The phonetic realization of tonal contrast in Dränjongke

LEE Seunghun J.[1], *HWANG, Hyun Kyung*[2], *MONOU, Tomoko*[3], *KAWAHARA, Shigeto*[4]

[1]International Christian University, Japan & University of Venda, South Africa
[2]RIKEN, Japan & International Christian University, Japan
[3]Mejiro University, Japan & International Christian University, Japan
[4]Keio University, Japan

seunghun@icu.ac.jp, hyunkyung.hwang@riken.jp, tomoko.monou@gmail.com, kawahara@icl.keio.ac.jp

Abstract

Dränjongke is a Tibeto-Burman language spoken in Sikkim, India. The language has been described as having a two-way tonal contrast [1], but how the tonal contrast is realized phonetically has not been explored in detail. To fill this gap, we report an acoustic analysis of Dränjongke based on the speech of 12 native speakers, focusing on how the tonal contrast in Dränjongke manifests itself in syllabary readings. The results suggest that (i) in syllables consisting of only vowels (e.g. [a]), the tonal differences manifest themselves at the left edge of syllables, and more surprisingly, (ii) in syllables with sonorant onsets (e.g. [na]), the tonal differences primarily appear during the sonorant intervals, and can be neutralized during the vocalic intervals.

This paper discusses some theoretical implications of the observed patterns. First, we propose that phonologically, tones can be directly associated with consonants, a proposal independently motivated by an analysis of phonological consonant-tone interaction patterns in various languages [2]. Second, we address the general issue of how supralaryngeal gestures and laryngeal gestures are coordinated, the issue known under the rubric of "articulatory binding" [3]. We argue that articulatory binding—a principle that requires laryngeal gestures to be coordinated with supralaryngeal gestures to maximize their perceptual salience—is violable.

Index Terms: Dränjongke, Tibeto-Burman languages, tonal contrast, phonetic alignment, articulatory binding

1. Introduction

Dränjongke (a.k.a. "Bhutia", "Hloke" or "Sikkimese") is a Tibeto-Burman language spoken in Sikkim, India. The language has been described as a language with a two-way tonal contrast (H(igh) vs. L(ow), [1]), but how the tonal contrasts are realized phonetically has not been explored in detail in previous literature—we report a part of our general research project which aims to fill this gap.

This paper has two specific aims. One aim is more or less descriptive—we report an acoustic experiment which explores the phonetic realization of the lexical tonal contrast in Dränjongke, whose phonetic properties have not been studied using instrumental techniques. The results of the experiment based on the production of 12 speakers' syllabary readings show that the tonal contrast manifests itself primarily at the left edge of syllables. A rather surprising result is that in CV-syllables in which onset consonants are sonorants (e.g. [na]), the tonal contrast primarily appears during the consonantal intervals. For some speakers, f0 differences are neutralized in vocalic intervals.

These observations lead to the second aim of the paper. Based on the results of the acoustic study, we propose that tones in Dränjongke are aligned to the left edge of the syllables, even when the leftmost segments are consonants. We explore the implication of the results in terms of (i) how tonal contrasts should be represented phonologically in Dränjongke, and (ii) how laryngeal and supralaryngeal gestures may be coordinated in the light of "articulatory binding", which requires that laryngeal gestures should be bound to supralaryngeal gestures in such a way that the consequences of the laryngeal gestures are most perceptible [3].

2. Method

The data discussed in the paper come from recordings of Dränjongke, based on the fieldwork conducted by the first author during the summer of 2017. The language reportedly has a two-way H vs. L tonal contrast [1]; tones are not contrastive in syllables whose onsets are obstruents (although different obstruent types affect f0 of the following vowel—we are unable to report these patterns due to space limitation).

2.1. Recording

Twelve speakers (two female) participated in the recording session. They were all school teachers from primary and secondary schools.

Dränjongke has eight vowels, described as "a", "ä", "e", "o", "ö", "u" and "ü" by [1]. Those syllables with a sonorant onset included [la], [ra], [ma], [na], [ɲa], and [ŋa]. Each speaker read the whole list of syllables, consisting of all eight vowels in isolation, and the combination of all consonants plus [a]. They repeated this list five times.

2.2. Acoustic analysis

The current analysis targeted the syllabary reading of vowel-only tokens and sonorant-vowel tokens. Most speakers produced H-toned [ra] as breathy; such tokens were excluded from the analysis, as their f0 contours were not measurable. Both sonorant and vowel intervals are annotated using Praat. Time-normalized tonal contours are extracted by calculating average f0 values within 10 equally-timed windows, both within the sonorant and vocalic intervals.

