The impact of prosodic boundary and information structure on tonal coarticulation in spontaneous Cantonese

Gao, X.; Voeten, C.; Liberman, M.

DOI
10.21437/SpeechProsody.2024-80

Publication date
2024

Document Version
Final published version

Published in
Speech Prosody 2024

License
Other

Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

The impact of prosodic boundary and information structure on tonal coarticulation in spontaneous Cantonese

Xin Gao[1], *Cesko Voeten*[123], *Mark Liberman*[1]

[1]University of Pennsylvania, [2]University of Amsterdam, [3]Fryske Akademy

kauhsin@sas.upenn.edu, cvoeten@sas.upenn.edu, myl@cis.upenn.edu

Abstract

Pre-low Raising (PLR) is a tone coarticulation phenomenon, in which the pitch of a high-tone syllable is higher before a low tone than before a high tone. Although PLR is observed in various languages in read speech [1, 2], its occurrence in spontaneous speech remains understudied. The current study examines the presence of PLR in spontaneous Hong Kong Cantonese using the CantoMap Map-Task dataset [3] and explores the influences of prosodic boundaries and information structure (the givenness of a word in the discourse) using generalized additive mixed models (GAMMs).

The findings confirm the presence of PLR in spontaneous speech. Moreover, the study demonstrates that PLR is stronger in the absence of a prosodic boundary, indicating the substantial influence of prosodic boundaries on tonal coarticulation. Furthermore, the study examines how the f0 of the following syllable impacts the f0 realization of the preceding syllable, revealing a correlation between the phonetic context and the realization of tonal coarticulation.

Index Terms: tonal coarticulation, Pre-low Raising, prosodic boundary, givenness

1. Introduction

Following the definition of coarticulation as a contextual variation influenced by the phonetic contexts [4, 5], we consider tonal coarticulation as the variation in the realization of fundamental frequency (f0) of a specific tone, dictated by the effects of its surrounding context. In the current study, we investigate how different contextual elements affect the realization of f0 for a specific tone. The analysis investigates the phonological and phonetic characteristics of the following syllable and considers the impact of prosodic and information structure components within the context.

Research suggests that the strength of prosodic boundaries significantly influences tonal coarticulation. For instance, research has shown that in Tianjin Mandarin, the f0 realization of neutral tone is affected by the tone of the subsequent syllable; however, this effect of tonal coarticulation is dependent on the strength of the syntactic-prosodic boundary between the neutral tone and the following tone – a Subject-Predicate boundary can block the tonal coarticulation effect from the subsequent tone [6]. Similarly, in Standard Mandarin, the strength of the syntactic boundary impacts the anticipatory effect on tonal articulation in sequences with a neutral tone, as indicated by [7].

Moreover, givenness is related to prosodic reduction, as discussed in [8]. Specifically, the repetition of a word or word sequence, or the prior activation of a motor plan for a particular word, is a significant factor contributing to prosodic reduction [9]. There are significant differences in the phonetic realization of a word when it is mentioned for the second time, compared to its first mention, as evidenced in studies on Indian English and Korean [10]. Additionally, verbal repetition has been found to cause a phonetic reduction in word onsets as well as duration [11]. Furthermore, a production experiment demonstrated a phonetic reduction in the f0 peak of given constituents [12]. Given these findings, it is reasonable to infer that givenness also affects tonal coarticulation.

2. “Pre-low Raising” in Cantonese

“Pre-low Raising” (PLR) describes a phenomenon where the f0 realization of a high-tone syllable is raised when preceding a low tone, as opposed to when it precedes a high tone [1]. This phenomenon is cross-linguistically observed in various tonal languages, including Cantonese [1, 13, 14], Ibibio [15], Mandarin [16], Thai [17, 14], Triqui [18], Vietnamese [19], as well as in pitch-accent languages like Japanese [20].

Research also indicates that other factors modulate PLR. For example, speech rate and vowel quality influence the degree of tonal coarticulation in PLR with a sentence-reading experiment [14]. While PLR is well-documented across languages in read speech, its presence in spontaneous speech is less understood. Therefore, our study examines PLR in conversational Hong Kong Cantonese, focusing on the effects of prosodic factors and information structure, specifically boundaries and givenness.

Cantonese is a tonal language in which each syllable is associated with a lexical tone. In Hong Kong Cantonese, there are six unchecked tones, as illustrated in Figure 1. To narrow down the scope of our research and focus more on the exploration of factors influencing tonal coarticulation, in the current study, we examine the f0 realization of Tone 1 (T1) in contexts where it precedes either another T1 or a Tone 4 (T4). Namely, it involves comparing the F0 in the first syllable of the “T1-T1” and “T1-T4”.

