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

Effects of coda consonants on preceding vowel F0 in Eastern Khmu

James Kirby, Rasmus Puggaard-Rode, Francesco Burroni, Sireemas Maspong

Spoken Language Processing Group, Institute for Phonetics and Speech Processing, LMU München
{jkirby|r.puggaard|francesco.burroni|s.maspong}@phonetik.uni-muenchen.de

Abstract

This paper documents the acoustic effects of onset and coda consonants on the F0 trajectory of the preceding vowel in the Austroasiatic language Eastern Khmu. In addition to an onset voicing contrast, Eastern Khmu permits syllables ending in sonorants, unreleased plosives, a voiceless glottal fricative, and a glottal stop. While glottal stop and fricative codas are frequently implicated in tonogenesis, their coarticulatory effects on the pitch of the preceding vowel are reported to vary. Here, we analyze data from 20 speakers of Eastern Khmu producing monosyllables in which both onset voicing and coda type were varied. Growth curve analysis indicates that, relative to sonorant-final syllables, the presence of a non-sonorant final raises F0 of the preceding vowel for syllables with both voiced and voiceless onsets. Onset F0 effects are also visible in syllables with voiceless onsets, regardless of coda type. Coda-induced variations in the shape of the F0 trajectory are detectable early in the F0 excursion, a conclusion validated by a neural network classifier. We propose that the co-intrinsic effects of codas can impose a laryngeal setting affecting the F0 trajectory of the entire preceding syllable. We discuss the implications of these findings for classical models of tonogenesis.

Index Terms: microprosody, tonogenesis, coarticulation, laryngeals, coda consonants, co-intrinsic pitch, Khmu

1. Introduction

Tonogenesis, or the emergence of lexical tone, is often assumed to be the result of the transphonologization of microprosodic intrinsic or co-intrinsic F0 perturbations [1, 2]. While the microprosodic effects of prevocalic voicing on vowel F0 are well-documented [3, 4, 5, 6], less work has considered the effects of postvocalic consonants on the F0 trajectory of the preceding vowel. Unlike prevocalic voicing contrasts, there are few if any instances of postvocalic voicing constrast transphonologizing into F0-based contrasts, and phonetic evidence for a consistent effect of postvocalic obstruent voicing on vowel F0 is minimal at best [4]. Postvocalic /h/ and /ʔ/, on the other hand, are implicated in the emergence of lexical tone in a typologically diverse range of languages [7, 8, 9, 10, 11]. For example, the rising and falling tones of Vietnamese and Chinese are argued to be related to historical final glottal stops and fricatives, respectively [7, 12]. [7] sketched phonetic motivations for these correspondences (translation from [13]):

> The final fricative became a laryngeal **h** produced by an abrupt slackening of the larynx. The slackening of the vocal folds produced a drop in the pitch of the preceding vowel, i.e. a falling tone...[a] glottal stop following a vowel is produced by an increase in vocal fold tension (the opposite of what we have seen for final **h**)...the increase in vocal fold tension in anticipation of the coda glottal stop produces a rising tone.

This quote suggests that the laryngeal settings accompanying coda /h/ and /ʔ/ may exert predictable effects on the pitch contours of preceding vowels, much as voicing contrasts in onsets affect the pitch height of the following vowel. One study supporting this model with phonetic data is [4], which presents data from four speakers of an unspecified variety of Arabic showing the average F0 trajectories over the last 100 msec of the vowel in a C_1VC_2 syllable spoken in a carrier phrase. F0 preceding /ʔ/ rose between 9 and 48 Hz while F0 preceding /h/ fell by 25 to 50 Hz. A subsequent perception experiment demonstrated effects of these magnitudes to be perceptible.

