Tonal effects on vowel duration in Bangkok Thai

Francesco Burroni[a*], *Teerawee Sukanchanon*[b*]

[a]Department of Linguistics and Cognitive Science Program, Cornell University, USA
[b]Southeast Asian Linguistics Research Unit, Faculty of Arts, Chulalongkorn University, Thailand
fb279@cornell.edu, t.sukanchanon@gmail.com

Abstract

We investigated tonal effects on vowel duration in two experiments with 37 speakers of Bangkok Thai. For long vowels in open syllables, the pattern of tonal effects on vowel duration is {Mid,Low} < {Falling,High,Rising}; for closed syllables with short vowels and sonorant codas, the pattern is {Rising} < {Mid,Low} < {High} < {Falling}. Our results do not align with hypothesized universal patterns or with previous reports on Bangkok Thai. Our findings are better understood by referring to f0 control mechanisms. Finally, we found that the tonal effects are mediated by syllable structure in line with diachronic changes in vowel length.

Index Terms: tone, vowel duration, vowel length, word duration, fundamental frequency, Thai

1. Introduction

A claim often repeated in the literature,e.g., [1], and going back to [2], is that the effects of tone on vowel duration are hypothesized to be universally inversely proportional to average fundamental frequency (f0) values. Vowels in syllables with higher f0 have shorter vowel duration, while those in syllables with lower f0 have longer duration. Gandour [2] based his account on diachronic data from Thai varieties, where long vowels emerged in association with rising and non-high level tones, and short vowels emerged in association with falling and high level tones. Gandour's claim was also in line with Abramson's findings on Bangkok Thai (BKKT) [3], where the Mid (M) and Low (L) tones, tones with relatively low average f0, are longer than the Falling (F) and High (H) tones, tones with relatively high average f0. Note, however, that none of the BKKT tone is level, hence, the relationship between their names and f0 contour should be taken with a grain of salt, Figure 1.

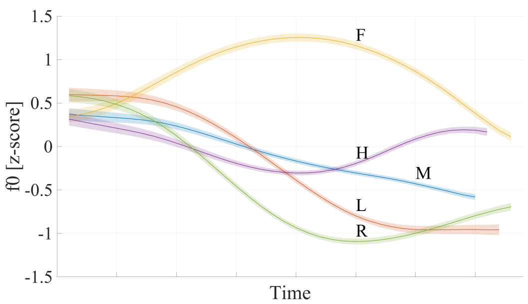

Figure 1: *Average BKKT tonal contours from exp. 1*

Subsequent work showed that Gandour's hypothesis may not hold for all languages. One study has shown that among the three level tones in Taiwanese, the durational effect hierarchy is L < M < H [4]. One study of Cantonese [5] has shown that the duration effect hierarchy is {L,H} < ML < M (braces are used to indicate sets of tones that do not differ). There are reasons to question whether Gandour's hypothesis and the order of durational effects reported in [3] are correct, even for BKKT. The first issue is based on physiological considerations. The durational effects of tone reported by [3], {F,H} < {M,L}, are the opposite of what f0 control mechanisms would predict; as it has been reported that f0 falls are faster/shorter than f0 rises [6], [7]. If the effects of tone on duration are related to f0 control, the M and L tones, which only have a falling component, should be associated with shorter vowel durations. A second issue comes from diachronic data. In BKKT, the words that underwent vowel lengthening are almost only words with tones that are relatively high in the f0 range, the F and H tones [8]. What emerges from diachronic data is also that all lengthened words have /a/ as nucleus and sonorants as codas, for the most part, /-j/ and /-w/. BKKT vowels also underwent shortening. Syllables ending with nasals /-m, -n, -ŋ/ were typically shortened when occurring with the L and F tones [8]. Tone-coda interaction effects suggest that the durational effects of tone may be conditioned by phonological factors, like syllable structure and coda types. A third issue that has not been explored is whether durational effects of tone may similarly affect the short and long vowels of BKKT, given that the language has a phonological contrast for vowel length.