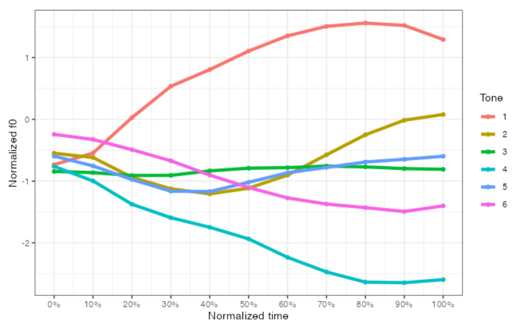

Figure 1: *Time-normalized f0 contours of the six unchecked tones in Hong Kong Cantonese from this study’s dataset [3].*

Because T1 is the highest tone while T4 is the lowest tone in

10.21437/SpeechProsody.2024-80

Hong Kong Cantonese,[1] if PLR phenomenon exists, we would expect to observe that the f0 realization of the first T1 in "T1-T1" is lower than that in "T1-T4".

Since the f0 realization of a syllable is influenced by the tone category of the following syllable, it seems reasonable to suggest that this variation stems from the abstract phonological influence of the context. This implies an abstract processing of dissimilatory tonal coarticulation from the speaker's perspective to enhance contrast. However, [13] considered a perceptual account for PLR as unlikely due to its occurrence in the high-register of Cantonese, where ambiguity is minimal. Instead, they suggested a physiological rationale: PLR facilitates achieving a high maximum f0 change velocity for the second syllable's f0 target. A lower f0 target in the second syllable requires the first syllable to rise higher. In this case, we should expect PLR's dissimilatory effect to stem from the phonetic context of the subsequent syllable.

To assess the sources of variations in PLR, we examine the effects of both phonological contexts and phonetic contexts on tonal articulation in this study.

3. Methods

3.1. Dataset

In this study, we used CantoMap, a corpus containing approximately 13 hours of conversational data in Hong Kong Cantonese, collected from 40 speakers. This data was gathered through a map task specifically designed to prompt natural conversation. The corpus is manually transcribed and includes segmented word boundaries [3].

The map task includes a small number of nonce proper names. Considering their limited quantity, we chose to exclude them from our analysis. A total of 2,458 "T1-T1" and 898 "T1-T4" sequences were analyzed in the current study.

3.2. Data analysis

An in-house Cantonese Forced Aligner was used to synchronize the audio recordings with their transcripts, achieving alignments at both the phone and syllable levels [21]. Due to technical limitations, sentence-final particles and code-switching words were omitted from our dataset.

After obtaining the forced-aligned results, f0 measurements from the "T1-T1" and "T1-T4" sequences were taken at every 10% interval of sonorant segments. For each speaker, we normalized their f0 by converting the raw f0 values into semitones. This was achieved by dividing the raw f0 values by the median f0 of the respective speaker.

In addition to f0 measurements, our data analysis involved extracting two additional types of information from the corpus. This includes word boundary information, particularly identifying if a word boundary exists between disyllabic sequences, and givenness information, determining whether a sequence is mentioned for the first time in the current conversation. Both these factors served as predictors in our analysis.

Statistical modeling of the data was done using two generalized additive mixed models ('GAMM'; [22]): one for the phonological context, and one for the phonetic context. Both models had the normalized f0 as the dependent variable.

The phonological model included fixed effects for tonal category (T1 through T6, deviation-coded), prosodic boundary (present vs absent, deviation-coded), and givenness (first vs non-first mention, deviation-coded). Smooth terms, defined as thin-plate regression splines, were added for 'time'; since there are eleven time samples per trial, these were allowed maximally eleven basis functions. We additionally included smooth terms for 'speech rate' and 'location'; these were again defined as thin-plate regression splines and were afforded 30 basis functions (which we verified to be sufficient by means of `mgcv`'s `gam.check` function) and are not of interest, but serve to control for these potential nuisance factors. In addition, we included random effects for speakers by tonal category, random smooths for speakers along time by tonal category (with the customary first-derivative penalty), random smooths for speakers by speech rate, and random smooths for speakers by location (both with first-derivative penalties).

