Voice quality is not an obligatory stage in tonogenesis: A case study of Eastern Khmu

Sireemas Maspong

Department of Linguistics, Cornell University
Institute for Phonetics and Speech Processing (IPS), LMU Munich
s.maspong@phonetik.uni-muenchen.de

ABSTRACT

This paper examines the role of voice quality in the development of tonal and vowel height contrasts that accompany tonogenesis. The study focused on Eastern Khmu because it maintains a contrastive onset voicing, unlike other varieties of Khmu which have undergone tonogenesis and lost this contrast. The study shows that pitch and vowel height differences on following vowels accompany the onset voicing contrast without any concomitant voice quality difference. These findings suggest that voice quality is not an obligatory factor in deriving tonal and vowel height contrasts. Additionally, the study suggests that a tonal or vowel height language can emerge directly from a reanalysis of onset-induced acoustic perturbations such as pitch and vowel height without a stage where voice quality is contrastive.

Keywords: Tonogenesis, Register, Tone, Voice quality, Vowel height, Khmu

1. INTRODUCTION

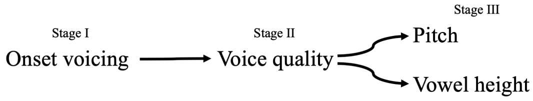
Cross-linguistically, particularly in the Austroasiatic language family, there exist languages in which onset voicing in the proto-language corresponds to present (f0-based) tonal and vowel height contrasts. The proto-onset voicing also corresponds to present register contrasts in some other languages. Register typically consists of a binary contrast, high and low register, that encompasses laryngeal (such as voice quality and pitch) and supralaryngeal properties (such as vowel quality) [1, 2]. The contrasts are summarized in Table 1.

Contrasts	*pa	*ba	Example
Tonal (f0)	Higher f0	Lower f0	N. Khmu
Vowel height	More open vowels	More close vowels	Standard Khmer
Register	Higher f0 Modal voice More open vowels	Lower f0 Breathy voice More close vowels	Mon

Table 1: Summary of contrasts conditioned by onset voicing

The present contrasts emerge from the loss of onset voicing [3] through a process called tonogenesis [4] or registrogenesis [5]. Some scholars propose that a stage with contrastive voice quality, which is a part of register contrast, is an obligatory intermediate stage between the stage with onset voicing contrast and the stage with f0-based tonal contrast or vowel height contrast [e.g., 3, 6]. In other words, voice quality is considered an obligatory stage in emergence of pitch contrasts [7, 8] or vowel height contrasts [8]. However, recent studies have shown that a pitch contrast can derive directly from the loss of onset voicing contrast without going through an intermediate stage with a voice quality contrast [9]. Although evidence for the emergence of vowel height contrast has not been found, the possibility of deriving contrasts without an obligatory intermediate stage suggests alternative pathways, as summarized in Fig. 1.

(i) Voice quality is a crucial intermediate stage



(ii) Voice quality is *not* a crucial intermediate stage

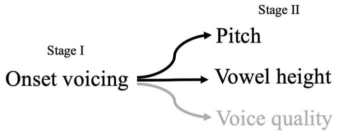


Figure 1: Two possible pathways of tonogenesis/registrogenesis

In this paper, I aim to investigate whether voice quality is an obligatory factor in deriving f0-based tonal contrast and vowel height contrast. There are two opposing hypotheses: (i) voice quality is an obligatory factor, and (ii) voice quality is not an obligatory factor.

To explore this, I focus on Eastern Khmu, an Austroasiatic language at the first stage of tonogenesis where onset voicing contrasts are still present. According to Premsrirat [10], there are Eastern Khmu dialects with onset voicing contrast (Stage I in Fig. 1-i), Northern Khmu dialects with register (voice quality-based) contrast (Stage II in Fig. 1-i), and Northern and Western Khmu dialects with tonal (f0-based) contrast (Stage III in Fig. 1-i).

If voice quality is an obligatory factor (Fig. 1-i), we would expect that in Eastern Khmu, when the difference in voice quality of the vowel following voiced and voiceless onsets is not present, there should not be differences in pitch and vowel height. On the other hand, if voice quality is not an obligatory factor (Fig. 1-ii), the differences in pitch and vowel height can be present, even if the difference in voice quality is not present. The predictions in the case that the voice quality difference is not present are summarized in Table 2.