However, the correlation between coda type and F0 direction predicted by [7] and reported by [4] does not seem to be nearly as cross-linguistically robust as the co-intrinsic effects of onset voicing on the following F0 height. For example, in Lhasa Tibetan, coda /ʔ/ lowers, rather than raises, the F0 of the preceding vowel [14]. Similarly, a study of the effects of final /h/ and /ʔ/ in six Southeast Asian languages found that while coda /h/ seemed to induce a falling contour over the preceding vowel, so too did a final glottal stop [15]. In Itunyoso Trique, F0 falls before final /h/, but is unperturbed before final /ʔ/ [16]. In general, the effect of coda /h/ appears to be more consistent across studies than the effect of coda /ʔ/, but here too there are exceptions. In Punjabi, vowels which were followed by /h/ or by breathy-voiced stops in Middle Indo-Aryan developed high, not low tones [17], and historical coda /h/ developed into high tones in Tsat [10] and Rục [18].

The typological variability reviewed above suggests that the physiological implementation of laryngeal codas may be considerably more complex than [7] assumed. [19] surmised that final glottals would raise or lower depending on if they are produced with abrupt closure (resulting in tense/constricted voice) or more gradual, incomplete constriction (resulting in creaky voice), respectively, a prediction also made by [11, 20] and for which there is now considerable acoustic and articulatory evidence (see e.g. [16, 21, 22]). Similarly, a final glottal fricative may be realized as either breathy [ɦ] or non-breathy [h], with correspondingly different effects on the preceding F0 trajectory. There is also increasing historical evidence that in at least some tone languages, the effects of final laryngeals was mediated by onset-conditioned differences in phonation type [23, 24, 18], but how the articulation of onsets and codas interact in terms of their co-intrinsic effects on pitch remains poorly understood.

1.1. The present study

The aim of the present study is to contribute new empirical data on the effects of final laryngeal consonants on preceding vowel F0, with the goal of better understanding how the articulation of these segments can influence both the magnitude and temporal extent of the preceding F0 trajectory. For this study, we focus on Khmu `[kɟg]`, an Austroasiatic language spoken in Laos and neighboring regions of China, Thailand, and Vietnam. While some varieties of Khmu have transphonologized

10.21437/SpeechProsody.2024-66

onset voicing into a pitch-based contrast [25, 26, 27], Eastern Khmu (EK) is a phonologically conservative variety which retains a voicing contrast in both initial obstruents and sonorants [28, 29]. Like other Khmu varieties, EK permits syllables to end in a vowel or final sonorant (so-called *smooth* syllables), an unreleased obstruent /p t c k/ (so-called *checked* syllables), a final glottal fricative /h/, or a final glottal stop /ʔ/. Thus in EK it is possible to have four-way near-minimal quadruplets contrasting only in coda, e.g. /buː/ 'puffy, swollen', /buːc/ 'liquor', /buh/ 'carry with a tumpline', /buʔ/ 'breast, milk'. This rich inventory of finals, together with a voicing contrast in onset position and a lack of lexical tone, makes EK a natural laboratory in which to study how the laryngeal properties of onsets and codas interact in their effects on the F0 trajectory of the syllable.

2. Methods and materials

2.1. Speech materials

Our master wordlist of 125 monosyllables included 72 smooth syllables, 14 checked syllables, 11 syllables with a coda /h/ and 28 syllables with a coda /ʔ/. The range of syllable onsets was diverse and designed to study several other research questions not considered here (see [29, 30]). In this work, we set aside syllables beginning with aspirated or preglottalized onsets and restrict analysis to the 105 items beginning with either voiced /b d g m n ŋ l r j/ or voiceless /p t k m̥ n̥ ŋ̊ l̥ r̥ j̊ s/ onsets.

2.2. Participants

20 speakers of EK (12 female and 8 male, 21–69 at time of recording) were recorded using a headset condenser microphone in a quiet, sound-treated booth in Vientiane, Lao PDR in January 2020. Each speaker produced each target item twice in isolation and twice in the carrier phrase /ʔoʔ cə law ___ ʔan klɔh/ (1sg IRR speak ___ SBJV clearly) "I will say ___ clearly". Participants produced the EK form in response to an oral prompt of the Lao gloss by an assistant. Recordings were made direct to laptop using the SpeechRecorder software [31].