Against this background, we formulate the following research questions: (I) Is the order reported by [3] for tonal effects on vowel duration still observed in contemporary BKKT? (II) What is the relationship, if any, between the durational effects of tone, f0 control mechanisms, and diachronic changes in vowel length? (III) Are the effects mediated by other phonological factors, e.g., vowel length and syllable structure? We make the following hypotheses and associated predictions. For (I), if the hypothesis of [2], an inverse relationship between tonal durational effects and average f0, is correct; then, we expect to observe the durational effect hierarchy reported in [3]. The M, L, and R tones, which have relatively low average f0, should be associated with longer vowels. However, the diachronic changes in vowel length and known differences between the speed of f0 rises and falls may suggest otherwise. For (II), if the order of tonal duration effects reported in [3] is correct, we do not have any clear relationship between f0 control mechanisms and vowel duration. We also do not have any clear association with diachronic changes. If, however, f0 control mechanisms are one of the factors determining vowel duration and, moreover, diachronic changes reflect synchronic phonetic variation, we may expect an order of tonal durational effects like {M,L,F} < {H,R} (f0 falls are shorter and f0 rises are longer [7]) or {R} < {M,L} < {F,H} (R induces shortening, while F and H induce lengthening, as diachronic data show). For (III), the diachronic data suggest that

* Equal contribution.

certain combinations of tones and codas result in phonological lengthening or shortening diachronically. We may, thus, expect to observe traces of these interactions in synchronic variation in vowel duration. We have investigated these questions by studying the vowel /a/, the only vowel in BKKT that has undergone phonological lengthening diachronically. We have studied two environments: /Ca:/ (long /a/ in open syllables) and /Ca(G|N)/ (short /a/ in closed syllables with glide or nasal codas); as these are among the only syllabic configurations where all 5 BKKT tones (M, L, F, H, R) can be licensed.

2. Methods

For experiment 1 (exp. 1), data were collected from 20 speakers (range=19-52, mean=28.4, std=12.8), and, for experiment 2 (exp. 2), from 17 speakers (range=20-23, mean=21.2, std=0.8). Participants in exp. 1 were students or staff at a North American university; participants in exp. 2 were students at a university in Bangkok. All speakers were screened for nativeness in BKKT by a native speaker trained in phonetics. They reported no speech or hearing impairments. In exp. 1, participants produced [ma:] with all five tones of BKKT (M, L, F, H, R). The target was embedded in a carrier sentence with a fixed number of words and syllables, Table 1.

Table 1. *Stimuli in exp. 1*

w1	w2	w3	w4	w5
dū:	mî:	mā:	bōn	bɔ̄n
	mǐ:	mà:		dā:w
		mâ:		lāŋ
		má:		
		mǎ:		

Disyllabic tonal combination of w2 and w3 are intended to represent nonce words. F/R in w2 were chosen because the dataset was originally designed to study tonal coarticulation in separate work by one of the authors. Participants were instructed so that the F/R initial word of the combination represents an imaginary animal while the second one a fur pattern. Thus, the disyllabic tonal combinations represent noun-noun compounds, meaning "(I) look at a mî:/mǐ: (with fur pattern) mā:/mà:/mâ:/má:/mǎ:". Participants completed 10 blocks in which they produced random combinations of 2 (F/R) × 5 (M/L/F/H/R) × 3 (distractors) = 30 unique stimuli. w5 of the carrier sentence was varied at every trial to function as a distractor. The number of tokens collected in exp. 1 was 20 (participants) × 30 (unique stimuli combinations) × 10 (blocks) = 6000. In exp. 2, the targets were 45 unique monosyllabic words with syllable structure /Caŋ/ and /CaG/ (where C = /p, t, k/ and G = /j, w/) combined with all 5 tones of BKKT. The word list was designed to maximize the number of nonce words. An example of all targets with /p/ onset is given in Table 2. Identical combinations were elicited with /k/ and /t/ onsets, yielding 3 (onsets) × 3 (rhymes) × 5 (tones) = 45 unique targets.

