Level	3.6% (18)	/	0.2% (1)	
Total	93.4% (469)	3.6% (18)	2.6% (13)	0.4% (2)

(b)

Non-final tone	Final tone		
	Falling	Dipping	Level-rising
Level-rising	79.6% (390)	13.7% (67)	6.5% (32)
Falling	0.2% (1)	/	/
Level	/	/	/
Total	79.8% (391)	13.7% (67)	6.5% (32)

Table 2: *The tonal patterns in Chao's notation*

The language group	The major tonal pattern	Other variants of tonal patterns
The dialect group	[334+31]	[334+412], [33+32], [334+223], [42+44], [42+42], [22+223]
The Mandarin group	[334+31]	[334+413], [334+223]

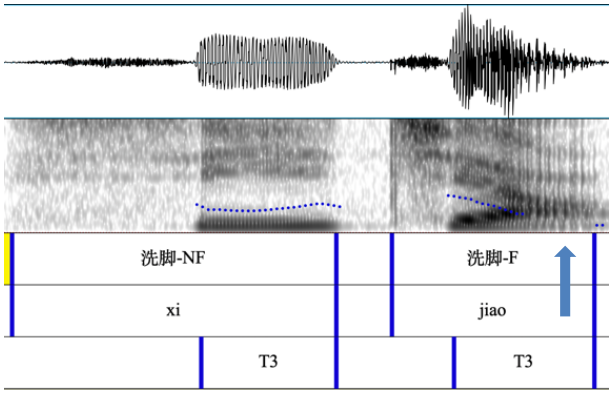


Figure 1: *Samples of creaks of T3*

References

[1] Chiew P. S. 2021. A preliminary study of third tone sandhi in Malaysian Mandarin. *Journal of Social Science and Humanities 5*(2), 1-14.

[2] Chua, C. L. (2003). *The emgerence of Singapore Mandarin: a case study of language contact*. The University of Wisconsin-Madisonl PhD thesis.

[3] Guo, X. 2017. An outline of Malaysia Huayu (马来西亚华语概说). *Global Chinese*, *3*(1), 61-83.

[4] Huang, T. 2016. *Topics in Malaysian Mandarin phonetics and phonology*. National Tsing Hua University PhD thesis.

[5] Khoo, K. U. 2014. Research on Juluan Mandarin in Malaysia: a special case study of social language variation (马来西亚“居銮华语”调查研究：一个特殊社会语言变异个案分析). In Khoo K. U. (Ed.), *A collected of papers of Malaysian Mandarin* (*马来西亚华语研究论集*), 183-207. Kuala Lumpur : Center for Malaysian Chinese Studies.

[6] Wang, X. M. 2019. On the four perspectives of Global Chinese studies: classic Chinese, dialects, Putonghua and foreign languages (“古、方、普、外”—论全球华语研究的四个视角). Global Chinese, 5*(1)*, 45-57.

[7] Wu, Z. J. 2008. *The linguistic essays of Wu Zongji* (*吴宗济语言学论文集*). Beijing: The Commercial Press.

[8] Yin. H. 2020. Acoustic study of tone 3 sandhi in Beijing and Taiwan Mandarin. *International Journal of English Linguistics*, *10(2)*, 1923-8703.

Tonal Variation of Southern Min Dialect: A Case Study of Klang Hokkien in Malaysia

Poh Shin Chiew & Meng Huat Chau[2]
[1 2] *University Malaya, Faculty of Languages and Linguistics*
chiewpohshin@um.edu.my, chaumenghuat@yahoo.co.uk

Tone plays a significant role in the phonological system of various Chinese varieties, alongside vowel and consonant segments. Understanding the evolution of ancient tones, which trace back to Middle Chinese (MC) around AD 600, holds crucial importance in the field of Chinese dialectology. Many non-Mandarin dialect groups, such as Min and Yue, have retained the MC tonal system and preserved ancient stop codas. Nevertheless, Chinese tone categories are currently undergoing merging, leading to increased tonal variation, especially in multilingual societies like Malaysia and Singapore.

Previous studies have observed that certain varieties of the southern Min dialect are experiencing tonal integration and stop coda weakening in Singapore and Penang, Klang of Malaysia [1] [2] [3]. The primary focus of this study is to further explore the tonal variation in Klang Hokkien (KH), which is spoken in the Klang district near Malaysia's capital. It aims to examine how the younger generation produces these lexical tones and gain insights into the dynamic changes of its tonal system with compared to the older generation [3].

The KH speech data were collected remotely due to pandemic-induced movement restrictions, which also provide participants with the flexibility to record at their convenience. All participants obtained their consent, and completed a background survey, along with self-recordings, following specific guidelines on a designated site. The recordings were saved in a non-lossy format and underwent thorough screening before further data analysis. For the purposes of this study, partial speech samples from the database were analyzed. This study comprised 15 female participants, all of whom were born and raised in the Klang district and spoke the southern Min dialect. The participants had an average age of 22 years (SD=4).

A list of 39 monosyllabic morphemes was used to examine the KH lexical tones in isolation, commonly referred to as citation forms. The list includes eight MC tones (T1-T8), which follows the standard method in Chinese dialectology research and is consistent with the previous KH study for improved comparability. All syllables consist of onset and rhyme, specifically CV and CVN for smooth tones (T1-T6), while CVP ended with an oral or glottal stop coda for check tones (T7-T8). A total of 487 valid tokens for analysis, excluding those that did not read out and read in Mandarin. The pitch and duration of each rhyme were examined by referring to spectrograms and auditory perception. In order to provide a clearer and more concise representation of the tonal variation, the pitch features have been categorized into low (L), mid (M), and high (H) levels instead of using Chao's 5-point tone letters.

The initial findings revealed that the citation form of the younger generation in the current study was largely similar to the older generation (Table 1), but it exhibited some variations. Referring to Figures 1 and 2, the realisations of T1, T2, T3, and T7 echoed the previous tone system but showed slight differences in pitch level and contour. The merging of T4, T5, and T6 was observed as stated, but around 20%-30% of realisations conflicted with other smooth tones. The T8 realisations showed a high degree of variation, with the emergence of a high falling tone similar to T7 and a mid-rising tone similar to T2. The study also confirmed the distinction of long-short features between smooth tones and checked tones. However, the rhyme duration of CVP ended with a glottal stop coda exhibited noticeable lengthening (median=200ms), which extended from the weakening and reduction of the stop coda (Figure 3). In terms of comprehensive pitch and duration features, the variation of T8 in KH was the most obvious in this study. In summary, the current findings show that the distinction between smooth and checked tones persists among the younger generation; however, tone reduction is particularly noticeable, especially in the T8 checked tone.

Table 1: *Klang Hokkien citation tones produced by the older generation ([3])*

MC Tone Categories	T1	T2	T3	T4	T5	T6	T7	T8
Notation (Chao's 5-point tone letters)	33	24	53		31		53	33

Proceedings of the Second International Conference on Tone and Intonation, pages 105–106
November 18–20, 2023.

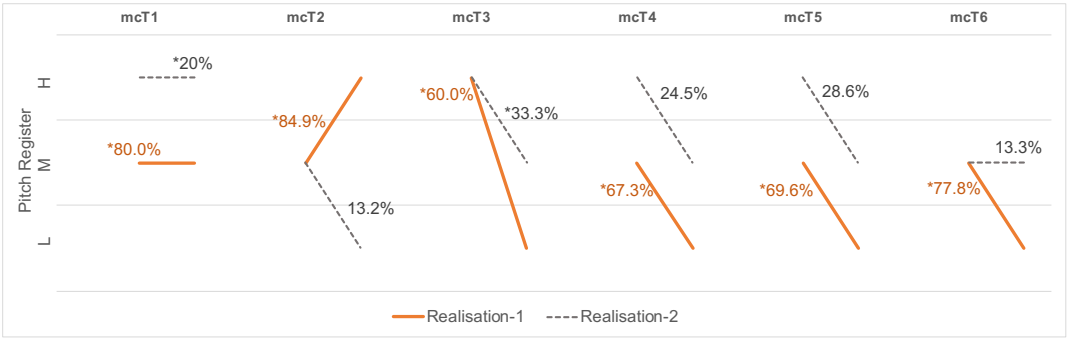


Figure 1: *Pitch realisation of lexical tone in Klang Hokkien : smooth tones(L=low, M=mid, H=high; *mark indicated the realisation similar to the citation forms reported in Chiew, 2019 [3])*