2.3. Acoustic analysis and preprocessing

F0 was calculated at 5 ms intervals with speaker-specific pitch floor and ceiling values throughout the sound files using PraatSauce [32], a suite of Praat scripts designed for extracting F0, formants, and various spectral measures. F0 measures were removed if they fell above or below three standard deviations of the mean within the same vowel category of the same speaker. Remaining F0 measurements were converted to semitones using the speaker mean as the reference level.

The start and end of the nuclear vowel were segmented manually in Praat [33], as well as various landmarks relating to voicing implementation at the syllable onset. We used these manual annotations to determine vowel onset, but due to the uncertainty associated with coda boundaries, we estimated vowel offset automatically using *convex hull segmentation*, a programmatically determined intensity cutoff [34, 35].

8Following [34, 35], we first computed the convex hull function for the window containing the intensity trajectories of the entire syllable rime. This function exhibits a monotonically non-decreasing behavior from the beginning of the segment to its point of maximum intensity, and then becomes monotonically non-increasing until the end of the window. We searched for the maximum distance between the intensity trajectories and the convex hull function. If the maximum distance exceeded a threshold of 3 dB, the window was narrowed. The end of the new window was set to the location of the maximum distance, and the convex hull was recomputed for this adjusted window. This iterative process continued until the maximum distance fell below the threshold of 3 dB. The endpoint of the final window was selected as the boundary between the vowel nucleus and the consonantal coda.

The resulting collection of sound files and annotations were stored as an EMU database [36], and subsequent work with the database was accomplished using the `emuR` package in R [37].

2.4. Statistical analysis

The F0 trajectory data were statistically analyzed in R [37]. Visual inspection of the empirical data suggested that the shape variation in pitch contours can be captured using a third-order polynomial, so we analyzed the data using growth curve analysis (GCA) [38]. For the GCA analysis, we further filtered the wordlist to remove open smooth syllables, resulting in a list of 92 items ending in some type of coda, and considered syllables produced in carrier phrase context only.

Using the `lme4` library [39], we fitted the F0 to a linear mixed-effects regression using the following predictor variables (reference levels are in boldface when relevant): CODA CLASS (***sonorant***, *checked*, /h/, or /ʔ/); ONSET CLASS (***voiced*** or *voiceless*); REPETITION (***first*** or *second*); VOWEL HEIGHT (***low***, *mid*, or *high*); speaker GENDER; and vowel DURATION. DURATION values were standardized. GENDER was sum coded, since there is no meaningful reference level for this variable. Interactions were further included between CODA CLASS:GENDER and CODA CLASS:ONSET CLASS; no other interactions were found to improve the model fit.

The F0 trajectory shapes were modeled by including orthogonal polynomial functions corresponding to each pitch trajectory as predictor variables, capturing the linear slope as well as quadratic and cubic curvature; interactions between these third-order polynomial terms were modeled for each predictor variable and interaction included in the model. GCA models are very informative insofar as they estimate effects and summary statistics separately for each shape parameter captured by the polynomial functions. This is useful if one wants to determine whether, for example, syllables with coda /h/ differ from syllables with sonorant codas in terms of their constant F0 level, or the linear slope of their F0 contour, or the quadratic curvature of their F0 contour, etc.

The final model includes by-participant random intercepts, and by-participant random slopes for the third-order polynomial functions. It was not possible to fit the model with a more elaborate random effects structure.

2.5. Signal chopping neural network analysis

To locate in time information concerning different coda classes, we supplemented our GCA analyses with a signal chopping neural network (NN) analysis [40]. With this analysis we predict EK coda classes (5 classes: *no coda*, *sonorant*, *checked*, /h/, or /ʔ/) based on F0 contours, ΔF0 contours (i.e., the difference between adjacent measures), both raw and z-scored by speaker (4 features $\times$ n samples in each trajectory). The contours have been time-warped to their median to prevent classification based on durational information. Numerical embeddings for categorical variables "subject" and "elicitation context" were also used as input features $(2 \times n)$. Input features are, thus, $6 \times n$ matrices.