Table 2. *Example stimuli with p onset in exp. 2*

syllable structure		tones				
onset	rhyme	1	2	3	4	5
p	-aŋ	pāŋ	pàŋ	pâŋ	páŋ	pǎŋ
	-aj	pāj	pàj	pâj	páj	pǎj
	-aw	pāw	pàw	pâw	páw	pǎw

The targets were embedded in a carrier sentence [dū: khām wâ: X bōn Y] "Look at the word X on Y", in which X is the target and Y is one of eight disyllabic meaningful words acting as a distractor. The distractors are [pʰē:.dā:n] 'ceiling', [kām.phɛ̄:ŋ] 'wall', [thā:ŋ.dɤ̄:n] 'path', [kā:ŋ.kē:ŋ] 'trousers', [lām.pʰō:ŋ] 'speaker', [bān.dāj] 'stairs', [sāʔ.pʰā:n] 'bridge', and [kāʔ.lā:] 'coconut shell'. Exp. 2 was divided into 8 blocks. Each block contained all unique targets, pseudo-randomized, with a constraint that targets with the same tone will not appear consecutively. The distractors were evenly distributed across all tokens. The number of collected tokens in exp. 2 was 17 (participants) × 45 (unique stimuli) × 8 (blocks) = 6120.

For both experiments, participants sat in a sound-attenuated room in front of a computer monitor. A custom MATLAB GUI for each experiment was used to present stimuli in the form of pictures and phrases in Thai orthography. Simultaneously audio was collected, using a head-mounted microphone at a sampling rate of 44.1 kHz and 24 bits per sample. Speaker-specific monophone HMMs were trained in Kaldi [9] and used to perform forced alignment separately by speaker. The training data were hand-segmented in PRAAT [10] using the waveform, spectrogram, and changes in intensity/formant trajectories. Boundaries between vowels and glides were identified as the midpoint of the first formant transition. Boundaries between vowels and nasals were based on the appearance of antiformants and spectral/amplitude changes. Following forced alignment, a proportion of segmented data was randomly selected to be visually inspected and hand corrected. The manually checked alignments were used to retrain HMMs and realign all the trials. This process was repeated until the automatically segmented data were comparable to manual annotation. Trials that contained disfluencies were excluded. The word [kâw] in exp. 2 was excluded, as participants consistently produced it with a phonologically long vowel. In total, we analyzed 5977 tokens from exp. 1 and 5956 tokens from exp. 2. The main dependent variable we analyzed is vowel duration of the targets as defined by segmental boundaries. For exp. 1, given the identical segmental composition of all targets, we also analyzed word duration. For exp. 1, linear mixed effect regressions were conducted to assess the effect of tone, a categorical variable (with reference set as the M tone), on vowel duration. Random intercepts and slopes by speakers were also included. For exp. 2, we had two categorical variables, tone (reference M tone) and coda type (reference /-ŋ/), as well as their interaction. For random effects in exp. 2, we compared various structures and found speakers and onsets intercepts and slopes to be the maximal structure justified by a loglikelihood ratio test. The final statistical model for each dependent variable was selected in a pruning stepwise procedure, that is, we compared nested models and removed one fixed effect at a time and checked for significance *via* loglikelihood ratio tests.