Differences	(i) Voice quality is obligatory	(ii) Voice quality is *not* obligatory
Voice quality	✗	✗
Pitch	✗	✓
Vowel height	✗	✓

Table 2: Predictions in the case that the voice quality difference is *not* present for the hypothesis

The findings of the acoustic study we present indicate that the predicted differences in voice quality are not present in the vowels following voiced and voiceless onsets, yet differences in pitch and vowel height are observed. These results suggest that pitch and vowel height can undergo transphonologization directly from onset voicing, bypassing an intermediate stage where voice quality is contrastive.

2. METHODS

2.1 Data collection

The study was conducted with 19 native speakers of E. Khmu (14F; 5M). All participants are bilingual in Thai. None of the participants reported any speech or hearing impairments. The target words consisted of monosyllabic or sesquisyllabic Khmu words with /iː, uː, eː, oː, aː/ vowels in the main syllables. The onsets of the target syllables were limited to alveolar and velar consonants, and included voiced stops, voiceless (unaspirated) stops, and voiceless aspirated stops. The participants were prompted to produce the target words in a carrier sentence with five repetitions, using Thai translations of the target words displayed on a screen and pronounced by a native Thai speaker. This method was chosen due to the lack of a writing system in E. Khmu and the possibility of some participants being unable to read Thai. Audio recordings of the participants were made using a Shure SM10A head-mounted microphone and an audio interface Roland OCTA-CAPTURE connected to a PC laptop, with a sampling rate of 44.1 kHz.

2.2 Data processing

The acoustic measurements extracted from the vowel of the target words were fundamental frequency (f0), First Formant (F1), and the amplitude difference between the first and second harmonics (H1-H2) and Cepstral Peak Prominence (CPP), which have been reported to successfully distinguish voice quality [11]. To extract f0, a cross-correlation algorithm in Praat [12] was used (window size: 13.3 ms for male, 10 msec for female; time step: 3.3 ms for male, 2.5 ms for female). For F1, the Berkform toolbox [13], a MATLAB toolbox for formant tracking with a Robust LPC algorithm, was used. H1-H2 was obtained from the inverse-filtered FFT spectra using a function in the Voicebox toolbox [14] in MATLAB. Finally, CPP was obtained from real cepstra extracted by a function in the Signal Processing toolbox [15] and calculated following the method in [16]. The window size and time step of F1, H1-H2, and CPP tracking were three times the size of those used for f0 tracking.

2.3 Data analysis

The acoustic trajectories of target vowels were analyzed using Generalized Additive Mixed Models (GAMMs) fitted with the *R* package `mgcv` [17]. Each acoustic measure (f0, F1, H1-H2, CPP) was used as a dependent variable, while onset voicing was the main predictor. Prior to the analysis, each acoustic trajectory was time-warped to have 40 datapoints using MATLAB's linear interpolation function. The subject was included as random smooths. Places of articulation (alveolar and velar) were included as a fixed effect in the null models with F1 as a response. For the corresponding alternative, an interaction with onset voicing was also included. Place of articulation is included only for model with vowel /iː, uː, aː/, since they have a full cross set of the onset voicing and the place of articulation.

Separate models were fitted for each vowel phoneme: /iː, uː, eː, oː, aː/. In total, there were 20 models (4 cues x 5 vowels) that were each compared with a corresponding null model without the fixed effect of onset voicing using a likelihood ratio test.

3. RESULTS

3.1 F0

Based on Table 1, we expect that vowels following voiced onsets will exhibit a lower f0 compared to vowels following voiceless onsets. The results reveal