3. Results

Figure 1 shows the tonal contours of H-toned and L-toned tokens of vowel-only syllables. Although specific patterns are different for each speaker, especially in terms of differences in the shapes of syllable-final boundary tones, we generally observe clear separation between H-toned syllables (solid, red lines) and L-toned syllables (dotted, blue lines) at the beginning of syllables.[1] Speaker 11 is perhaps exceptional in that the difference between the two tones is generally small (< 10 Hz). Other speakers show differences as large as 20-30 Hz at the left edge of the syllables.

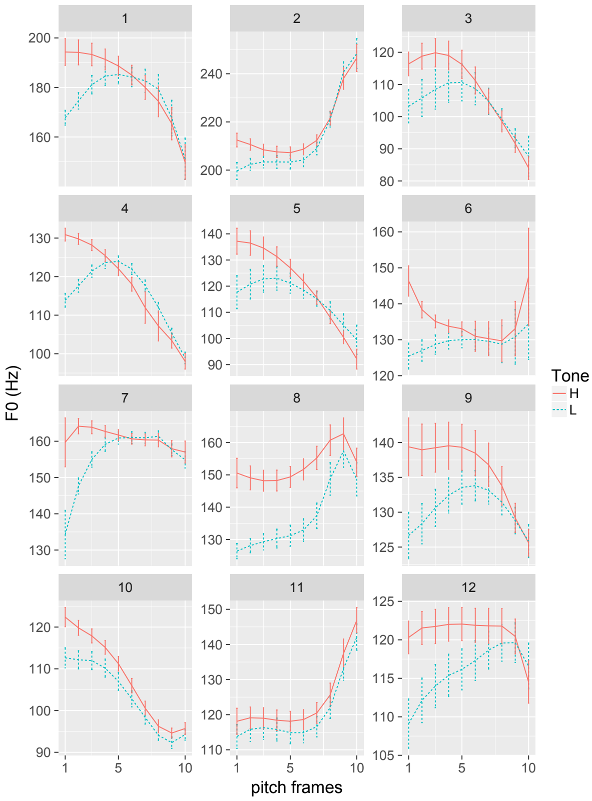

Figure 1: *The tonal contours of H- and L-tokens of vowel-only syllables for each speaker. H-toned syllables are shown with red solid lines and L-toned syllables are shown with blue dotted lines. Speakers 1 and 2 are female speakers. The y-axis scales are adjusted for each speaker. The error bars represent 95% confidence intervals.*

To assess the statistical significance of the difference between H-toned syllables and L-toned syllables, the average f0 values of the entire vocalic intervals were calculated. Figure 2 is a boxplot representation of averaged f0 values in the H vs. L tone conditions for each speaker. Since there is a substantial degree of interspeaker variation, within each speaker, a linear mixed model was run with the f0 values as the dependent variable, the tone type as a fixed independent variable and repetition as a random variable. The analysis was implemented using R, with `lme4` package [4]; the *p*-values were calculated using `lmerTest` package [5]. The results appear in Table 1, which shows that all speakers, including Speaker 11, robustly implemented the H-L tonal contrast in vowel-only tokens.

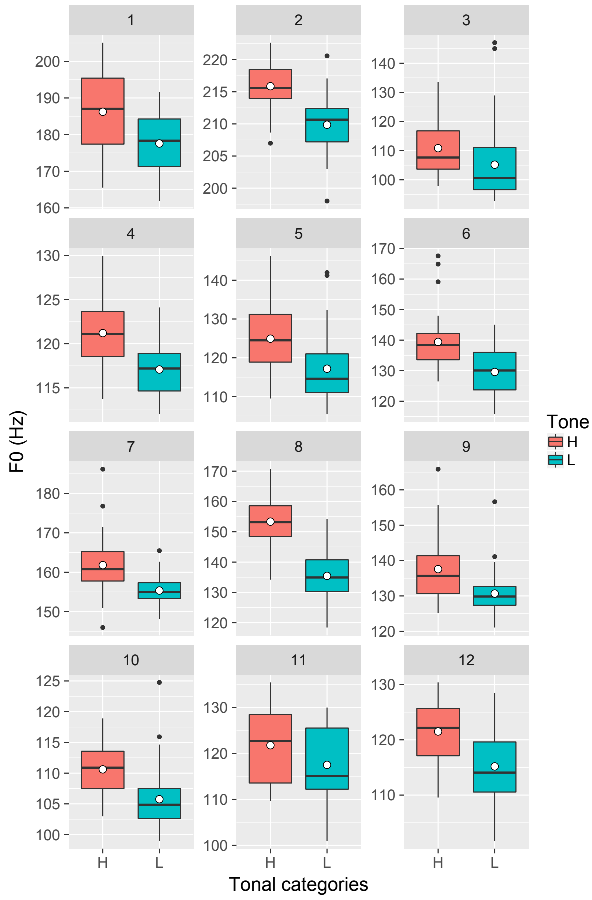

Figure 2: *The boxplot of the average f0 values in the H-toned and L-toned syllables. White dots represent averages. Vowel-only tokens.*