3. Results

In exp. 1, we found that tone has a significant effect on long vowel duration in BKKT ($\chi^2_{(4)} = 22.7$, $p < .0001$). The M and L tones are systematically associated with shorter vowels than the F, H, and R tones. No significant differences were observed between M *vs.* L or among F *vs.* H *vs.* R. The differences between {M,L} and {F,H,R} range between –2 and –20 ms. Since the average vowel duration for the L and M tones is estimated at ~224 ms, the difference in duration associated with the presence of {F,H,R} tones represents an increase in duration in the range of 1-10% of the vowel. A full pairwise comparison with 95% confidence intervals (CI) for the differences in vowel duration among all tones is presented in Table 3. Each cell represents the difference between the tone in each row minus the tone in each column. Longer duration of vowels associated with {F,H,R} *vs.* {M,L} is evident by inspecting the distributions of all tokens in boxplots, Figure 2.

Table 3. *V: duration differences between each tone*

	M	L	F	H	R
M		*ns*	–6, –19	–4, –21	–2, –18
L	*ns*		–6, –18.5	–5.5, –20	–4, –17
F	6, 19	6, 18.5		*ns*	*ns*
H	4, 21	5.5, 20	*ns*		*ns*
R	2, 18	4, 17	*ns*	*ns*	

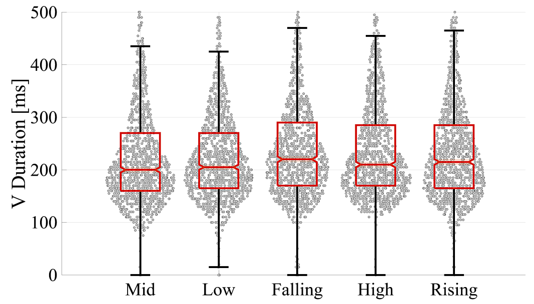

Figure 2: *V: duration in all tone conditions*

Word duration is also affected by tone category ($\chi^2_{(4)} = 23.6$, $p < .0001$) in the same way as vowel duration. M *vs.* L toned words are not different from each other. Nor are F *vs.* H *vs.* R toned words. However, {M,L} toned words are shorter than {F,H,R} toned words. 95% CI range between 5-25 ms; corresponding to an increase in duration for {F,H,R} toned words in the range of 2% to 8% of total word duration, estimated at ~314 ms.

In exp. 2, a significant effect of tone-coda interactions on vowel duration was observed ($\chi^2_{(10)} = 945.1$, $p < .0001$). Like in exp. 1, {M,L} are associated with shorter vowel duration than {F,H}. There are no differences between M *vs.* L or between F *vs.* H. However, we found that R toned vowels are significantly shorter than vowels associated with all other tones. Table 4 shows a full pairwise comparison with 95% CI for the differences in vowel duration among all tones (beige color indicates marginally significant differences).

Table 4. *V duration differences between each tone*

	M	L	F	H	R
M		*ns*	-2, -11	*0, -6*	1, 9
L	*ns*		-3, -13.5	-1, -8	0, 6
F	2, 11	3, 13.5		*ns*	6, 18
H	*0, 6*	1, 8	*ns*		4, 12
R	-1, -9	-0, -6	-6, -18	-4, -12	

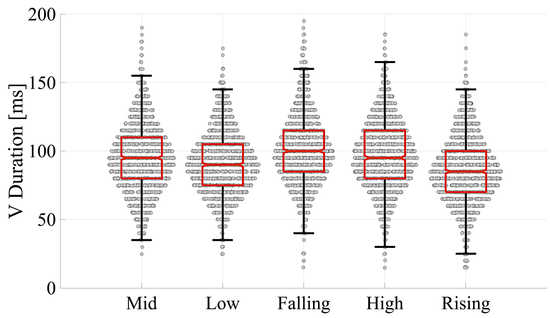

Figure 3: *V duration in all tone conditions*

For codas, vowel duration in syllables with bilabial glide coda (/-w/) is longer than in syllables with palatal glide (/-j/) and velar nasal (/-ŋ/) codas. Among the three codas, vowel duration in syllables with /-j/ is the shortest, Table 5 and Figure 4.