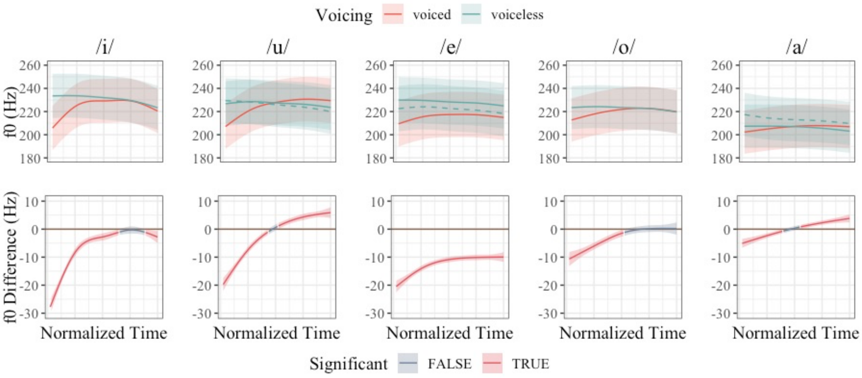


Figure 2: GAMM fitted f0 trajectories following voiced (red) and voiceless stops (green) (top) and the predicted mean difference across onset categories (bottom); dashed lines represent vowels following voiceless aspirated stops

that, as predicted, f0 values are lower following voiced onsets, at least at the onset of all vowels. Vowel /eː/ maintains the difference in f0 between following voiced and voiceless onsets throughout the entire trajectory, although the differences are more significant at the onset of the trajectories. The vowels following voiceless aspirated stops pattern with those following voiceless stops, as illustrated in Fig. 2 above.

3.2 F1

According to Table 1, vowels following voiced onset are expected to display lower F1 (higher vowel) than those following voiceless onsets. The results indicate that F1 values are lower after voiced onsets, as expected, at least at the beginning of all vowels. Vowels /iː, eː, oː/ show a sustained difference in F1 between following voiced and voiceless onsets throughout the entire trajectory, with larger differences at the onset of the trajectory. Vowels following voiceless aspirated stops display similar patterns to those following voiceless stops, as illustrated in Fig. 2.

Moreover, the effect of voicing persists despite the presence of place of articulation being included as a fixed effect. Vowels following velar onsets have lower F1 than those following alveolar onsets, but the effect of voicing remains evident, as shown in Fig. 3 below.

3.3 H1-H2

According to Table 1, we expect vowels following voiced onset to display higher H1-H2 (breathy voice) than vowels following voiceless onsets (modal voice). The results show that only vowel /oː/ displays higher H1-H2 following voiced onsets, as expected, and retains the difference in H1-H2 throughout the whole trajectory. However, unexpected patterns are observed for the other vowels. Vowels /eː, aː/ display higher H1-H2 values following voiced onsets, but not at the beginning of the trajectories. The trajectories of vowels /iː, uː/ display lower H1-H2 following voiced onsets than following voiceless onsets (breathy following voiceless onsets). The difference very small (~1–2 dB) for all vowels. Vowels following voiceless aspirated stops, again, pattern with those following voiceless stops, except for vowel /aː/. See Fig. 4 on the next page.

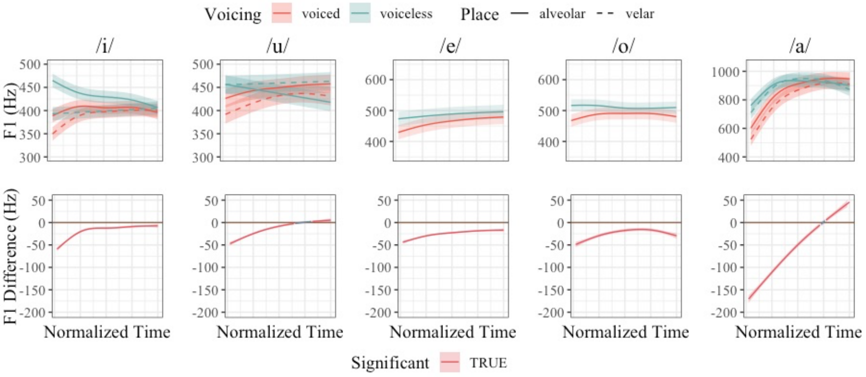


Figure 3: GAMM fitted F1 trajectories following voiced (red) and voiceless stops (green) (top) and the predicted mean difference across onset categories (bottom); dashed lines represent vowels following velar stop