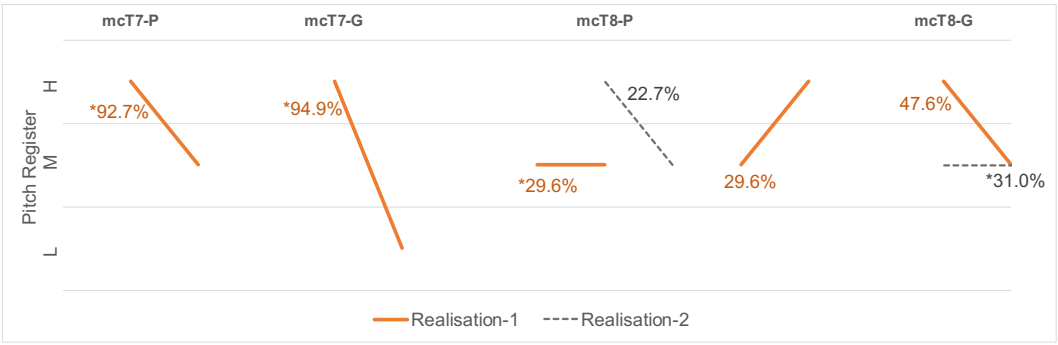


Figure 2: *Pitch realisation of lexical tone in Klang Hokkien: checked tones*

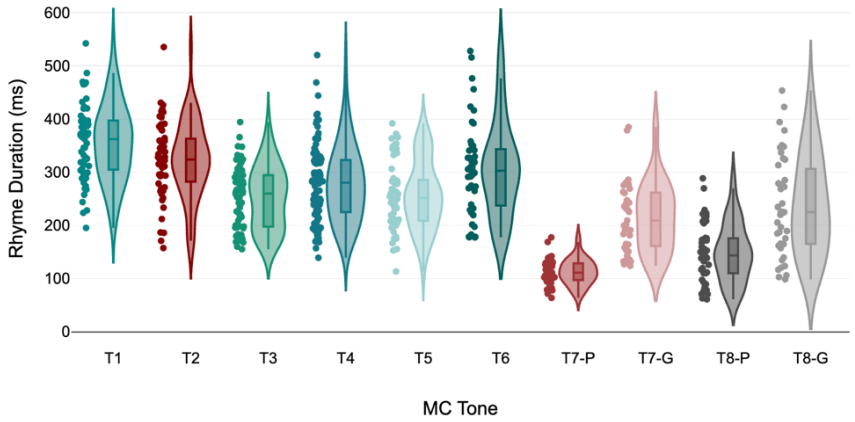


Figure 3: *Rhyme duration (msec) of smooth tones and checked tones in Klang Hokkien (P=syllable ended with oral stop coda, G=syllable ended with glottal stop coda)*

References

[1] Zhou, C.J., & Chew, C.H. 2002. *A dictionary of Southern Min dialect in Singapore (新加坡闽南话词典)*. Beijing: China Social Sciences Press.

[2] Chuang, C. T., Chang, Y. C., & Hsieh, F. F. (2013). Complete and not-so-complete tonal neutralization in Penang Hokkien. In Lee, W.S. (Ed.), *Proceedings of the International Conference on Phonetics of the Languages in China*. Hong Kong: City University of Hong Kong, 54-57.

[3] Chiew, P.S. 2019. Lexical Tone and Tone Sandhi in Klang Hokkien. In Yap, N.T., & Setter, J. (Eds.), *Speech Research in a Malaysia Context.* Serdang: UPM Press, 99-119.

Pitch Accents, Boundary Tones and Tune Compositionality

Stella Gryllia[1] & Amalia Arvaniti[1]
[1]*CLS, Radboud University,*
stella.gryllia@ru.nl, amalia.arvaniti@ru.nl

A recurrent issue in the study of intonation relates to whether contours should be treated as gestalts [1, 2] or composites of independent elements [3, 4]. We contribute to this debate by examining a corpus of wh-questions (N = 2135) which were elicited from 18 Greek speakers using a discourse completion task (DCT). DCTs involved two scenarios that participants heard and saw on screen: Scenario A presented a situation ending with an information-seeking question (1a); Scenario B presented a situation in which the wh-question was used as an implicit statement (1b); cf. [5, 6]. The expected tune in response to Scenario A is autosegmentally analyzed as a L*+H pitch accent on the utterance-initial wh-word, followed by a L- phrase accent and a H% boundary tone. The expected tune in response to Scenario B is analysed as L+H* L- L% [5].

(1a) Scenario A, leading to genuine question *: Your best friend went away for a few days' vacation and has asked you to look after her house. She has given you instructions on everything, but suddenly you remember that she hasn't told you anything about her flowers. You call her and say*:
[ˈpote na poˈtiso ta luˈluđja] "When should I water the flowers?"
(1b) Scenario B, leading to implicit statement: *When your dad does the grocery shopping, he gets stingy and does not buy enough of anything. One day you have many friends at home and run out of milk although you had warned him about it. When a friend asks for a double cappuccino, you tell your dad:*
[meˈti na tu ˈftçakso ˈtora toŋ gapuˈtsino] "What will I make him cappuccino with now?"

We used functional principal component analysis (FPCA), a data-driven, dimension-reduction method to analyse the pitch contours. In FPCA curves are modelled as B-splines and based on this modelling FPCA returns the dominant modes of curve variation called Principal Components (PCs). By definition, each PC presents an independent mode of variation. Thus, if curve changes associated with the tonal elements posited by AM are captured by different PCs, this is evidence that these elements are independent of each other which supports compositionality. This is confirmed in our analysis. As shown in Figure 1, the pitch movement associated with each of the posited tonal elements is captured by a different PC: PC1 captures the shape of the fall (as a consequence of peak height and alignment), PC2 captures the extent of the initial rise and subsequent peak alignment (the difference between L*+H and L+H*), and PC4 captures the difference between a final rise (H%) and flat F0 (L%).

The scores of the first four PCs were statistically analysed in R (R: 4.2.1 [7]) using linear mixed effects models; base model formula lmer (PC#~ SCENARIO +(1| ITEM)+(1+CONTEXT| SPEAKER), data = data, REML=FALSE) (lme4: 1.1.26 [8]). We focus on PC1, PC2 and PC4 because conditional R^2 was the highest for these three PCs; this measure indicates that they explain more variance than PC3 (not discussed here), see Table 1. Specifically, PC1 captures 39.1% of the curve variability; tunes uttered after scenario A are significantly different from tunes uttered after scenario B, see Table 1. As shown in Figure 1, lower scores (blue curves) fall more abrupt compared to higher scores (red curves) that fall more smoothly. Conditional R^2 indicates that 24% of the variance was due to the combined effects of SCENARIO, SPEAKER, and ITEM. PC2 captures 31.0% of the curve variability; the tunes differ significantly between the two scenarios; higher scores (red curves) lead to earlier peak relative to lower scores (blue curves). Conditional R^2 indicates that 28% of the variance was due to the combined effects of SCENARIO, SPEAKER, and ITEM. PC4 captures 8% of the curve variability; there is a significant difference between the two scenarios. Higher scores (red curves) lead to a rise relative to lower scores (blue curves) that remained flat. Conditional R^2 indicates that 23% of the variance was due to the combined effects of SCENARIO, SPEAKER, and ITEM.

Given that each PC presents an independent mode of variation, we can conclude that tunes are composites of independent elements; PC1 captured the shape of the fall, PC2 captured the extent of the initial rise and the peak alignment and PC4 captured differences related to the boundary tone. This provides prima facie evidence for tune compositionality.

Proceedings of the Second International Conference on Tone and Intonation, pages 107–108
November 18–20, 2023.

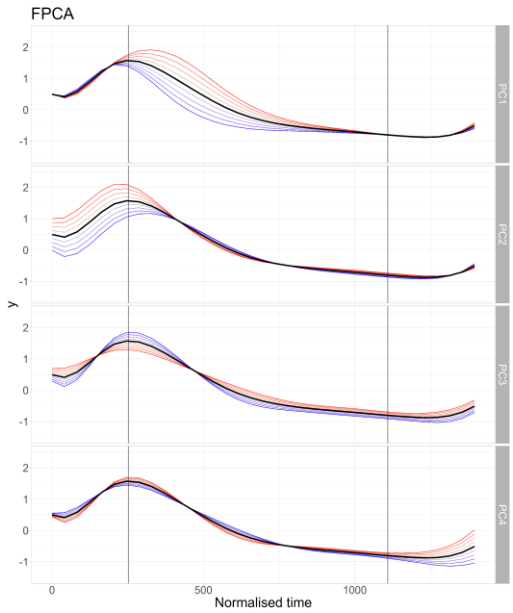


Figure 1: PC1, PC2, PC3, & PC4 curves modelling the wh-question corpus (solid black line = mean curve; red curves = higher PC scores; blue curves = lower PC scores); the first vertical line represents the offset of the first vowel after the accented vowel of the wh-word, while the second vertical line represents the onset of the stressed vowel of the last word (landmark registration).