Table 5. *V duration differences between each coda*

	/-ŋ/	/-j/	/-w/
/-ŋ/		1, 6	-16, -21
/-j/	-1, -6		-20, -25
/-w/	16, 21	20, 25	

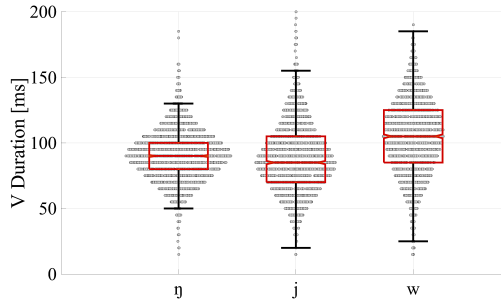

Figure 4: *V duration in all coda conditions*

We also observed coda and tone interaction effects on vowel duration. Significant differences in vowel duration due to tone-coda interactions along with 95% CI are presented in Table 6.

Table 6. *significant differences in vowel duration due to interactions (baseline is M tone with /-ŋ/ coda)*

	/-j/	/-w/
L	*ns*	-1, -8
F	7, 14	-3, -11
H	*ns*	*ns*
R	*ns*	-7, -14

The interactions can be interpreted as follows. /-j/ plus F is longer by [7, 14] ms than it is expected by adding the effects of /-j/ and F together. Notice how /-j/ coda plus F "stands out" more from the other /-j/ realizations even though the mean of all /-j/ realizations is lower than the other two coda conditions, orange arrows in Figure 5. A reverse effect is observed with /-w/, for which cooccurrence with the F, H, or R tones does not result in the degree of lengthening expected by adding the effects of coda and tone. In all these conditions the predicted vowel durations are shorter by [-8, -1] ms, [-11, -3] ms, and [-7, -14] ms respectively than predicted by a purely additive model. Note how /-w/ plus L/R are much further away from /-w/ plus M/H compared to their distances in the other two coda conditions, blue arrows in Figure 5. Finallly, note that /-w/ plus F does not "stand out" as much as expected from /-w/ with other tones, rightmost orange arrow in Figure 5.

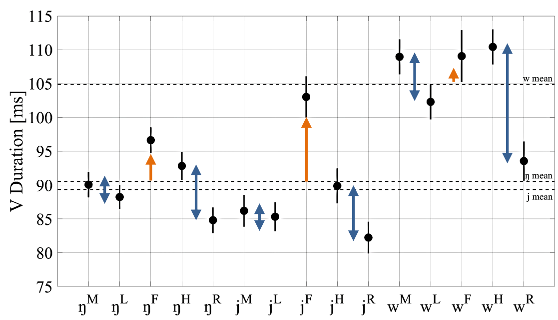

Figure 5: *Mean and 95% CI for all tones × codas*

4. Discussion

In exp. 1, we found that {M,L} toned long vowels in open syllables are systematically shorter than those in {F,H,R} toned syllables. In our data, the relative order of durational effects of tone on long vowels, {M,L} < {F,H,R}, is the *reverse* of what was reported by Abramson for BKKT [3]. These effects of tone on vowel durations are also problematic for the idea that such effects are universally inversely proportional to average f0 height of the tone [2]. In this respect, our data on long vowels in open syllables show that two tones with relatively low average f0 height, M and L, are linked to the shortest vowels. A more promising way to interpret our results is in terms of a transparent relationship between the durational effects of tone on vowels and f0 control mechanisms. It has been reported that, at least for untrained speakers, i.e., non-singers, f0 falls are produced faster than f0 rises [6], [7]. If the production of vowels and tones is temporally coordinated, so that their executions are time-locked or closely timed, our durational findings can be understood as follows. Vowels associated with the M and L tones, which comprise only faster f0 falling components are shorter; as these tones are produced with active f0 lowering laryngeal activity [11], [12]. While vowels with the F, H, and R tones, which all comprise slower f0 rising components either in the first half (F) or the second half (H, R) of the contour, Figure 1, are longer. Additionally, our findings are also in line with diachronic BKKT evidence. Vowel lengthening in BKKT is sporadic but observed almost exclusively in connection with F and H tones. This is in line with the longer duration we have observed for the {F,H} toned vowels. A problematic finding for this link between diachronic lengthening and synchronic phonetic variation due to tone is the longer duration for R toned vowels. Diachronically, some words with R tones are shortened, not lengthened, at least in BKKT (but see [2] for lengthening effects of R tones in other Thai varieties). This issue brings us to the interactions between tones and the presence of codas.