In the signal chopping analysis, we fed the signals matrices and then "chopped" versions of them (first 100%, then 90%, 80% ... to 10%) to an NN architecture whose task was to predict coda category. We then determined at what percentage of the signal the neural network performance exceeded chance accuracy, suggesting that information regarding these categories is already present at that point in the signal. We trained 10 networks for each step, for a total of 10 steps × 10 networks = 100 networks. We thus can have a mean and standard error for accuracy from 10% to 100% of the signals.

The NNs were trained using a randomly selected 80% of the data, while 10% was used for validation and hyper parameter tuning, the final 10% was used as a test set and the accuracy we report is based on that 10%. During training, the number of tokens for each coda category was balanced via oversampling to obtain a data set where above-chance accuracy cannot be due to differences in the number of sampled tokens by category.

The network architecture we chose is a Bidirectional Long Short Term Memory (BiLSTM) [41] with self-attention [42]. We used a BiLSTM layer with 100 hidden units, followed by a dropout layer to prevent over-fitting (with dropout probability = .75), followed by a multi-head self-attention layer (with 12 heads and 24 channels), followed by a fully connected layer, and finally by a classification layer. Our initial learning rate was set at $5e-4$ with a drop-factor of .9 every 1 epochs. Other hyperparameters include: a minibatch size of 128, a validation patience of 100, and a maximum number of epochs of 1000.

Chance accuracy was calculated based on a binomial formula that takes into account both the number of classes and the number of samples, following the recommendation of [43]. We calculated chance level for a p-value $< 1e-5$, 1014 test samples, and 5 classes.

3. Results

3.1. Growth curve analysis

The GCA model structure summarized in §2.4 expands to more than 60 fixed effects terms, so for reasons of space, we do not provide the full model summary here but focus on effects of particular interest. Each of the categorical predictors mentioned above as well as DURATION have a significant influence (at the conventional $p < 0.05$ level) either on the constant F0 level or on at least one of the shape parameters.

By weighing a third-order polynomial function by the fitted effects for the polynomial terms, we can recreate the F0 trajectories predicted by the model for different levels of categorical variables. Figure 1 shows the predicted level and trajectory differences for the interaction between CODA CLASS and ONSET CLASS faceted by ONSET CLASS; Figure 2 shows the same data faceted by CODA CLASS. The intercept is estimated at -1.71 semitones ($SE = 0.14, t = -12.45, p < 0.01$), and the shape parameters predict a negative slope ($\hat{\beta} = -2.28, SE = 0.54, t = -4.23, p < 0.01$), negative quadratic curvature ($\hat{\beta} = -1.49, SE = 0.16, t = -9.51, p < 0.01$), which can in this case be interpreted as a steeper drop in the final half of the trajectory than predicted by the slope alone, and a slight negative cubic curvature ($\hat{\beta} = -0.24, SE = 0.11, t = -2.23, p = 0.03$); this corresponds to the shape and position of the black line in the “voiced onset” facet of Figure 1. To give an example of how the levels are calculated, syllables with coda /h/ are estimated at a higher level ($\hat{\beta} = 0.27, SE = 0.02, t = 13.17, p < 0.01$), with no stable difference in linear slope ($\hat{\beta} = -0.03, SE = 0.11, t = -0.26, p = 0.79$), a less negative quadratic curvature (i.e., less steep trajectory; $\hat{\beta} = 0.41, SE = 0.11, t = 3.6, p < 0.01$), and a less negative cubic curvature ($\hat{\beta} = 0.48, SE = 0.11, t = 4.25, p < 0.01$); these values match the level and shape of the red line in the ‘voiced onset’ of Figure 1.