PC1	**Est.**	**SE**	**df**	**t-value**	**Pr(>|t|)**
(Intercept)	3.426	1.505	21.288	2.29	*
Scenario: B	-6.855	1.515	18.004	-4.525	***
Marginal R^2: 0.08		*Conditional R^2: 0.24*			
PC2	**Est.**	**SE**	**df**	**t-value**	**Pr(>|t|)**
(Intercept)	-3.65	1.13	22.95	-3.23	**
Scenario: B	7.31	0.41	2108.0	18.05	***
Marginal R^2: 0.11		*Conditional R^2: 0.28*			
PC3	**Est.**	**SE**	**df**	**t-value**	**Pr(>|t|)**
(Intercept)	-0.82	0.53	23.49	-1.544	0.136
Scenario: B	1.633	0.46	17.99	3.585	**
Marginal R^2: 0.02		*Conditional R^2: 0.09*			
PC4	**Est.**	**SE**	**df**	**t-value**	**Pr(>|t|)**
(Intercept)	0.91	0.71	19.497	1.277	0.217
Scenario: B	-1.81	0.67	18.014	-2.694	*
Marginal R^2: 0.03		*Conditional R^2: 0.23*			

Table 1: Results of LMEMs for PC1, PC2, PC3, PC4 and PC5; $p < .05$ = *, $p < .01$ = **, $p < .001$ = ***; the marginal (an estimate of the proportion of variance explained by fixed factors) and conditional (an estimate of the proportion of variance explained by the combined effect of fixed and random factors) R^2 of each model are presented below the relevant model in italics

References

[1] Hirst, D., & Di Cristo, A.1998. A survey of intonation systems. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems a Survey of Twenty Languages*, 1-44.

[2] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3-4), 220-251.

[3] Pierrehumbert, J. & Hirschberg, J. B. 1990. The meaning of intonational contours in the interpretation of discourse. *Intentions in Communication*, 271-311.

[4] Ladd, D. R. 2008. *Intonational Phonology*. Cambridge University Press.

[5] Baltazani, M., Gryllia, S., & Arvaniti, A. 2020. The Intonation and Pragmatics of Greek wh-Questions. *Language and Speech*, 63(1), 56–94. https://doi.org/10.1177/0023830918823236

[6] Gryllia, S., Baltazani, M., Arvaniti, A. 2018. The role of pragmatics and politeness in explaining prosodic variability. *Proc. 9th International Conference on Speech Prosody* 2018, 158-162, https://doi.org/10.21437/SpeechProsody.2018-32

[7] R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: http://www.r-project.org/.

[8] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 1.1, 1–48. [Online]. Available: https://www.jstatsoft.org/v067/i01

Exploring the Utility of Automatically Generated References for Assessing L2 Prosody

Mariana Julião[1,2], Helena Moniz[1,3] & Alberto Abad[1,2]
[1]*INESC-ID,* [2]*IST,* [3]*CLUL - Lisbon*
mariana.juliao@inesc-id.pt

Prosody assessment is a challenging yet essential aspect of effective communication, particularly in the context of spoken language. Prominence, denoting the emphasis placed on specific words within a phrase, significantly contributes to speech intelligibility. While modern e-learning platforms have recently incorporated prosody assessment, the existing approaches remain simplistic, relying mainly on imitation tasks. In this study, we explore the possibility of a novel system designed to evaluate the quality of spoken sentences in terms of prominence. Our system, Goodness of Prosody (GoP) [Fig1] comprises two main branches: an acoustic-based branch for prosody classification, which identifies prominent words, and a text-based branch for prosody prediction, which determines the expected prominent words.

The current version of GoP utilizes forced alignment, leveraging readily available text data. However, future iterations can integrate an Automatic Speech Recognition (ASR) module to extend its usability in scenarios where text availability is limited. For prosody classification, we trained a prosodic event detector using English radio news speech (BURNC [1]) to detect pitch accents. When selecting an acoustic classifier, our primary consideration was not only its overall performance but also its independence from the speaker's proficiency levels. We compared classifiers based on wavelets, a CNN-only model, and a CNN+LSTM model. Our results demonstrated that the CNN+LSTM model outperformed the others and exhibited minimal variation across different proficiency levels.

To predict prosody from text, we followed the example of [4], who released a full annotation of LibriTTS on prominence and boundaries after classifying the corresponding speech. We annotated two large corpora (LibriTTS [5] and VCTK [6]) by classifying them with our best acoustic model, previously described. With this, we have generated labels for the text. Also based on the work of [4], we trained a network to label text for prominence. This consisted of a fully connected layer which got BERT [7] embeddings as input, and which learns the prominence labels previously assigned by the acoustic classifier. The full process is described in [Fig2].

One of the primary challenges in prosody assessment is the wide range of possible variations in speech production for the same text-intention pairs. This variability also applies to prominence, as it is often left to the speaker's sensitivity to decide which words to emphasize. To address this issue, we developed different text-based classifiers to predict the accentedness of each word in an utterance, providing reference standards for the assessment of spoken productions. We trained four different text-classifiers, each trained on a distinct corpus or corpus partition. By generating multiple references for the same utterance and selecting the best one, we aimed to account for the diverse possibilities.

Finally, we examined the agreement between the L2-corpus labels and the generated text references. In our evaluation framework, LeaP [8], the speaker proficiency levels were not absolute; rather, they indicated whether the speaker had undergone prosody training or language immersion. Consequently, we limited our comparisons to "before" and "after" categories, excluding comparisons with other speakers. We observed no correlation between speaker proficiency and reference matching [Tab2], regardless of utterance length, except for native speakers who tended to exhibit more mismatches. In particular, we notice that after going abroad the accuracy improved approximately 2%, but when comparing before and after a prosody training course, the accuracy worsened 4%. Further investigation of utterance and reference mismatches revealed that native speakers often over or under-emphasized words in unexpected ways, highlighting the subjective nature of prominence. Additionally, we identified some shortcomings in the acoustic classifier, indicating areas for improvement.

This study contributes to the advancement of prosody assessment for e-learning platforms by proposing the GoP system, which combines acoustic-based classification and text-based prediction to evaluate the prominence of spoken sentences. The findings shed light on the challenges associated with prosody assessment, particularly in L2 speech, and provide insights into the potential enhancements needed for future iterations of the GoP system.

Proceedings of the Second International Conference on Tone and Intonation, pages 109–110
November 18–20, 2023.

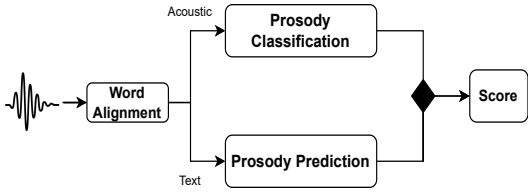


Fig1: Goodness of Prosody model.

Level		Acc.
going abroad	before (e1)	0.889
	after (e2)	0.905
prosody training course	before (c1)	0.861
	after (c2+c3)	0.819
superlearner	sl	0.861
native	na	0.797

Tab2: Accuracy between the closest generated reference and the manual labeling of prominence. Average per utterance, speaker and level.

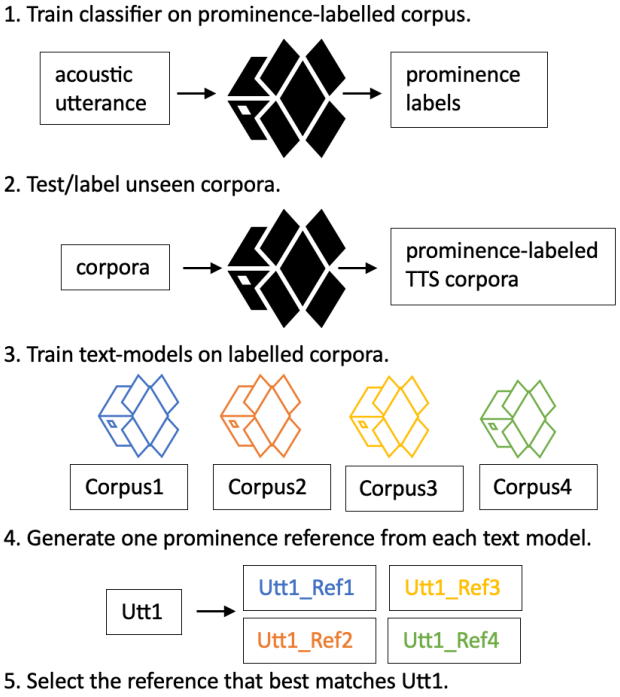


Fig2: Reference generation process.

References

[1] Ostendorf, M., Price, P.J. & Shattuck-Hufnagel, S. 1995. The Boston University radio news corpus. *Linguistic Data Consortium*, pp.1-19.

[2] Suni, A., Šimko, J., Aalto, D., & Vainio, M. 2017. Hierarchical representation and estimation of prosody using continuous wavelet transform. *Computer Speech & Language*, *45*, 123-136.

[3] Nielsen, E., Steedman, M., & Goldwater, S. 2020. The role of context in neural pitch accent detection in English. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 7994-8000).

[4] Talman, A., Suni, A., Celikkanat, H., Kakouros, S., Tiedemann, J., & Vainio, M. 2019. Predicting Prosodic Prominence from Text with Pre-trained Contextualized Word Representations. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics* (pp. 281-290).