In exp. 2, we studied short vowels combined with (sonorant) codas. We found the following differences among durational effects of tone on vowels: {R} < {M,L} < {F,H}. In the presence of codas, we observed that the R tone results in shorter vowel duration, compared to M and L, while H and F still results in the longest vowel duration, just like for long vowels in open syllables. Against the background of these findings, we note again that the order {M,L} < {F,H} is the reverse of the one reported in [3] and of what is predicted by an inversely proportional relation between average f0 height and vowel duration [2]. Moreover, the durational effects {M,L} < {F,H} can again straightforwardly be explained in terms of faster f0 falls and slower f0 rises. In particular, tones that comprise only an f0 fall (M, L) are shorter than other tones. The shortest duration of the R-toned vowels is, however, slightly surprising since this tone involves both an f0 fall and an f0 rise. This last finding suggests caution in attributing all effects purely to physiological considerations, like f0 control mechanisms. Phonological structure also seems to play a role. Whatever the exact explanation behind our findings, all of these effects closely mirror diachronic changes in BKKT vowel length [8]. The F and H tones, associated with the longest vowels, are known to correlate with diachronic lengthening; while the R tone, associated with the shortest vowels when codas are present, correlates with diachronic shortening. This is exactly what we observe diachronically with short and long vowels, respectively. Finally, we also observed interactions between tonal effects and coda types that also closely mirror diachronic changes. Lengthening of words with /-j/ and F tone is consistent with an overrepresentation of this combination among words that underwent diachronic lengthening. Similarly, shortening of words with /-w/ and R is a well-attested phenomenon diachronically in BKKT, e.g., [tʰɛ̆ːw] > [tʰɛ̆w] ‘line, queue’.

Some limitations of this work and future avenues for research should also be discussed. First, we only examined the effects of tone on one vowel quality, /a/ both long and short. More vowel qualities should be studied to investigate the possibility that some effects may be vowel specific. For instance, mid vowels have been reported to be more prone to shortening effects in BKKT, accordingly, tonal effects on them may be different. Second, we reported on data from /Caː/ and /CaG/N/ environments, a logical next step is to examine /CaːG/N/ to further verify that short and long vowels behave the same in closed syllables. The extent to which other Thai varieties may or may not conform to the picture presented in this paper is also another avenue for future research. Finally, an investigation that focuses on individual differences in the strength of tonal durational effects may help shed further light on the rationale behind this phenomenon and its (non-)physiological underpinnings. Notwithstanding these limitations, the data presented in this paper suggest a consistent ordering for BKKT tonal durational effects {M,L} < {H,F}; while the effects of the R tone are dependent on syllable structure. Syllable structure also mediates the effects of other tones, like L and F, on vowel duration. The synchronic phonetic effects we observed are mostly in line with phonological vowel lengthening observed diachronically in BKKT. Given that the amount of lengthening due to tone is of moderate size, up to ~20%, we find it unlikely that phonological changes are purely the result of synchronic phonetic biases. However, the synchronic variation we have reported can reasonably be considered a precursor for sound change. The systematic phonetic biases present in the production of the vowel /a/ may lead it to move more and more towards /aː/ realizations, in association with certain tones and coda types, through iterations in the perception production loop. Eventually, this drift in duration space may lead to a new phonological length categorization for exemplars of this vowel diachronically.