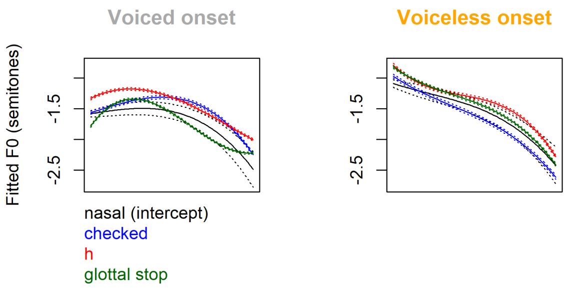

Figure 1: *Model-predicted F0 trajectories by onset class. Dotted lines indicate standard errors.*

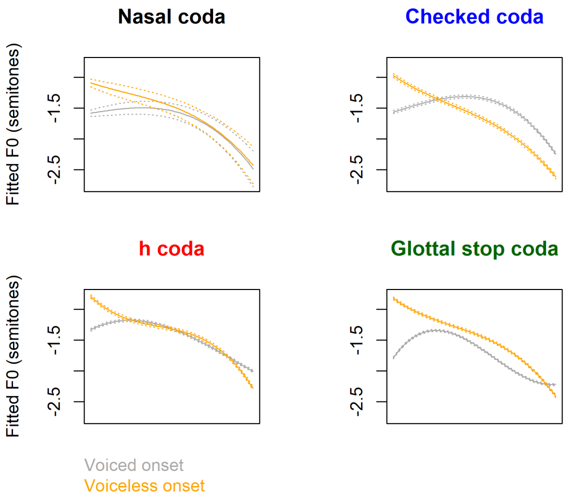

Figure 2: *Model-predicted F0 trajectories by coda class. Dotted lines indicate standard errors.*

In syllables with voiced onsets (the reference level of the ONSET CLASS variable), syllables with coda /h/ are predicted as having a higher F0 than the baseline nasal coda throughout the trajectory, beginning very early in the trajectory. Syllables with checked codas are also predicted as having higher F0 than the baseline throughout the trajectory, although this difference is less pronounced. There is no predicted level difference between /ʔ/ and the baseline, although the model does predict a shape difference, particularly in cubic curvature.

Voiceless onsets exert an influence on F0 that is particularly prominent early in the trajectory. /h/ and /ʔ/ codas are both estimated as having higher F0 than the nasal baseline in the early stages of the trajectory. For /h/, but not for /ʔ/, this effect largely continues throughout the trajectory. Generally, coda-induced divergences from the nasal baseline are smaller after voiceless onsets, perhaps because F0 shape variability early in the trajectory is constrained by the onset-induced F0 perturbation.

In addition to the coda and onset perturbations, VOWEL HEIGHT exerts a large influence on F0. *High* vowels in particular have a much higher level (estimated at 2.3 semitones higher than the intercept), and a less negative slope but a steeper fall towards the end of the trajectory; *mid* vowels have less prominent effects in the same directions. REPETITION also exerts an influence on F0, such that the second repetition of an item has a higher F0 level, steeper negative slope, and more negative quadratic curvature. DURATION does not affect overall F0 level, but affects all shape parameters, such that longer vowels have a steeper negative slope, more negative quadratic curvature, and less negative cubic curvature. There is no main effect of GENDER on either the level or shape of F0, but GENDER interacts with CODA CLASS on multiple levels, with particularly male speakers having a much steeper predicted negative slope before /h/; we have no explanation for this at present.

3.2. Neural network analysis

As is evident in Figure 3, the NNs achieve a final accuracy of $\approx 70\%$ in predicting coda classes based on F0 contours (both raw and z-scored), their first derivative, and subject and elicitation information. As expected, the neural network accuracy gets better and better towards the coda, which is the source of the F0 perturbation. Interestingly, the NN accuracy is almost $\times 2$ chance already at the first time step (10%) of the signals. This fact suggests that different coda classes can be differentiated based on their F0 and F0 rate of change very early in the trajectory; thus, their perturbation, albeit clearly originating from the coda, seem to be a non-local effect.