[5] Zen, H., Dang, V., Clark, R., Zhang, Y., Weiss, R.J., Jia, Y., Chen, Z. & Wu, Y. 2019. LibriTTS: A Corpus Derived from LibriSpeech for Text-to-Speech. *Proc. Interspeech 2019*, pp.1526-1530.

[6] Veaux, C., Yamagishi, J. & MacDonald, K. 2017. CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit. *University of Edinburgh. The Centre for Speech Technology Research (CSTR)*, *6*, p.15.

[7] Kenton, J. D. M. W. C., & Toutanova, L. K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT* (pp. 4171-4186).

[8] Gut, U. 2012. The LeaP corpus: A multilingual corpus of spoken learner German and learner English. *Multilingual corpora and multilingual corpus analysis*, *14*, 3-23.

The Prosodic Profile of Black Mountain Mönpa

Gwendolyn Hyslop
The University of Sydney
gwendolyn.hyslop@sydney.edu.au

Black Mountain Mönpa is an under-described Tibeto-Burman language of central Bhutan, spoken by approximately 500 speakers. The language has been difficult to classify within the Tibeto-Burman family, with van Driem (1995) first assessing the language to belong to the East Bodish branch (close to Tibetan) and then later determining it to be an isolate within the family (van Driem 2011). Hyslop (2016) confirms several unusual phonetic features – found primarily in the lexical domain of plants – offering further support to the isolate status of the language. Gerber (2020) identifies some phonological features that group Black Mountain Mönpa with other languages in the region as belonging to an old sprachbund. This talk presents a prosodic overview of Black Mountain Mönpa, with an aim to further our understanding of the historical development of the region while also contributing to prosodic typology. Data for this analysis come from a few months' fieldwork in 2010 and 2023. There are no other publications pertaining to the language.

Black Mountain Mönpa has a rich set of phonemic contrasts. At the purely suprasegmental level, the language contrasts high versus low tone in monosyllables, as evidenced by the near minimal pair shown in (1) and (2). Current data suggest the contrast is only made following sonorant consonant onsets. A handful of words show phonemic nasalization of vowels, as evidenced by (3). Vowels may be long or short, as shown in (4). Data suggest the contrast is only available in open syllables.

In addition to tone, Black Mountain Mönpa displays a phonemic glottalization which can occur in the first syllable, as in (5) or in a later syllable, as in (6). While we represent this with a phonemic glottal stop, acoustically this contrast is usually realized as creaky voice across the entire syllable; see Figure 1.

Multisyllabic words can show initial stress, as in (7), or final stress, as in (8). Many of the words that fall into the latter category contain an initial vowel only, suggestive of the sesquisyllabic syllable type often associated with Austroasiatic languages (Matisoff 1973); however, instead of a schwa, this vowel is a fully realised low, back vowel.

In summary, Black Mountain Mönpa shows a prosodic profile that is also unique in the regional context. The apparent word tone on initial sonorant-initial only syllables appears to be like the tonal system of East Bodish languages, as does the minimal vowel length contrast (e.g. Hyslop 2017:§2-3) while nasalized vowels are found in Dzongkha, a Tibetic language spoken west of the region (e.g. van Driem 1998). Both Dzongkha and the East Bodish languages also have (exclusively) word-initial stress. Neither word-final stress nor glottalized syllables are found in adjacent languages, to our knowledge, though glottalized segments do occur in Dzongkha and other languages in Bhutan.

(1)	*lé*	'catch'	*là:*	'come'
(2)	*ɲí*	'seven'	*ɲè*	'fish'
(3)	*dõ*	'hole'	*rò*	'chase'
(4)	*ço*	'SFP'	*ço:*	'sichuan pepper'
(5)	*ˈhoʔma*	'3.SG.MSC'		
(6)	*sɔχoˈlaʔ*	'leafy green type'		
(7)	*ˈkɣgɣ*	'hen'		
(8)	*aˈpɛŋ*	'grandmother'	*aˈpʰøt*	'cotton'

Proceedings of the Second International Conference on Tone and Intonation, pages 111–112
November 18–20, 2023.

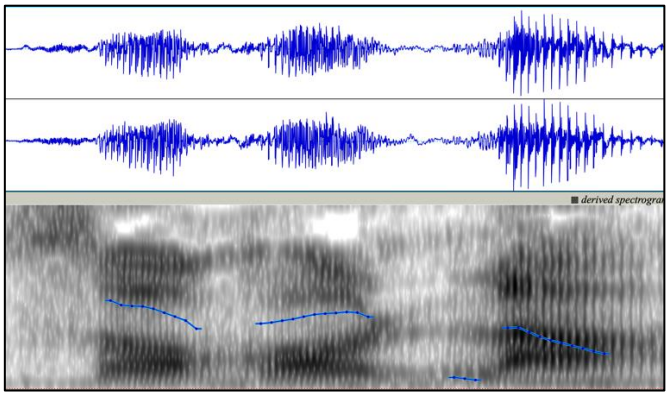


Figure 1: *Wave and spectrogram of* sɔχoˈlaʔ, *showing glottalised final syllable*.

References

[1] Driem, George van. 1995. “Black Mountain Verbal Agreement Morphology, Proto-Tibeto-Burman Morphosyntax and the Linguistic Position of Chinese.” In *New Horizons in Tibeto-Burman Morphosyntax*, edited by Yoshio Nishi, James A. Matisoff, and Yasuhiko Nagano, 229–59. Senri Ethnological Studies 41. Osaka: National Museum of Ethnology.

[2] Driem, George van. 1998. *Dzongkha*. Leiden, The Netherlands: Research CNWS, School of Asian, African, and Amerindian Studies.

[3] Driem, George van. 2011. “Tibeto-Burman Subgroups and Historical Grammar.” *Himalayan Linguistics* 10 (1): 31–39.

[4] Gerber, Pascal. 2020. “Areal Features in Gongduk, Bjokapakha and Black Mountain Mönpa Phonology.” *Linguistics of the Tibeto-Burman Area* 43 (1): 55–86. https://doi.org/10.1075/ltba.18015.ger.

[4] Hyslop, Gwendolyn. 2016. “Worlds of Knowledge in Central Bhutan: Documentation of ’Olekha.” *Language Documentation & Conservation* 10: 77–106.

[5] Hyslop, Gwendolyn. 2017. *A Grammar of Kurtöp*. Languages of the Greater Himalayan Region 18. Leiden: Brill.

[6] Matisoff, James. 1973. “Tonogenesis in Southeast Asia.” In *Consonant Types and Tone*, edited by Larry Hyman, 72–95. Southern California Occasional Papers in Linguistics. Los Angeles: Linguistics program, University of Southern California.

Exploring Rhythmic Patterns in Deori and Mising using Read Speech Data

Krisangi Saikia & Shakuntala Mahanta
Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati, INDIA
krisangisaikia@iitg.ac.in, *smahanta@iitg.ac.in*

This study explores the temporal organization and rhythmic characteristics of two Tibeto-Burman languages spoken in Assam, India: Deori and Mising. Deori, belonging to the Boro-Garo branch, is an endangered language on the brink of obsolescence. In contrast, Mising, from the Tani branch, is classified as a low-resourced language. Understanding the speech rhythm of these languages is essential for comprehending language perception, comprehension, and production. By investigating the rhythmic properties of Deori and Mising, this research sheds light on the unique temporal patterns that shape these languages despite their varying degrees of vulnerability and available resources.

The data for the current study comprises scripted sentences (translated versions of the story *The North Wind and the Sun* in the respective languages). Sixteen native speakers (8 each from Deori and Mising), aged between 21 and 36 years, participated in two production experiments. Subjects comprise an equal number of males and females for both languages. Each participant was asked to produce the story four times, ensuring a natural speech rate and intonation pattern. The best three repetitions produced by each speaker were considered for final analysis. The translated story comprises roughly 11 sentences for each language with varied syllable lengths (ranging between 6 to 12 syllables per sentence). The recorded speech data were annotated at the phoneme level in Praat 6.1.06, delineating vocalic and consonantal intervals based on auditory and acoustic cues according to standard segmentation criteria [Turk et al. (2006), Frota & Vigário (2001)]. The *Correlatore* program (version.2.3.4) was used to extract different rhythmic metrics, including Cmean, Vmean, %V, ΔC, ΔV, Varcos (Varco-V, Varco-C), and the PVI (nPVI, rPVI) from the annotated speech data. The *speaking rate* also influences rhythm measures. The speech rate is calculated in terms of the time taken syllables per second and segments per second. The values of these matrices were plotted against each other using the *ggplot* package (Figures 1 and 2, for example) in the R software (version 4.2.2 (R Core Team, 2022).