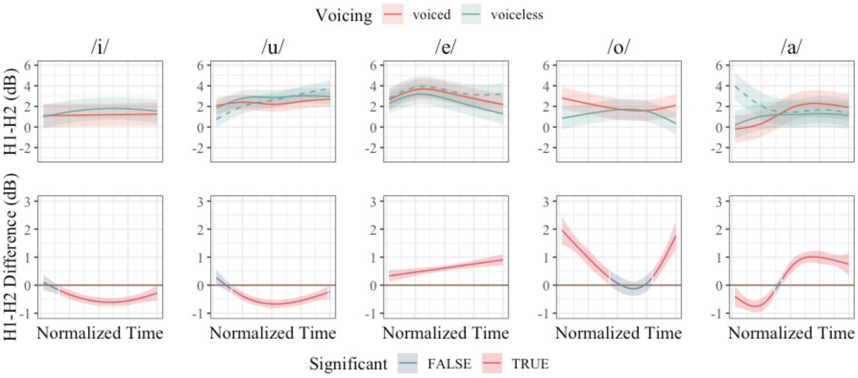


Figure 4: GAMM fitted H1-H2 trajectories following voiced (red) and voiceless stops (green) (top) and the predicted mean difference across onset categories (bottom); dashed lines represent vowels following voiceless aspirated stop

3.4 CPP

According to Table 1, it was expected that vowels following voiced onset would display lower CPP (breathy) than vowels following voiceless onsets (modal). However, the results show that only vowel /e:/ displays lower CPP following voiced onsets than following voiceless onsets, and retains the difference in CPP throughout the whole trajectory. The difference is more prominent in the middle of the trajectories. Unexpected patterns (breathy following voiceless onsets) are observed for CPP trajectories of the other vowels. These trajectories display higher CPP following voiced onsets than following voiceless onsets, either at the beginning of the trajectories or throughout the whole trajectories. The difference is minimal (~1 dB) for all vowels. The corresponding figures can be found in Fig. 5 below.

4. DISCUSSION

The findings of this study indicate that the E. Khmu variety exhibits differences in f0 and F1 values across vowels following voiced and voiceless onsets. However, while some vowels exhibit differences in voice quality across onset voicing categories, these differences are not consistent and do not align with the predictions as outlined in Table 2-ii. Thus, I conclude that voice quality is not an obligatory factor in deriving a tonal contrast based on f0 and a vowel height contrast based on F1 in this dialect.

These results suggest that an onset voicing language such as the E. Khmu could directly develop a tonal contrast or a vowel height contrast from a reanalysis of f0 and F1 differences, similar to other Austroasiatic languages, without the need for a prior contrast in voice quality. Moreover, the saliency of the observed differences at the beginning of the vowels indicates that they are likely an effect of onset voicing, a pattern that has been observed in other languages [e.g., f0: 18, 19, 20; F1: 21, 22]. Previous research has attributed f0 and F1 differences accompanying voicing to articulatory mechanisms that either facilitate or hinder voicing, such as larynx lowering, tongue root advancement [23, 24, 25, 26], or vocal fold tension [27]. We thus hypothesize that onset voicing in E. Khmu may involve one or more of these mechanisms.

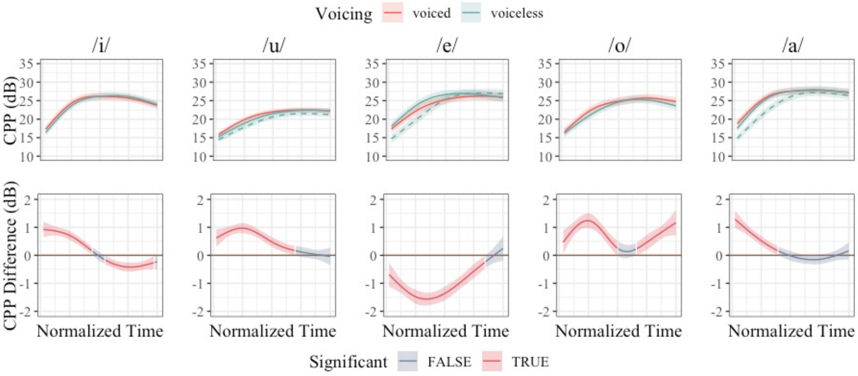


Figure 5: GAMM fitted CPP trajectories following voiced (red) and voiceless stops (green) (top) and the predicted mean difference across onset categories (bottom); dashed lines represent vowels following voiceless aspirated stops

5. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Professor Sam Tilsen, Professor James Kirby, Professor Abby Cohn, and Professor John Whitman for their valuable input and stimulating discussions. I am also thankful to the members of the Cornell Phonetics Lab for their generous support and helpful feedback throughout the research process.