Table 1: *Results of linear mixed model analyses. Vowel-only tokens.*

Speaker	**t-value and p-value**
1	t = -4.371, p <.001
2	t = -5.803, p<.001
3	t = -2.213, p <.05
4	t = -5.469, p <.001
5	t = -4.623, p <.001
6	t = -6.555, p <.001
7	t = -5.108, p <.001
8	t = -10.60, p <.001
9	t = -4.108, p <.001
10	t = -5.090, p <.001
11	t = -4.747, p <.001
12	t = -5.183, p <.001

Figure 3 shows the normalized tonal contours of sonorant-vowel syllables. We observe that much of the difference between the H-toned and L-toned tokens appears during the consonantal syllables (the first 10 pitch frames). Indeed, for

[1] Since the language's tonal contrast is sometimes characterized as "a register contrast" [1], we also explored various measures of voice quality, such as H1-A1, H1-A2, NHR, but did not find a consistent difference that holds among these speakers.

Speakers 2, 3, 5, 6, 8, 9, 10, the f0 differences appear to be neutralized—not observed—during the vocalic intervals (the last 10 pitch frames).

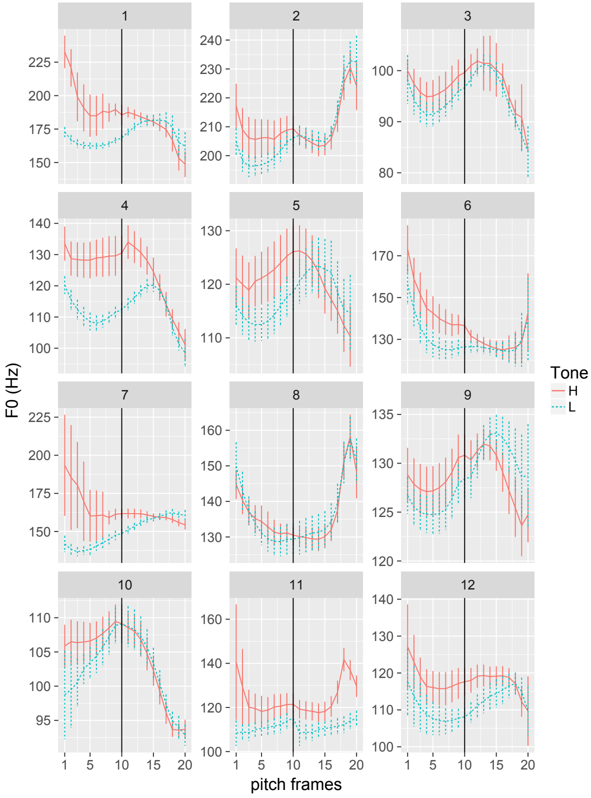

Figure 3: *The tonal contours of H- and L-tokens for each speaker for sonorant-vowel tokens. The first 10 frames correspond to sonorants; the last 10 frames correspond to vowels. The y-axis scales are adjusted for each speaker. The error bars are 95% confidence intervals.*

To assess the statistical significance of the separation between H-tones and L-tones, the average f0 values of both the sonorant intervals and vowel intervals were calculated. The results appear in Figure 4 and Figure 5, respectively. Again, within each speaker, a linear mixed model was run with the f0 values as the dependent variable, the tone type as a fixed independent variable and repetition as a random independent variable. Table 2 summarizes the results for the sonorant interval; Table 3 does so for the vocalic intervals.