To validate the visual observations of these plots, we conducted a *Pearson correlation* test to examine how utterance length, speaking tempo, and rhythm metrics interact. It's worth noting that some researchers have argued that the rhythm metrics proposed in the literature are inadequate for classifying languages into distinct rhythmic classes [1]. Researchers have also argued in favor of different rhythm matrices and multiple methods of calculating those metrices. The *correlation* test enabled us to identify the most consistent and highly correlated rhythm metrices. Rate of articulation, in terms of segments per second, has a negative correlation on Deori and nPVI-V, ΔV. However, there is a robust inverse relationship between Varco-C for both length and syllable per second. In the case of Mising, there is a significant negative correlation between the rate of articulation and the values of ΔV, ΔC, and rPVI-C.

Results also indicate that the rhythm component pairs, viz., (ΔV, varco-V), (ΔV, %V), (ΔC, %V), (ΔC, varco-C), (ΔC, rPVI-C), (nPVI-V, rPVI-V), (rPVI-C, varco-C), (nPVI-V, varco-V) and (nPVI-V, %V) are highly correlated to each other for Deori; whereas, the rhythm component pairs such as (ΔV, varco-V), (ΔV, %V), (ΔC, varco-C), (nPVI-C, rPVI-C), (nPVI-V, rPVI-V), (ΔC, rPVI-C), (ΔC, %V), and (nPVI-V, varco-V) are highly correlated to each other for Mising. The average values of different rhythm matrices are compared with those of rhythm correlates for the languages examined by [2] [see Table 1] indicates that Deori and Mising cluster with syllable-timed rhythm classes. Similarly, the values proposed by [4] [see Table 2] also indicate a syllable-timed rhythm class for Deori and slightly moving towards stress-timed rhythm class for Mising. In Figure. 1. we plot the values of nPVI-V and rPVI-C with other languages [4]. It can be seen that Deori is tends to cluster with the syllable-timed language (French). Whereas the nPVI-V values for Mising are higher and showing the tendency of moving towards stress-timed language (English). We can see in Figure.2 the %V and ΔC, the results show that Deori and Mising are very close to each other and likely to form a cluster and are placed in between prototypical syllable-timed (Spanish,French and Catalan) and Mora-timed (Japanese).

Proceedings of the Second International Conference on Tone and Intonation, pages 113–114
November 18–20, 2023.

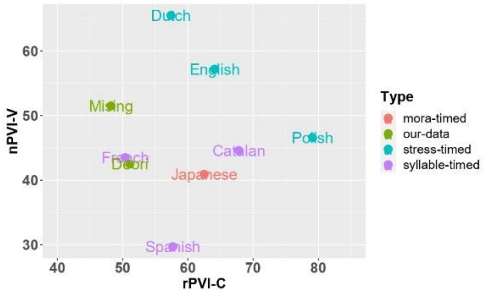


Figure 1 : *rPVI-C and nPVI-V for Deori and Mising read speech comparison to analyzed by [4]*

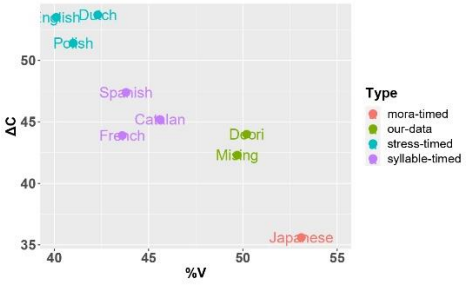


Figure 2 : *ΔC and %V for Deori and Mising read speech comparison to langauges analyzed by[2]*

Table 1: *The values of Rhythm correlates for [British English (Stressed-timed), French (Syllable-timed) and Japanese (Mora-timed)] as proposed by [4] are compared with the values for Deori and Mising.*

Language	%V	Vnpvi	Crpvi	ΔC	ΔV
British English	41.1	57.2	64.1	56.7	46.6
French	50.6	43.5	50.4	42.4	35.5
Japanese	45.5	40.9	62.5	55.5	53.0
Deori	50.2	42.5	51.1	44	37.7
Mising	49.7	51.5	48.2	42.3	43.9

Table 2: *The values of Rhythm correlates for [British English, Polish, Dutch (Stressed-timed), Spanish, French, Catalan (Syllable-timed) and Japanese (Mora-timed)] along with their standard deviation as proposed by [2] are compared with the values for Deori and Mising.*

Language	%V(SD)	ΔV(SD)	ΔC(SD)
English	40.1 (5.4)	4.64 (1.25)	5.35 (1.63)
Polish	41.0 (3.4)	4.23 (0.67)	5.33 (1.18)
Dutch	42.3 (4.2)	3.11 (0.93)	5.37 (1.5)
Spanish	43.8 (4.0)	3.32 (1.0)	4.74 (0.85)
French	43.6 (4.5)	3.78 (1.21)	4.39 (0.74)
Catalan	45.6 (5.4)	3.68 (1.44)	4.52 (0.86)
Japanese	53.1 (3.4)	4.02 (0.58)	3.56 (0.74)
Deori	50.2 (5.7)	3.77 (1.64)	4.41 (0.75)
Mising	49.7 (4.8)	4.39 (1.22)	4.23 (0.68)

References

[1] Arvaniti, A. 2012. "The usefulness of metrics in the quantification of speech rhythm," *Journal of Phonetics*, vol. 40, no. 3, pp. 351–373.

[2] Ramus, F., Nespor, M. & Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. Cognition, 73/3, 265-29

[3] Dellwo, V. 2006. Rhythm and speech rate: A variation coefficient for deltaC. Language and Language Processing: *Proceedings of the 38th Linguistic Colloquium*, Piliscsaba 2003, ed. by Pawel Karnowski Imre Szigeti, 231–241. Frankfurt: Peter Lang

[4] Grabe, E. & Low, E.L. 2002. Durational variability in speech and the rhythm class hypothesis. In: Gussenhoven, C., Warner, N. (eds), *Papers in Laboratory Phonology* 7, Berlin: Mouton de Gruyter, 515-546

The Roles of F0 Range/Slope and Duration in Cueing the Mandarin Rising or Falling Tones

Wei Zhang[1] & Wentao Gu[2,3]
[1]*Department of Linguistics, McGill University, Canada*
[2]*School of Communication Sciences and Disorders, McGill University, Canada*
[3]*School of Chinese Language and Literature, Nanjing Normal University, China*
wei.zhang16@mail.mcgill.ca, wtgu@njnu.edu.cn

It has been well known that rising/falling pitch is employed to distinguish the rising (R) or falling (F) tones from the high-level (H) tone in Mandarin [1], but which F0 cue–F0 range or F0 slope–is primary or more critical to perception of these dynamic tones is still inconclusive. Since F0 range and F0 slope are closely related through the variable 'duration', the research question is equivalent to which F0 cue is associated with a perceptual boundary that is less dependent on duration. Also, duration itself may serve as a secondary cue due to intrinsic durational differences among isolated syllables of the four tones, specifically, the F tone is the shortest, while the R tone has a duration comparable to, or marginally longer than, the H tone [2].

To elucidate this issue, first of all, we took the H-F tonal contrast as the test case (since H-R and H-F are basically symmetric in F0), and recruited 30 native speakers of Mandarin (15F, 15M) to conduct two-alternative forced choice (2AFC) identification tests on two types of two-dimensional H-F tonal continua, one of which, as shown in Fig. 1A, varied along F0 range and duration ('F0 range continuum'), while the other, as shown in Fig. 1B, varied along F0 slope and duration ('F0 slope continuum'). The identification rates of the F tone for each continuum are shown in Fig. 2. Analysis with mixed-effects logistic models revealed a significant interaction between F0 slope and duration in the F0 slope continuum, but not between F0 range and duration in the F0 range continuum. Moreover, at each duration step we calculated sharpness and position of the perceptual boundary, of which the ratios relative to the values at 100 ms, as illustrated in Fig. 4, are approximated by linear regression to indicate the rates of change with duration. Results suggest that F0 range is the primary cue as it results in a more robust (less duration-dependent) perceptual boundary than F0 slope. Meanwhile, position of the perceptual boundary in the F0 range continuum is not fully independent of but shifting towards the F tone mildly with duration, suggesting that duration (or equivalently, F0 slope) plays a secondary role in identifying the H-F tonal contrast.

There are two ways to interpret this supplementary effect. On the one hand, aside from the primary cue of F0 range, there might be a potential threshold in F0 slope to ensure an identifiable falling pitch. If so, the effect should apply almost equally to the H-F and H-R contrasts – thus the perceptual boundary of the H-R contrast will shift towards the R tone with a longer duration. On the other hand, this effect may be specific to the H-F contrast, mainly attributed to the effect of duration itself (like the role of duration in identifying vowels /ɪ/-/i/ in English) – the F tone is inherently shorter than the H tone, resulting in a perceptual boundary closer to the F tone with a longer duration. If this is true, the effect will be missing or even reversed in the H-R contrast because the R tone has a duration comparable to, or marginally longer than, the H tone.

To further clarify which interpretation better accounts for the supplementary effect of duration in tone identification, we recruited another set of 30 native speakers of Mandarin (21F, 9M) to conduct 2AFC identification tests on an H-F and an H-R tonal continua which were both two-dimensional F0 range continua, with shared step sizes in F0 range and duration.