6. REFERENCES

[1] Brunelle, M., Kirby, J. 2016. Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass* 10(4). 191-207.

[2] Brunelle, M., Tạ, T. 2021. Register in languages of Mainland Southeast Asia: the state of the art. In: Sidwell, P., Jenny, M. (eds.), *The Languages and Linguistics of Mainland Southeast Asia: A Comprehensive Guide*. De Gruyter Mouton, 683–706.

[3] Haudricourt, A. 1965. Les mutations consonantiques des occlusives initiales en môn-khmer. *Bulletin de la Société de Linguistique de Paris* 60(1). 160–172.

[4] Matisoff, J. 1973. Tonogenesis in Southeast Asia. In: Hyman, L. (ed.), *Consonant Types and Tone*. Linguistic Program, University of Southern California, 73–95.

[5] Diffloth, G. 1982. Registres, devoisement, timbres vocaliques: Leur histoire en Katouique. *Mon-Khmer Studies* 11. 47–82.

[6] Svantesson, J., House, D. 2006. Tone production, tone perception and Kammu tonogenesis. *Phonology* 23(2). 309–333.

[7] L.Thongkum, T. 1990. The interaction between pitch and phonation type in Mon: phonetic implications for a theory of tonogenesis. *Mon-Khmer Studies* 16–17. 11–24.

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

[9] Coetzee, A., Beddor, P., Shedden, W., Wissing, D. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *J. Phon.* 66. 185–216.

[10] Premsrirat, S. 2004. Register complex and tonogenesis in Khmu dialects. *Mon-Khmer Studies* 34. 1–17.

[11] Keating, P., Kuang, J., Garellek, M., Esposito, C. 2021. *A cross-language acoustic space for vocalic phonation distinctions*. https://linguistics.ucla.edu/people/keating/Keating-etal_ms_Nov2021.pdf

[12] Boersma, P., Weenink, D. 2020. Praat: doing phonetics by computer. http://www.praat.org/

[13] Yao, Y., Tilsen, S., Sprouse R., Johnson, K. 2010. Automated Measurement of Vowel Formants in the Buckeye Corpus. *UC Berkeley PhonLab Annual Report* 6.

[14] Brooke, M. 1997. Voicebox: Speech Processing Toolbox for MATLAB. http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

[15] Signal Processing Toolbox for MATLAB. https://www.mathworks.com/products/signal.html.

[16] Hillenbrand, J., Cleveland, R., Erickson, R. 1994. Acoustic correlates of breathy vocal quality. *J. Speech and Hearing Research* 37. 769–778.

[17] Wood, S. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society* 73(1). 3–36.

[18] Kirby, J., Ladd, D. 2015. Stop voicing and F0 perturbations: Evidence from French and Italian. *Proc. 18th ICPhS.*

[19] Kirby, J. 2018. Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *J. Phon.* 71. 326–354.

[20] Schertz, J., Khan, S. 2020. Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. *J. Phon.* 81. 100979.

[21] Esposito, A. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59. 197–231.

[22] Kingston, J., Diehl, R., Kirk C., Castleman, W. 2008. On the internal perceptual structure of distinctive features: The [voice] contrast. *J. Phon.* 36(1). 28–54.

[23] Bell-Berti, F. 1975. Control of pharyngeal cavity for English and voiceless stop. *J. Acoust. Soc. Am.* 57(2).456–461.

[24] Ahn, S. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *J. Phon.* 71. 451–467.

[25] Westbury, J. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *J. Acoust. Soc. Am.* 73(4). 1322–1336.

[26] Westbury, J., Keating, P. 1986. On the naturalness of stop consonant voicing. *J. Ling.* 22(1). 145–166.

[27] Löfqvist, A., Baer, T., McGarr, N. 1989. The cricothyroid muscle in voicing control. *J. Acoust. Soc. Am.* 85(3). 1314–1321.