5. Conclusion

In this paper, we have shown that contemporary BKKT vowel duration, for both long and short vowels, is systematically influenced by tonal categories. For long vowels in open syllables, we observed an order {M,L} < {F,H,R}. For short vowels in syllables with sonorant codas, we observed an order {R} < {M,L} < {H} < {F}, as well as interactions between tones and different coda types. This ordering of tonal effects on duration are both in contrast with previous reports [3] and problematic for the idea that vowel duration is inversely related to average f0 height [2]. We have argued that different tonal effects may be better understood if vowel duration is related to differences in the speed of f0 falls, which are faster, *vs.* rises, which are slower. In other words, longer vowel durations may be associated with tones that have a slower f0 rising component and shorter durations with tones that consist entirely of a faster f0 falling component. Crucially, these effects are not purely physiological, as they are also mediated by syllable structure and coda types. Finally, the durational effects we reported are in line with observed diachronic changes in phonological vowel length that are known to have taken place in the history of BKKT. Phonetic synchronic variation due to tonal effects on vowel duration may, thus, be considered fertile soil for phonological sound change diachronically.

6. Acknowledgements

We want to thank the members of Quantitative Methods for the Study of Language course taught at Chulalongkorn University in the Spring Semester of 2022 for their comments on the second experiment presented in this paper. We also want to thank Sireemas Maspong and Pittayawat Pittayaporn for their comments on an earlier draft of this paper. We also wish to thank three anonymous SST reviewers whose comments helped improving the quality of our work. Finally, it is our pleasure to acknowledge financial support from the Cognitive Science Program at Cornell University (FB), and from the Southeast Asian Linguistics Research Unit and the Faculty of Arts at Chulalongkorn University (TS).

7. References

[1] A. C. L. Yu, "Tonal effects on perceived vowel duration," in *Laboratory Phonology 10*, C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée, Eds. De Gruyter Mouton, 2010, pp. 151–168. doi: 10.1515/9783110224917.2.151.

[2] J. Gandour, "On the Interaction between Tone and Vowel Length: Evidence from Thai Dialects," *Phonetica*, vol. 34, no. 1, pp. 54–65, 1977, doi: 10.1159/000259869.

[3] A. S. Abramson, *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments*, vol. 28.2. Bloomington: Indiana University, 1962.

[4] E. Zee, "Duration and intensity as correlates of F0," *Journal of Phonetics*, vol. 6, no. 3, pp. 213–220, Jul. 1978, doi: 10.1016/S0095-4470(19)31153-2.

[5] Q.-M. Kong, "Influence of Tones upon Vowel Duration in Cantonese," *Lang Speech*, vol. 30, no. 4, pp. 387–399, Oct. 1987, doi: 10.1177/002383098703000407.

[6] J. J. Ohala and W. G. Ewan, "Speed of Pitch Change," *The Journal of the Acoustical Society of America*, vol. 53, no. 1, pp. 345–345, Jan. 1973, doi: 10.1121/1.1982441.

[7] J. Sundberg, "Maximum speed of pitch changes in singers and untrained subjects," *Journal of Phonetics*, vol. 7, no. 2, pp. 71–79, Apr. 1979, doi: 10.1016/S0095-4470(19)31040-X.

[8] P. Pittayaporn, "Sound changes in Thai vowel length: the significance with respect to modern spelling," Chulalongkorn University, 2016.

[9] D. Povey *et al.*, "The Kaldi speech recognition toolkit," *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*, Jan. 2011.

[10] Paul Boersma and D. Weenink, *Praat: doing phonetics by computer*. 2022. [Online]. Available: https://www.praat.org.

[11] D. M. Erickson, "A physiological analysis of the tones of Thai.," Ph.D. Dissertation, University of Connecticut, 1976.

[12] D. Erickson, "Thai Tones Revisited," *Journal of the Phonetic Society of Japan*, vol. 15, no. 2, pp. 74–82, 2011.