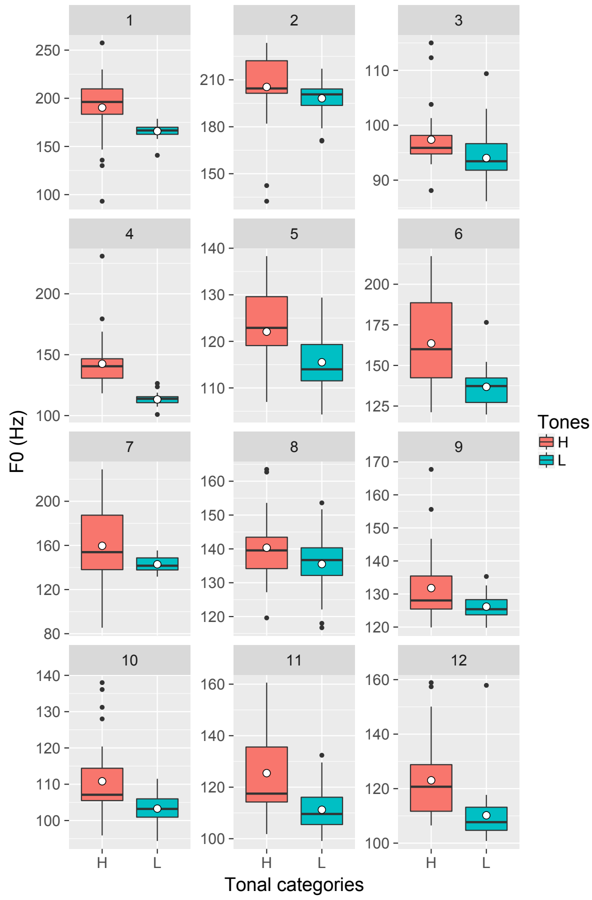

Figure 4: *The boxplot of f0 values in the H-toned and L-toned syllables. The sonorant interval in sonorant-vowel syllables.*

Table 2: *Results of linear mixed model analyses. The sonorant interval in sonorant-vowel syllables.*

Speaker	***t*-value and *p*-value**
1	$t = -4.032$, $p < .001$
2	$t = -1.583$, *n.s.*
3	$t = -2.380$, $p < .05$
4	$t = -6.727$, $p < .001$
5	$t = -3.128$, $p < .01$
6	$t = -4.951$, $p < .001$
7	$t = -2.611$, $p < .05$
8	$t = -2.128$, $p < .05$
9	$t = -2.74$, $p < .01$
10	$t = -3.567$, $p < .001$
11	$t = -4.338$, $p < .001$
12	$t = -3.954$, $p < .001$

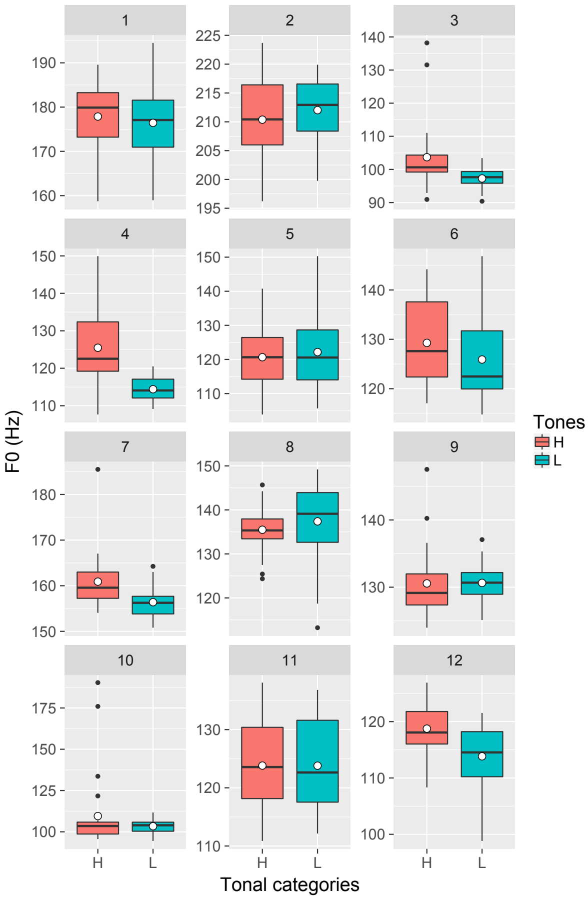

Figure 5: *The boxplot of f0 values in the H-toned and L-toned syllables. The vowel interval in sonorant-vowel syllables.*

Table 3: *Results of linear mixed model analyses. Vowel interval in sonorant-vowel syllables.*

Speaker	***t*-value and *p*-value**
1	t = -0.688, *n.s.*
2	t = 1.486, *n.s.*
3	t = -3.369, p <.01
4	t = -5.984, p <.001
5	t = 0.735, *n.s.*
6	t = -2.964, p<.01
7	t = -3.681, p <.001
8	t = 1.10, *n.s.*
9	t = 0.139, *n.s.*
10	t = -1.522, *n.s.*
11	t = 0.211, *n.s.*
12	t = -3.996, p <.001

During the sonorant interval, all the speakers but Speaker 2 showed a significant separation between the H-toned syllables and L-toned syllables (Table 2). Excluding two outlier points yielded a significant result for this speaker too (t = -3.931, p < .001). Only less than half of the speakers (Speakers 3, 4, 6, 7, 12) kept the significant separation in the vowel intervals; Speakers 1, 5, 8, 9, 10 do not show a significant difference between H-toned an L-toned syllables (Table 3).