Figure 3 shows the identification rates of the F and R tones for the respective F0 range continua. Analysis with mixed-effects logistic models revealed significantly negative effects of duration in both continua, with a larger effect size in the H-F than in the H-R continuum. Moreover, we calculated the perceptual boundary position at each duration step, and the ratios relative to the values at 100 ms are illustrated in Fig. 5. The results of linear regression indicate that the boundary shifts mildly towards the F tone when duration is longer (with a significant correlation) in the H-F tonal continuum (consistent with the line in Fig. 4), but no significant correlation is observed in the H-R tonal continuum. This supports the second interpretation that the supplementary role of duration (or F0 slope) in tone identification is specific to the H-F tonal contrast due to the intrinsic shorter duration of the F tone, but not applicable to the H-R tonal contrast.

In summary, F0 range instead of F0 slope is the primary cue to identify the dynamic R/F tones from the level tone, while duration plays a secondary role in identifying the H-F instead of the H-R tonal contrast. This study will have broader implications for clarifying the roles of differing cues of tones in other languages.

References

[1] Gandour, J. T. 1983. Tone perception in Far Eastern languages. *J. Phon.* 11(2), 149–175.
[2] Lin, M. 1988. Perceptual cues for tones in Standard Chinese, *Zhongguo Yuwen (Chinese Language)* 204, 182–193.

Proceedings of the Second International Conference on Tone and Intonation, pages 115–116
November 18–20, 2023.

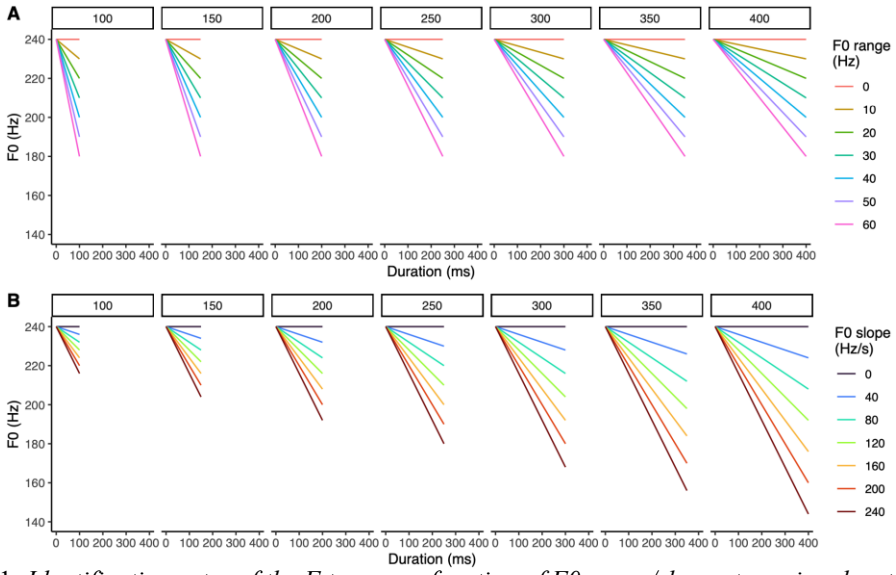


Fig. 1: *Identification rates of the F tone as a function of F0 range/slope at varying durations.*

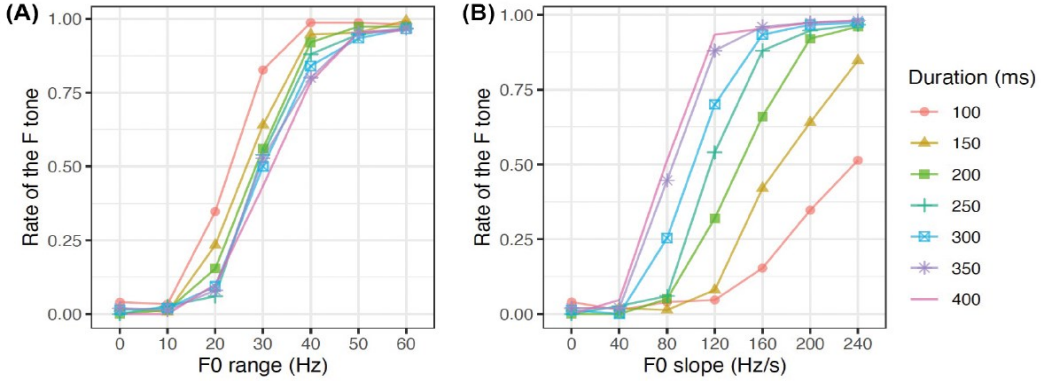


Fig. 2: *Identification rates of the F tone as a function of F0 range/slope at varying durations.*

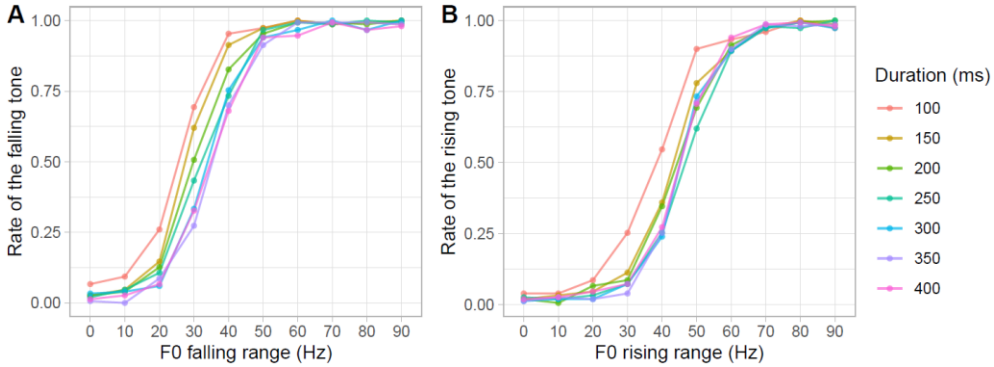


Fig. 3: *Identification rates of (A) the F tone and (B) the R tone as a function of F0 range at varying durations.*

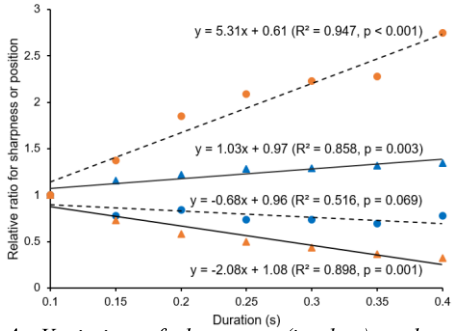


Fig. 4: *Variation of sharpness (in dots) and position (in triangles) of the perceptual boundary at all duration steps in the H-R tonal continua. The solid/dashed lines indicate the results of linear regression for the F0 range/slope continuum.*

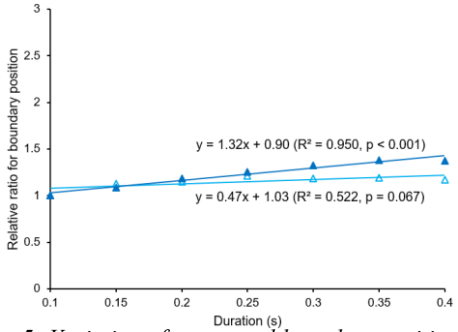


Fig. 5: *Variation of perceptual boundary position at all duration steps in the H-F (in solid triangles) and the H-R (in hollow triangles) tonal continua. The solid lines indicate the results of linear regression.*

Production of Mandarin Tones in Patients with Parkinson's Disease

Wentao Gu[1,2], Ping Fan[1,3], and Weiguo Liu[4]
[1]*School of Chinese Language and Literature, Nanjing Normal University, China*
[2]*School of Communication Sciences and Disorders, McGill University, Canada*
[3]*School of Liberal Arts, Anhui Normal University, China*
[4]*Nanjing Brain Hospital Affiliated to Nanjing Medical University, China*
wtgu@njnu.edu.cn, fpshida2010@126.com, liuweiguo1111@sina.com

Parkinson's disease (PD) is one of the neurodegenerative diseases in the middle-aged and elderly people, with typical motor impairments such as bradykinesia, hypokinesia, akinesia, muscle rigidity, and rest tremor. In addition, 70%-90% of PD patients also suffer from hypokinetic dysarthria, which may have developed for years before the appearance of obvious clinical motor symptoms and hence may be an indicator for early diagnose of PD. Hypokinetic dysarthria in PD is manifested in all dimensions of speech production. Especially, prosodic characteristics of PD speech include monotone, monoloudness, abnormal speech rate, and disfluency. In tone languages like Mandarin, monotone (i.e., a reduced pitch variability) in PD speech has been acoustically evidenced by a global analysis of F0 contours of continuous utterances [1–3], but F0 variations of lexical tones have not been specifically investigated.