To represent the phonetic context, we calculated the averaged normalized f0 within the initial 30% of the second syllable in disyllabic sequences. This approach is based on the hypothesis that the early part of the second syllable, rather than the entire syllable, has a more direct impact on the f0 realization of the preceding syllable.

The GAMM for the phonetic model included fixed effects for prosodic boundary and givenness as above. It similarly shared the general and random smooth terms for speech rate and location. The key term of interest, however, was different: instead of fitting a 1-D smooth over time by the six tonal categories, the phonetic model fitted a 2-D tensor product of 'f0 at 30%' and 'time'. This fits a two-dimensional surface (a 'bendy sheet', according to [22]), which bends nonlinearly over (coarticulatory) f0 values at 30% and time simultaneously. The two margins of the tensor product had 30 and 11 basis functions, respectively; furthermore, the tensor product was constructed by each of the four combinations of prosodic boundary and givenness. The random-effects part of the model comprised random effects for speakers by these same four levels of boundary and givenness. We also included random smooths for speakers along the f0 at 30% by these boundary-and-givenness combinations, and random smooths for speakers along time by boundary–givenness, with the appropriate penalties.[2] As before, we additionally included random smooths for speakers along speech rate and along location.

Based on exploratory fits, the two models were fit to scaled-t errors and included an autoregressive process of order 1 with $\rho = .3$. Significance of differences between conditions was assessed by means of Bayesian credible intervals, using the approach by [22]. Differences are considered significant wherever the credible interval excludes zero.

4. Results

The first question we address is the presence of PLR in conversational Cantonese. Figure 2 displays the f0 contours for "T1-T1" and "T1-T4" sequences, where the dashed line differentiates the first and second syllables. A noticeably higher f0 realization is observed for the first syllable in the "T1-T4" sequence compared to the "T1-T1", providing evidence of PLR in spontaneous conversational speech.

[1] Tone 6 (T6) is also a low-register tone in Hong Kong Cantonese; however, "T1-T6" sequences were not included in our analysis due to the prevalent use of the phrase *gan1 zyu6* (meaning "and then") in the corpus, which is predominantly found in utterance-initial positions.

[2] Due to technical limitations, it was necessary to split out the tensor product into these two separate 1-D splines; in current `mgcv` versions, fitting random tensor products is only possible in very specific circumstances, which did not apply to our model.

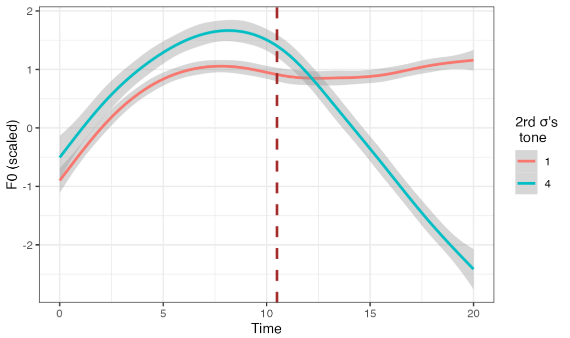

Figure 2: *F0 contours of "T1-T1" and "T1-T4" sequences.*

4.1. The phonological contexts

First, we explore whether and how the phonological context, namely the tone category of the second syllable, impacts the realization of PLR. Figures 3–5 present the relevant smooth curves fit by the GAMM, as well as the difference between them. Figure 3 illustrates the f0 difference in the first syllable between "T1-T1" and "T1-T4." Negative values on the y-axis signify a lower f0 in the first syllable of "T1-T1" compared to "T1-T4," effectively illustrating the PLR pattern.

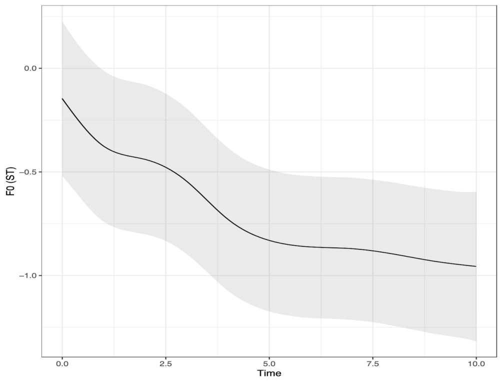

Figure 3: *F0 difference "T1-T1" minus "T1-T4" from GAMM: Negative values show lower f0 in "T1-T1".*

Figure 4 shows that prosodic boundary presence significantly affects PLR realization between "T1-T1" and "T1-T4." It displays the f0 differences in the first syllable of "T1-T1" and "T1-T4" sequences, with the x-axis representing the normalized time of the first syllable. Notably, the absence of an inter-syllable boundary amplifies the f0 difference in the first syllable between "T1-T1" and "T1-T4" sequences in the second half of the vowel. This suggests that phrasal boundary presence has a significant effect on the PLR realization in the phonological context.