4. Discussion and conclusions

Based on the tonal contours observed in Figure 1 and Figure 3, for the lexical tones in Dränjongke, it seems most natural to posit phonological representations in Figure 6. For the case of vowel-only syllables (Figure 6(a)), it is straightforward; the tone is associated with the vowel. For the case of sonorant-vowel syllables, for those speakers who showed f0 differences between H and L-toned only during the sonorant intervals (e.g. Speakers 1, 2 and 5), the tone is associated solely with a consonant, but not with a vowel, as in Figure 6(b). For those speakers who showed f0 differences in both the sonorant and vowel interval (e.g. Speakers 4 and 7), the phonological representation may be the one in Figure 6(c).

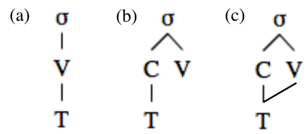

Figure 6: *Proposed phonological representations.*

We often assume that tones are associated with vowels (or moras or syllables) (see [6] for a cross-linguistic overview); the current data suggest that tones can be directly associated with consonants. This sort of representation seems rather radical; however, this kind of proposal is independently motivated by a cross-linguistic examination of the patterns of consonant-tone interaction, as argued by Lee (2008) [2] (cf. PENTA [7], [8]).

The results may also bear on a general issue of how supralaryngeal gestures and laryngeal gestures are coordinated, an issue that is discussed under the rubric of "articulatory binding". Kingston (1990) [3] argues that laryngeal gestures—such as voicing and aspiration—are often bound to consonant release, because the acoustic consequences of these laryngeal gestures are most audible in this position. However, the current results suggest that the requirement to align tones to the left edge of syllables can take precedence over the requirement to align tones to positions in which their acoustic consequences would be most audible (i.e. articulatory binding is violable) (see also [9], [10]).

We would like to close this paper by pointing out two tasks that should be addressed in future studies; (i) to examine whether the tonal alignment patterns observed in nonce syllabary reading here are observed in the production of real words, and (ii) to examine the tonal realization patterns in sentential contexts, especially how tonal contrasts interact with phrasal boundary tones. These issues, in addition to comparative phonetic work in Tibeto-Burman languages, are currently under active investigation.

5. Acknowledgements

This project is supported by Strategic Japanese-Swiss Science and Technology Programme of JSPS and SNSF. We are grateful for the Namgyal Institute of Tibetology, George van Driem and Donna Erickson for their help with this project.

References

[1] G. van Driem, *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language (2 vols.)*. Leiden: Brill, 2001.

[2] S. Lee, Consonant-tone interaction in Optimality Theory, Doctoral dissertation, Rutgers University, 2008.

[3] J. Kingston, Articulatory binding. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology*. Cambridge: Cambridge University Press, 1990. pp. 406-434.

[4] D. Bates, M. Maechler, B. Bolker, S. Walker, R.H.B. Christensen, H. Singmann, B. Dai., F. Scheipl, G. Grothendieck, and P. Green, lme4. R-package. https://cran.r-project.org/web/packages/lme4/lme4.pdf, 2017.

[5] A. Kuznetsova, P.B. Brockhoff, and R.H.B. Christensen, lmerTest. R-package. https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf, 2017.

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

[7] Y. Xu, A. Lee, S. Prom-on. & F. Liu. "Explaining the PENTA model: A reply to Arvaniti and Ladd," *Phonology*. vol. 32, pp 505-535, 2015.

[8] Y. Xu. "Understanding tone from the perspective of production and perception," *Language and Linguistics* vol.5, 757-797, 2004.

[9] D. Howe and. D. Pulleyblank, D. "Patterns and timing of glottalization," *Phonology*. vol. 18, pp. 45-80, 2001.

[10] J. Shaw and S. Kawahara "The lingual gesture of devoiced /u/ in Tokyo Japanese" *Journal of Phonetics*. vol. 66, pp. 100-118, 2018.