Speech disorders and gait disorders are both axial symptoms (i.e., disorders of body axis) in PD, thus having some common mechanisms. Like freezing of gait when turning in PD [4], formant transitions (F2 slopes) within diphthongs in speech are slower in PD than in healthy controls [5]. These not only share the same mechanisms with monotone in PD speech, but also lead us to wonder whether local pitch transition of lexical tones is a better indicator of PD in tone language speakers than the global F0 variability, and if so, in which tonal contexts the indicator will be more effective.

Thus, the present work investigated PD's production of four tones of Mandarin, i.e., T1 (HH), T2 (LH), T3 (LL) and T4 (HL). Two groups of Mandarin-speaking participants were recruited: 13 patients with Parkinson's disease (PD) who were at the modified H&Y stage of 1-3 without dementia, depression, anxiety or other neurological diseases, and 13 gender- and age-matched healthy controls (HC). For tonal coarticulation, it is known that carryover effect is primary while anticipatory effect plays a secondary role. Therefore, for each participant, 40 monosyllabic words (10 for each of four tones), 80 disyllabic words (5 for each of 16 tonal combinations), and 20 trisyllabic words (5 for each of the four combinations: T4-**T1**-T2, T1-**T2**-T2, T1-**T3**-T4, and T4-**T4**-T1, in all of which pitch targets switch from L to H or vice versa across syllable boundaries) were recorded. Growth curve analyses with quadratic polynomials were conducted on F0 contours. The means, slopes, and curvatures of F0 contours in all target syllables were then analyzed using linear mixed-effects models.

In monosyllabic words (Fig. 1), in the latter syllables of disyllabic words (Fig. 2), as well as in the intermediate syllables of trisyllabic words (Fig. 3), the PD group showed a lower slope for T2, a lower curvature for T3, and lower absolute slope and curvature for T4 than the HC group. Results indicated smaller F0 variations in the PD group, which coincided with subjective impression on PD's monotonous voice of tone, suggesting degraded F0 manipulation in PD. Moreover, in the di- and tri-syllabic words with reversed pitch targets across syllable boundaries (i.e., L-H or H-L, as shown in red in Figs. 2–3), the PD group exhibited significantly smaller F0 shifts at syllable boundaries than the HC group, while in the disyllabic words without any change of pitch target across syllable boundaries (i.e., H-H or L-L, as shown in green in Figs. 2–3), no significant difference in F0 shift was found between the two groups.

In sum, a reduced pitch variation in PD speech of Mandarin can be better exposed in polysyllabic words with reversed pitch targets across syllable boundaries. This suggests that for speakers of tone languages that generally have faster F0 variation than non-tone languages due to the existence of lexical tones, the words in particular tone sequences can be very effective materials not only for early diagnosis of PD but also for speech therapy in PD patients.

Proceedings of the Second International Conference on Tone and Intonation, pages 117–118
November 18–20, 2023.

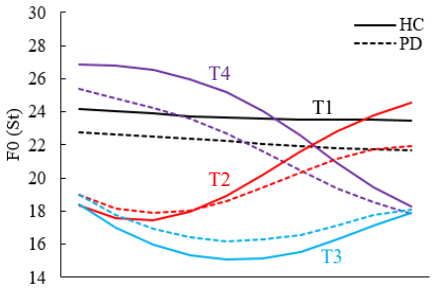


Figure 1: *F0 contours of four tones.*

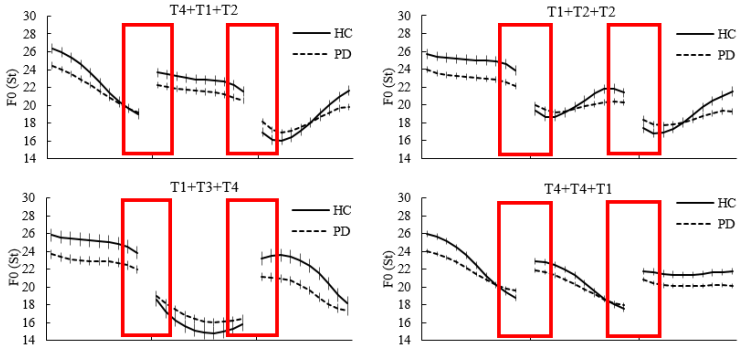


Figure 3: *F0 contours of trisyllabic words of 4 tonal combinations.*

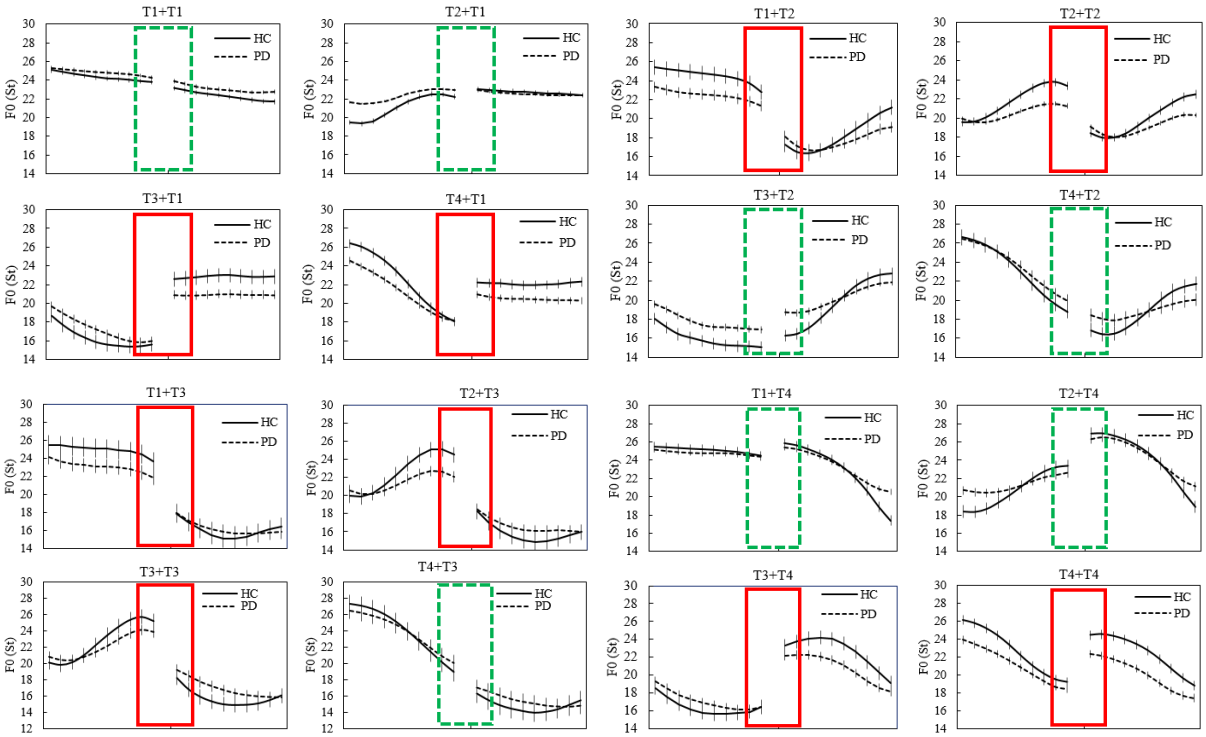


Figure 2: *F0 contours of disyllabic words of all 16 tonal combinations.*

References

[1] Gu, W., Fan, P., Liu, W. 2018. Acoustic analysis of Mandarin speech in Parkinson's disease with the effects of levodopa. In: Q. Fang *et al.* (Eds.), *Studies on Speech Production* (Revised selected papers from ISSP 2017; Lecture Notes in Artificial Intelligence, vol. 10733, Springer), 211–224.

[2] Fan, P., Gu, W., Liu, W. 2018. Acoustic analysis of Mandarin speech in patients with Parkinson's disease. *Chinese Journal of Phonetics* 9, 19–25. (In Chinese)

[3] Fan, P., Gu, W., Liu, W. 2022. Acoustic analysis of parkinsonian speech assessed as 'normal' in the UPDRS. *Journal of Audiology and Speech Pathology* 30(3), 262–265. (In Chinese)

[4] Mekyska, J., Galaz, Z., Kiska, T., *et al.* 2018. Quantitative analysis of relationship between hypokinetic dysarthria and the freezing of gait in Parkinson's disease. *Cognitive Computation* 10, 1006-1018.

[5] Chiu, Y.-F., Forrest, K. 2017. The interaction of lexical characteristics and speech production in Parkinson's disease. *Journal of Speech, Language, and Hearing Research* 60(1), 13–23.