Moreover, we examined whether givenness, as an information structure element, plays a role in PLR in spontaneous speech. Figure 5 suggests that later mentions resulting in larger f0 differences between T1-T1 and T1-T4 than first mentions, particularly towards the end of the first syllable of the sequence. credible interval for the difference includes zero throughout. This finding suggests that in our dataset, the givenness of information alone could not account for the variation in the phonetic realization of PLR in spontaneous and conversational speech.

In a midway recap, the GAMM for the phonological context

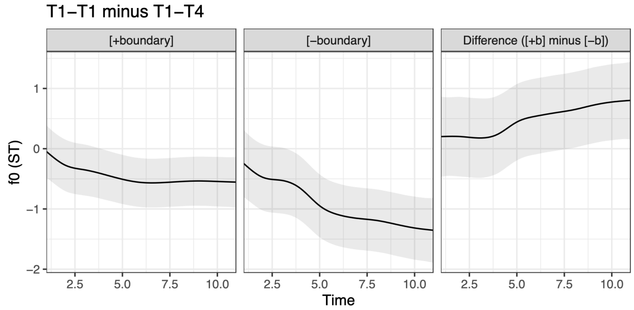

Figure 4: *Boundary effect on f0 difference "T1-T1" minus "T1-T4" sequences.*

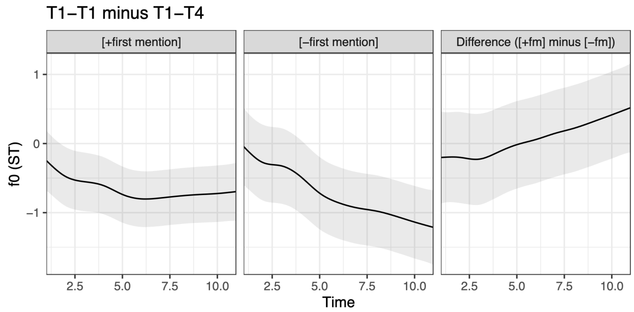

Figure 5: *Givenness effect on f0 difference "T1-T1" minus "T1-T4" sequences.*

reveals a higher f0 realization in the first syllable of T1-T4 sequences compared to T1-T1 in conversational Hong Kong Cantonese, confirming the presence of the PLR phenomenon. This effect is stronger when there is no prosodic boundary within the sequence.

4.2. The phonetic contexts

Next, we investigate whether the phonetic context, like the phonological context, can produce a similar dissimilatory effect on the f0 realization of the first syllable in disyllabic sequences.

Figures 6 and 7 visualize the tensor product smooth (i.e., the 2-dimensional surface of f0-at-30%-by-time values), as well as the difference between these two tensors, respectively as functions of prosodic boundary and of givenness. All instances analyzed in 4.1 are involved here, namely from both "T1-T1" and "T1-T4" sequences. The y-axis in the plots denotes the average f0 of the initial 30% of the second syllable, while the x-axis denotes the time point. Color reflects the effects on the first syllable's f0 realization; where the color is fully opaque (i.e., *not* translucent), the 95% credible interval excluded zero (i.e., the effect is significant).

Should a dissimilatory effect from the phonetic context influence the first syllable's f0, we would expect more f0 increase (shown by color) with lower f0 in the second syllable (y-axis values). However, figures 6 and 7 reveal the contrary: the first syllable's f0 increases more with higher f0 in the second syllable. This suggests that the phonetic context leads to assimilation, not dissimilation, in the first syllable's f0. Thus, the phonetic context's influence in our data is opposite to the phonological context's; while the latter shows PLR's dissimilatory effect, the former demonstrates assimilation.