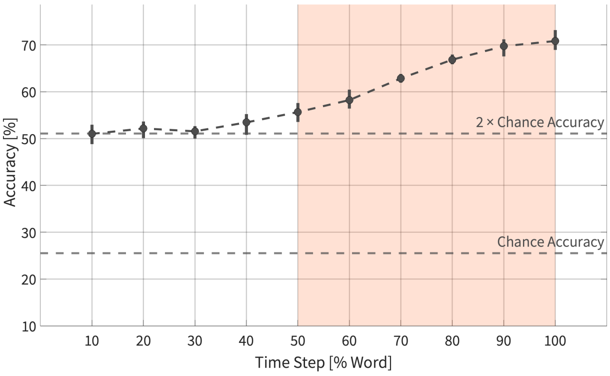

Figure 3: *Prediction of coda accuracy for different choppings.*

4. Discussion

Our acoustic analysis found that the syllabic F0 trajectory in EK was significantly raised before a coda /h/ and /ʔ/ when compared to syllables ending with sonorant codas (/m n ŋ/). The magnitude of the effect is greatest preceding /h/ and weakest preceding /ʔ/, with checked syllables having an effect somewhat in between. This was broadly the case for syllables with both voiced and voiceless onsets. However, the magnitude of the effect interacts with onset voicing, being more prominent when the onset was voiced than when it was voiceless. Checked syllables exert a smaller, but still measurable raising effect when the onset is voiced, but not when it is voiceless.

Although co-intrinsic effects of onset voicing were also observed, these were limited in magnitude (≈ 0.5 semitones) and temporally restricted to the beginning of the trajectory. Conversely, the effect of the checked and glottal fricatives of F0 was temporally distributed over the entire preceding vowel, to a greater or lesser degree. This is less obviously the case for syllables with coda /ʔ/, where the effect is strongest in the beginning of the excursion. At least in syllables with voiced onsets, F0 also appears to drop off at the end of syllables with coda /ʔ/, suggestive of creak, but we note that F0 estimation becomes difficult in the vicinity of many types of glottals. Overall, however, our data suggest that both glottal fricative and glottal stop codas in EK are gradually phased with respect to the preceding vowel, which would explain their relatively large coarticulatory effects on the vocalic F0 trajectory. This is arguably not especially surprising; EK is not a tone language, so there is no functional pressure to phase gestures in such a way as to maximize perceptual recovery of the F0 contour (cf. [16, 20]).

It is interesting to note that the F0 contour on all syllables in our study was falling, despite the fact that these were elicited in a carrier phrase context. We suspect this may simply reflect global F0 declination over the (short) utterance, but we note that it is consistent across speakers and utterances in our corpus.

These findings do not conform with the predictions of the classical tonogenetic model [7]; however, this may be due at least in part to a general failure to systematically distinguish abrupt from gradually phased /ʔ/ [16] and breathy [ɦ] from non-breathy [h] [44]. It may well have been the case that in the particular instances of pre-tonal Vietnamese and Chinese, coda /h/ syllables were breathy-voiced [ɦ] (with an accompanying pitch fall) while coda /ʔ/ syllables were tense/constricted [ʔ] (with an accompanying pitch rise), and that these effects transphonologized in a relatively straightforward manner (but cf. [18, 23, 24]). Our data simply suggest that the articulation of these segments in EK is rather different than predicted by the classical account. In particular, coda /h/ appears to be realized as non-breathy [h], and coda /ʔ/ may be realized as tense initially but creaky finally, which would explain the raised initial F0 but lowered final F0 (cf. [45]).

In this work, we have restricted our focus to F0, because this is the primary acoustic exponent of lexical tone. However, this is obviously far from the only acoustic property which can be affected by coda type. For instance, [46] points out that the percept of a coda /ʔ/ can be cued by a number of cues including duration, spectral tilt, and F0. F0 differences might therefore not need to be consistently present as phonetic precursors in order for tonogenesis to take place. Listeners may thus only need to learn that F0 *can* serve as a cue in order to select it as a target for (trans)phonologization. In ongoing work we are expanding our analysis to consider a broader range of acoustic cues and their interactions.