A new approach to the learning of tonal categories in tone and non-tone languages

Aoju Chen
Institute for Language Sciences, Utrecht University
aoju.chen@uu.nl

In the autosegmental-metrical (AM) theory, a mainstream framework for intonational phonology, the prosodic system of a language encompasses a hierarchical structure, composed of both a finite number of pitch patterns residing in words which are grouped into prosodic phrases, and form-meaning mappings between changes in pitch patterns and changes in meaning [1, 2]. Recent research on prosodic development in children acquiring a non-tone language has adopted the AM view on prosody and examined development trajectories of pitch accents, phrasing and intonational form-meaning mappings, focusing on what children can do at which age. Similarly, research on children acquiring a tone languages is mostly concerned with production of lexical tones and limited form-meaning mappings at different ages, albeit making no reference to the AM framework [3-5]. It has been found that children become increasingly attuned to native pitch patterns and less sensitive to non-native pitch patterns between 6 and 9 months and they have developed the inventory of lexical tones in a tone language and the inventory of pitch accents in a non-tone language at about 12 months. However, exactly HOW children acquire pitch accents in non-tone languages and lexical tones in tone languages) remains to be investigated. In this position paper, I propose a new approach to address this question.

In this approach, adopting the AM framework, I assume that lexical tones have the same phonological status as pitch accents, i.e. being the discrete building blocks of the intonational pattern of an utterance, and refer to both as tonal categories. Furthermore, I take the view that although tone and non-tone languages differ in the density of tonal distribution (i.e. tone residing in every word vs. in some words in an utterance) and tonal function, across languages tonal categories can be distinguished in three dimensions, i.e. direction of pitch change (e.g. fall vs. rise), pitch height (e.g., high level vs. mid-level), and alignment of the highest or lowest pitch (e.g., early fall vs late fall) [6, 7]. But languages can differ in the weighting between dimensions [7, 8]. For example, the most important dimension in tonal perception is direction of pitch change in Mandarin but pitch height in English. Essential to acquiring native tonal categories is thus to find out which dimension(s) of pitch variation is/are relevant to formation of native tonal categories.

Hence, I propose to answer the HOW question by studying (1) innate attunement to pitch height and direction of pitch change; (2) the role of prenatal exposure in formation of tonal categories; (3) the role of distributional learning as a mechanism for the learning of tonal categories in the last trimester of gestation (26-39 weeks) and in the first months after birth; and (4) the role of visual cues, in particular, head and neck movements, in the learning of tonal categories at 4 to 12 months. In what follows, I briefly explain the rationale behind each research direction.

Innate attunement: The auditory system of human and other mammals is sensitive to variation in pitch, duration and intensity. For example, both humans and rats tend to group sequences of sounds in terms high-low in pitch and intensity and short-long in terms of duration, known as the Iambic-Trochaic law [9, 10]. The innate sensitivity to the prosodic parameters raises the exciting possibility that children may use it to uncover the relevance of pitch height and direction of pitch change for distinguishing tonal categories.

Prenatal exposure: Infants appear to possess some knowledge of the dimensions along which tonal categories differ in their native language already at 4-5 months, most probably the dimensions of pitch height and direction of pitch change [11-14]. Interestingly, infants' sensitivity to pitch height, duration and intensity appears to already be molded into preliminary language-specific preferences at birth, presumably through prenatal exposure to speech [15]. This finding suggests that the learning of tonal categories may start in the third trimester of gestation when the fetus can hear and process low-frequency sounds with almost intact prosody transmitted through maternal abdomen [16, 17].

Distributional learning: Infants aged 6-8 months can learn non-native phonemes through distributional learning, i.e. a type of statistical learning that involves tracking distributional properties in the input. They interpret an acoustic parameter with bimodal distribution (i.e. the most frequent sounds are from the two ends of an acoustic continuum) as an indicator for the relevance of this acoustic parameter in categorising sounds, not an acoustic parameter with unimodal distribution (i.e. the most frequent sounds are from the middle of the continuum), and learn to discriminate novel sounds along the parameter with bimodal distribution only [18]. Distribution learning heightens 2-3 months olds' perception of non-native phonemes [19] and appears to support infants' learning of non-native tonal categories at 11-12 months [20]. The questions arise as to whether and how early distributional learning is available as a mechanism for the learning of tonal categories.

Visual cues: Infants usually not only hear but also see other people talking when interacting with them. Speech is typically accompanied by gestures; co-speech gestures are enhanced in infant-directed speech [21]. Visual information contributes to the learning of sounds in infancy if it contains sufficient category-related information [22, 23]. Visual cues facilitate the perception of lexical tones in adults unfamiliar with the tones [24, 25]. Lexical tones differ perceivably in head and neck movements when produced in isolation [26, 27]. For example, in Mandarin, dropping of head can signal the low tone. Visual cues may thus potentially support infants in working out the relevant prosodic dimension in which tonal categories differ in the auditory modality.

To sum up, I have outlined a new approach to the study of acquisition of tonal categories, bridging the divide in past research on children's production and perception of lexical tones in tone languages and pitch accents in non-tone languages and tackling the question of how children come to produce and perceive tonal categories in their prosodic system. In the talk, I will also suggest ideas for how to test hypotheses arising from this approach.

Proceedings of the Second International Conference on Tone and Intonation, pages 119–120
November 18–20, 2023.

References

[1] Ladd, D. R. (1996). Intonational phonology. Cambridge, UK and New York, NY: Cambridge University Press.

[2] Pierrehumbert, J. B., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In R. R. Cohen, J. Morgen, & M. E. Pollack (Eds.), Intentions in Communication (pp. 271–311). MIT Press.

[3] Sato, Y., Sogabe, Y., & Mazuka, R. (2009). Development of hemispheric specialization for lexical pitch–accent in Japanese infants. Journal of Cognitive Neuroscience, 22 (11), 2503-2513.

[4] Fikkert, P., Liu, L, & Ota, M. (2021) The Acquisition of Word Prosody. In C. Gussenhhoven, and A. Chen (eds), The Oxford Handbook of Language Prosody (pp.541-552). Oxford: OUP.

[5] Chen, A., Esteve-Gibert, N., Prieto, P. & Redford, M. (2020). Development in phrase-level from infancy to late childhood. In C. Gussenhoven & A. Chen (eds) The Oxford Handbook of Language Prosody (pp. 553-562). Oxford: OUP.

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

[7] Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. Language and Speech, 21, 1–33.

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

[9] Hayes, B. (1995). Metrical stress theory: Principles and case studies. In Proceedings of the eleventh annual meeting of the Berkeley linguistics (pp. 429–446). Chicago, IL: University of Chicago Press.

[10] Nespor, M., Shukla, M., Vijver Van de, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO et OV languages. Lingue E Linguaggio, 2, 1–29.

[11] Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. Infancy, 10(3), 241–265.

[12] Singh, L., & Fu, C. S. L. (2016). A New View of Language Development: The Acquisition of Lexical Tone. Child Development, 87(3), 834–854.

[13] Tsao et al. 2004, Tsao, F.-M. M., Liu, H.-M. M., Kuhl, P. K., Feng-Ming, T., & Huei-Mei, L. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. Child Development, 75(4), 1067–1084.

[14] 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.

[15] Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. Brain and Language, 162, 46–59.

[16] DeCasper, A. J., Lecanuet, J. P., Busnel, M. C., Granier-Deferre, C., & Maugeais, C. (1994). Fetal reactions to recurrent maternal speech. Infant Behavior and Development, 17(2):159-164

[17] Lecanuet, J. P., Gautheron, B., Locatelli, A., Schaal, B., Jacquet, A. Y., & Busnel, M. C. (1998). What sounds reach fetuses: Biological and nonbiological modeling of the transmission of pure tones. Developmental Psychobiology, 33, 203–219.

[18] Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. Developmental Science, 11, 122–134.

[19] Wanrooij, K., Boersma, P., & Van Zuijen, T. L. (2014). Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. PLoS ONE, 9(10).

[20] Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. Cognition, 133(2), 385–94.

[21] Smith, N. A., & Strader, H. L. (2014). Infant-directed visual prosody: Mothers' head movements and speech acoustics. Interaction Studies, 15, 38–54.

[22] Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. Cognition, 108(3), 850–5.

[23] Ter Schure, S., Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: Effects on infants' looking patterns and discrimination. Frontiers in Psychology, 7.

[24] Burnham, D., Reynolds, J., Vatikiotis-Bateson, E., Yehia, H., Ciocca, V ., & Haszard Morris, R. (2006). The perception and production of phones and tones: The role of rigid and non-rigid face and head motion. ISSP 2006, 7th International Seminar on Speech Production, (1954), 185–192.

[25] Hannah, B., Wang, Y., Jongman, A., Sereno, J.A., Cao, J., & Nie, Y. (2017). Cross-Modal Association between Auditory and Visuospatial Information in Mandarin Tone Perception in Noise by Native and Non-native Perceivers. Frontiers in Psychology, 8:2051.

[26] Attina, V., Gibert, G., Vatikiotis-Bateson, E., and Burnham, D. (2010). "Production of mandarin lexical tones: auditory and visual components," in Proceedings of the AVSP-2010, Hakone

[27] Chen, T. H., and Massaro, D. W. (2008). Seeing pitch: visual information for lexical tones of Mandarin-Chinese. J. Acoust. Soc. Am. 123, 2356–2366.

Author Index