The presence of phrasal boundaries and the givenness of in-

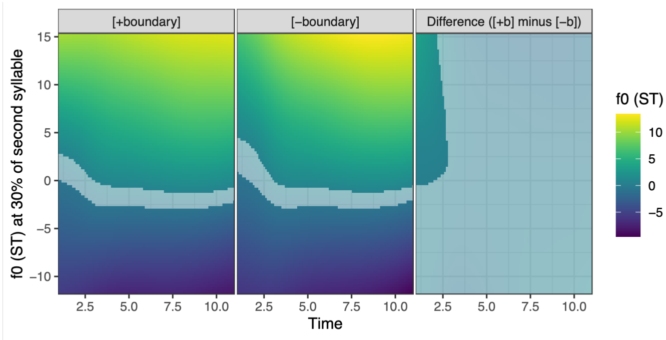

Figure 6: *Phonetic context impact on f0 by prosodic boundary.*

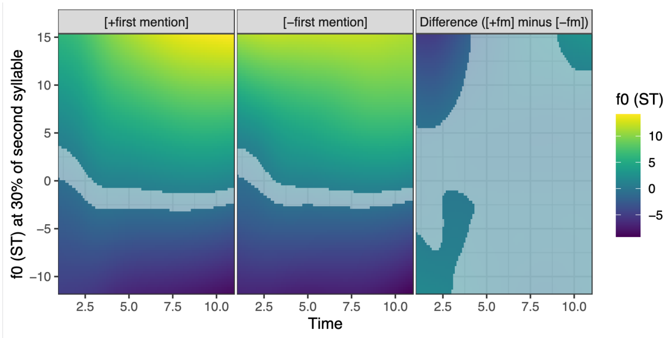

Figure 7: *Phonetic context impact on f0 by givenness.*

formation generally do not significantly affect how the phonetic context influences the realization of tonal articulation. This is evident in the "difference" panels (the right-most panels) in figures 6 and 7.

5. Discussion

5.1. PLR in spontaneous speech

This study explored the realization of tonal coarticulation in spontaneous speech using a map task corpus of Hong Kong Cantonese with conversational and spontaneous speech. The analysis involved comparing the f0 realization of the T1 syllable when placed before T1 or T4 syllables. First, we fit a GAMM to assess the effect of phonological tonal categories from the surrounding context. The results indicate significant f0 differences in the first syllable between "T1-T4" and "T1-T1" sequences, providing evidence for the presence of PLR, a phonological tonal coarticulatory phenomenon, in spontaneous speech. Furthermore, we investigated the effect of the phonetic context of surrounding syllables on the f0 realization. The results suggest that the phonetic context of surrounding syllables can influence f0 realizations.

5.2. Influence of prosodic and information structure on tonal coarticulation

Another goal of this research was to explore how tonal coarticulation is influenced by prosodic and information structure elements. Existing research suggests both prosodic boundaries and givenness influence phonetic realization [6, 10]. To address this research question, we first looked into the impact of a prosodic boundary on PLR, in the phonological context. This was investigated by comparing the PLR in disyllabic sequences in scenarios with and without a prosodic boundary between the syllables. The findings indicated that PLR is stronger in the absence of a prosodic boundary, indicating the influence of prosodic boundaries on tonal coarticulation. On the other hand, the givenness of a sequence does not significantly affect PLR realization. Additionally, the study examined how the phrasal boundary presence and givenness of a sequence affect the phonetic context effect on f0 realization. No significant effect of boundary or givenness has been observed in the data.

The absence of a notable effect of givenness on the phonological context (i.e., on PLR realization), and the lack of effects from boundary and givenness on the phonetic context could be attributed to the nature of a spontaneous and conversational speech corpus used in this study. Speech in such a corpus is subject to a multitude of influencing factors, making it difficult to control variables as precisely as in laboratory-based sentence reading experiments. Despite extensive efforts to model various influencing factors, the control remains incomplete. In this context, even if systematic effects of boundaries and givenness exist, they might not be detectable with the current modeling approaches. Further exploration of the interaction between prosody and information structure on phonetic feature realizations in spontaneous speech is necessary and will provide additional insights into this question.

5.3. Sources of variations: phonological vs. phonetic contexts

The current research also found that the phonological and phonetic contexts have opposite impacts on the f0 realization of the first syllable in disyllabic sequences in the current research. The influence of the phonological context aligns with the dissimilatory pattern of PLR: a higher tone in the subsequent syllable results in a lower f0 in the preceding syllable, and vice versa. However, the phonetic context showed an assimilatory effect, where a higher f0 in the subsequent syllable led to a higher f0 in the preceding syllable.

[13] proposed a physiological explanation for PLR, suggesting that raising the pitch in the first syllable enables a higher f0 velocity to achieve a lower f0 target in the subsequent syllable. Under the assumption that PLR is purely physiologically driven, one would expect that the phonetic context would induce a dissimilatory effect on the f0, mirroring the phonological context. However, the current study contradicts this expectation.