5. Conclusions

In this paper, we have presented an acoustic study of coda-induced F0 perturbations in Eastern Khmu. Using growth curve analysis, we found that both coda /ʔ/ and coda /h/ raise F0 compared to other codas in this language. These findings do not conform to the predictions of the classical tonogenetic model [7]. Results of a signal chopping neural network analysis further indicate that these effects span the entire preceding vowel. While this type of temporally distributed coarticulatory effect on F0 is precisely the kind of phonetic precursor that seems likely to transphonologize into lexical tone, the interactions observed between onset voicing and coda type suggest that the effects of laryngeal codas on the F0 and phonation of preceding vowels are more complex and language-specific than often assumed [18]. Further research on the language-specific implementation of laryngeal coarticulation and its effects on F0 (both acoustic and perceptual) is needed to better understand the phonetic underpinnings of tonogenesis.

Funded in part by ERC Starting Grant No. 758605.

6. References

[1] L. M. Hyman, “Phonologization,” in *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*, A. Julliand, A. M. Devine, and L. D. Stephens, Eds. Saratoga,CA: Anma Libri, 1976, pp. 470–418.

[2] A. Michaud and B. Sands, “Tonogenesis,” *Oxford Research Encyclopedia of Linguistics*, 2020.

[3] A. S. House and G. Fairbanks, “The influence of consonant environment upon the secondary acoustic characteristics of vowels,” *The Journal of the Acoustical Society of America*, vol. 25, no. 1, pp. 105–113, 1953.

[4] J.-M. Hombert, “Consonant types, vowel quality, and tone,” in *Tone: a linguistic survey*, V. Fromkin, Ed. San Diego: Academic Press, 1978, pp. 77–111.

[5] K. J. Kohler, “F0 in the production of fortis and lenis plosives,” *Phonetica*, vol. 39, pp. 199–218, 1982.

[6] H. M. Hanson, “Effects of obstruent consonants on fundamental frequency at vowel onset in English,” *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 425–441, 2009.

[7] A.-G. Haudricourt, “De l’origine des tons en viêtnamien,” *Journal Asiatique*, vol. 242, pp. 69–82, 1954.

[8] T.-l. Mei, “Tones and prosody in Middle Chinese and the origin of the rising tone,” *Harvard Journal of Asiatic Studies*, vol. 30, pp. 86–110, 1970.

[9] M. Dürr, “A preliminary reconstruction of the Proto-Mixtec tonal system,” *Indiana*, vol. 11, pp. 19–62, 1987.

[10] G. Thurgood, “Language contact and the directionality of internal drift: the development of tones and registers in Chamic,” *Language*, vol. 72, no. 1, pp. 1–31, 1996.

[11] J. Kingston, “The phonetics of Athabaskan tonogenesis,” in *Athabaskan prosody*, ser. Current Issues in Linguistic Theory, S. Hargus and K. Rice, Eds. Amsterdam: John Benjamins, 2005, no. 269, pp. 137–184.

[12] E. G. Pulleyblank, “The nature of Middle Chinese tones and their development,” *Journal of Chinese Linguistics*, vol. 6, no. 2, pp. 173–203, 1978.

[13] A.-G. Haudricourt, “The origin of tones in Vietnamese,” <halshs-01678018>, 2018.

[14] F. Hu and Z. Xiong, “Lhasa tones,” in *Proceedings of the 5th International Conference on Speech Prosody*, 2010.

[15] A. Thavisak, “The effects of glottal finals on pitch in SEA languages,” in *The 33rd International Conference on Sino-Tibetan Languages and Linguistics*. Bangkok: Ramkhamhaeng University, 2000, pp. 372–377.

[16] C. T. DiCanio, “Coarticulation between tone and glottal consonants in itunyoso trique,” *Journal of Phonetics*, vol. 40, no. 1, pp. 162—-176, 2012.

[17] T. K. Bhatia, “The evolution of tones in Punjabi,” *Studies in the Linguistic Sciences*, vol. 5, no. 2, pp. 12–24, 1975.

[18] T. T. Tạ, “Register and tone developments in Vietic languages,” PhD dissertation, University of Ottawa, 2023.

[19] M. Mazaudon, “Tibeto-Burman tonogenetics,” *Linguistics of the Tibeto-Burman Area*, vol. 3, no. 2, pp. 1–123, 1977.

[20] D. Silverman, “Phasing and recoverability,” PhD dissertation, University of California at Los Angeles, 1997.

[21] J. A. Edmondson and J. H. Esling, “The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies,” *Phonology*, vol. 23, no. 2, pp. 157–191, 2006.

[22] P. Keating, M. Garellek, and J. Kreiman, “Acoustic properties of different kinds of creaky voice,” in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 2015.

[23] K. J. Gregerson, “Tongue-root and register in Mon-Khmer,” *Oceanic Linguistics Special Publications*, no. 13, pp. 323–369, 1976.

[24] R. Gehrmann, “Desegmentalization: Towards a common framework for the modeling of tonogenesis and registrogenesis in Mainland Southeast Asia with case studies from Austroasiatic,” PhD dissertation, University of Edinburgh, 2022.

[25] S. Premsrirat, “Tonogenesis in Khmu dialects of Southeast Asia,” *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*, pp. 121–134, 2001.

[26] J.-O. Svantesson and D. House, “Tones and non-tones in Kammu dialects,” in *Proceedings of Fonetik 96, Swedish Phonetics Conference, TMH-QPSR 2/1996*, 1996, pp. 85–87.

[27] ——, “Tone production, tone perception and Kammu tonogenesis,” *Phonology*, vol. 23, no. 02, pp. 309–333, 2006.

[28] R. A. Osborne, “A phonology of Eastern Kmhmu’ with special reference to palatal continuant codas and neutralisation of vowel length contrast,” *Journal of the Southeast Asian Linguistics Society*, vol. 11, no. 2, pp. 67–86, 2018.

[29] J. Kirby, P. Pittayaporn, and M. Brunelle, “Transphonologization of onset voicing: revisiting Northern and Eastern Kmhmu’,” *Phonetica*, vol. 79, no. 6, pp. 591–629, 2023.

[30] J. Kirby, “Effects of voiceless and preglottalized nasals on f0 in Eastern Khmu (Kmhmu’ Am),” *Papers from the 30th Conference of the Southeast Asian Linguistics Society (2021)*, pp. 318–333, 2022.

[31] C. Draxler and K. Jänsch, “SpeechRecorder - a universal platform independent multi-channel audio recording software,” in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, 2004, pp. 559–562.

[32] J. Kirby, “praatsauce: Praat-based tools for spectral analysis [version 0.2.6],” https://github.com/kirbyj/praatsauce/, 2018.

[33] P. Boersma and D. Weenink, “Praat: doing phonetics by computer [version 6.4],” http://www.praat.org, 2023.

[34] P. Mermelstein, “Automatic segmentation of speech into syllabic units,” *Journal of the Acoustical Society of America*, vol. 58, no. 4, pp. 880–883, 1975.

[35] L. Eriksson, “Algorithms for Automatic Segmentation of Speech,” in *Lund University, Dept. of Linguistics Working Papers*, 1989, vol. 35, pp. 53–61.

[36] R. Winkelmann, J. Harrington, and K. Jänsch, “EMU-SDMS: Advanced speech database management and analysis in R,” *Computer Speech & Language*, vol. 45, pp. 392–410, 2017.

[37] R Core Team, “R: A language and environment for statistical computing.” Vienna, Austria, 2022.

[38] D. Mirman, *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press, 2014.

[39] D. Bates, M. Maechler, and B. Bolker, “lme4: Linear mixed-effects models using Eigen and S4.” 2014.

[40] S. Tilsen, “Detecting anticipatory information in speech with signal chopping,” *Journal of Phonetics*, vol. 82, p. 100996, 2020.

[41] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[43] E. Combrisson and K. Jerbi, “Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy,” *Journal of neuroscience methods*, vol. 250, pp. 126–136, 2015.

[44] G. Thurgood, “Vietnamese tone: revising the model and the analysis,” *Diachronica*, vol. 19, no. 2, pp. 333–363, 2002.

[45] E. Fischer-Jørgensen, “Phonetic analysis of the stød in Standard Danish,” *Phonetica*, vol. 46, no. 1, pp. 1–59.

[46] C. DiCanio, “Cue weight in the perception of Trique glottal consonants,” *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 884–895, 2014.