Therefore, our findings suggest that PLR is not purely a result of physiological factors. Because the processing of the second syllable's tone involves the handling of abstract categories, we propose that PLR likely involves a level of processing by the speaker that is more abstract in nature. The mechanisms of this processing require further exploration in future studies.

6. Conclusion

This study investigated tonal coarticulation in conversational speech, particularly focusing on PLR. By examining the f0 patterns of a high tone (T1) preceding either another high tone or a low tone (T4), we confirmed the existence of PLR in spontaneous speech. Our findings also revealed that prosodic structure significantly influence tonal coarticulation. Additionally, the study found opposite effects of phonological (dissimilatory) and phonetic (assimilatory) contexts on f0 variations.

7. References

[1] W. Gu and T. Lee, "Effects of tone and emphatic focus on F0 contours of Cantonese speech: A comparison with Standard Chinese," *Chinese Journal of Phonetics*, vol. 2, pp. 133–147, 2009.

[2] C. DiCanio, "Triqui tonal coarticulation and contrast preservation in tonal phonology," 2014.

[3] G. Winterstein, C. Tang, and R. Lai, "Cantomap: a Hong Kong Cantonese Maptask Corpus," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 2906–2913.

[4] R. G. Daniloff and R. E. Hammarberg, "On defining coarticulation," *Journal of Phonetics*, vol. 1, no. 3, pp. 239–248, 1973.

[5] J. J. Ohala, "Coarticulation and phonology," *Language and speech*, vol. 36, no. 2-3, pp. 155–170, 1993.

[6] Q. Li and Y. Chen, "Prosodically conditioned neutral-tone realization in Tianjin Mandarin," *Journal of East Asian Linguistics*, vol. 28, pp. 211–242, 2019.

[7] M. Shi, Y. Chen, and M. Mous, "Tonal split and laryngeal contrast of onset consonant in Lili Wu Chinese," *The Journal of the Acoustical Society of America*, vol. 147, no. 4, pp. 2901–2916, 2020.

[8] J. M. Kahn and J. E. Arnold, "A processing-centered look at the contribution of givenness to durational reduction," *Journal of Memory and Language*, vol. 67, no. 3, pp. 311–325, 2012.

[9] M. Wagner, "Information structure and production planning," 2014.

[10] R. E. Baker and A. R. Bradlow, "Second mention reduction in Indian, English, and Korean," *Journal of the Acoustical Society of America*, vol. 122, no. 5, p. 2993, 2007.

[11] C. L. Jacobs, L. K. Yiu, D. G. Watson, and G. S. Dell, "Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production," *Journal of Memory and Language*, vol. 84, pp. 37–48, 2015.

[12] S. Genzel, S. Ishihara, and B. Surányi, "The prosodic expression of focus, contrast and givenness: A production study of Hungarian," *Lingua*, vol. 165, pp. 183–204, 2015.

[13] A. Lee and Y. Xu, "Effect of speech rate on pre-low raising in Cantonese." Tonal Aspects of Languages, 2016.

[14] A. Lee, S. Prom-On, and Y. Xu, "Pre-low raising in Cantonese and Thai: Effects of speech rate and vowel quantity," *The Journal of the Acoustical Society of America*, vol. 149, no. 1, pp. 179–190, 2021.

[15] A. L. Coombs, "High tone processes in Ibibio," in *Proceedings of Meetings on Acoustics ICA2013*, vol. 19, no. 1. Acoustical Society of America, 2013, p. 060232.

[16] Y. Xu, "Contextual tonal variations in Mandarin," *Journal of phonetics*, vol. 25, no. 1, pp. 61–83, 1997.

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

[18] C. DiCanio, "The evolution of tonally conditioned allomorphy in Triqui: evidence from spontaneous speech corpora," *Linguistics Vanguard*, vol. 8, no. s5, pp. 545–556, 2022.

[19] M. Brunelle, "Northern and Southern Vietnamese tone coarticulation: A comparative case study," *Journal of Southeast Asian Linguistics*, vol. 1, no. 1, pp. 49–62, 2009.

[20] A. Lee, S. Prom-On, and Y. Xu, "Pre-low raising in Japanese pitch accent," *Phonetica*, vol. 74, no. 4, pp. 231–246, 2017.

[21] T. Lee and D. Tao, "Cantonese forced alignment," 2021.

[22] S. N. Wood, *Generalized additive models: an introduction with R*, 2nd ed. Boca Raton: Chapman & Hall/CRC, 2017.