The acoustic correlates of the four-way laryngeal contrast in Drenjongke stops

Seunghun Lee[1,2], Shigeto Kawahara[3], Celéste Guillemot[1], Tomoko Monou[1,3]
International Christian University[1], University of Venda[2], Keio University[3]
seunghun@icu.ac.jp[1], kawahara@icl.keio.ac.jp[2], celeste.guillemot@gmail.com[3], tomoko.monou@gmail.com[4]

ABSTRACT

Drenjongke (Bhutia) is a Tibeto-Burman language spoken in Sikkim, India, whose phonetic properties are understudied. This language has been reported to have a four-way laryngeal contrast: aspirated, voiceless, voiced, and "devoiced" [7]. The current experiment examined how these four types of consonants are distinguished acoustically. An acoustic analysis of twelve Drenjongke speakers shows that in addition to differences in VOT, there are systematic differences in F0 and F1 in the following vowel: aspirated and voiceless consonants show higher F0 than voiced and devoiced consonants; aspirated and devoiced consonants show higher F1 than voiceless and voiced consonants. Our analysis further suggests that high F1 after devoiced consonants is controlled, rather than an automatic consequence of long VOT. We conclude that Drenjongke speakers use at least three acoustic dimensions—VOT, F0 and F1—to distinguish the four-way laryngeal contrast.

Keywords: Drenjongke, laryngeal contrast, VOT, F0, F1

1. INTRODUCTION

Drenjongke (a.k.a. "Bhutia", "Hloke" or "Sikkimese") is a Tibeto-Burman language spoken in Sikkim, India by about 80,000 speakers. Although there is an impressionistic description of this language by van Driem [7], not much is known about the phonetic nature of this language. This paper examines one aspect of this language: its four-way laryngeal contrast, which is cross-linguistically rare. According to [7], this language has aspirated, voiceless, voiced and "devoiced" obstruents. The first two categories are classified as "H-register" consonants, while the last two categories are classified as "L-register" consonants. Some minimal quadruplets are shown below in (1), where devoiced consonants are shown with an apostrophe:

(1) Minimal quadruplets

[pø] 'incense' vs. [p^{h}o] 'stomach' vs. [bu] 'middle' vs. [b'u] 'son'
[ki] 'peace' vs. [k^{h}i] 'dog' vs. [gæ] 'eight' vs. [g'æ] 'row'
[tøŋ] 'to show' vs. [t^{h}oŋ] 'to see' vs. [do] 'stone' vs. [d'o] 'touch'
[ʈoŋ] 'to kill' vs. [ʈ^{h}om] 'bazaar' vs. [ɖu] 'dragon' vs. [ɖ'u] 'boat'

Particularly intriguing is the last category, "devoiced," whose acoustic properties are not clear even from van Driem's description. The current experiment thus explored how the four types of laryngeal categories are distinguished acoustically.

2. METHOD

The data reported is based on the fieldwork in Sikkim, India, which was conducted in the summer of 2017.

2.1. Speakers

Twelve native speakers of Drenjongke participated in the recording session. Speakers 1 and 2 were female speakers, and the remaining speakers were male. They were all school teachers from primary and secondary schools. All the speakers spoke Nepali and English in addition to Drenjongke (there are, unfortunately, no monolingual Drenjongke speakers). The age ranged from 25 years old to 55 years old, most of them being between 36 and 45 years old. Consent forms and demographic questionnaires were collected from each speaker before the recording session. Each participant was compensated for their time (800 Indian Rupee).

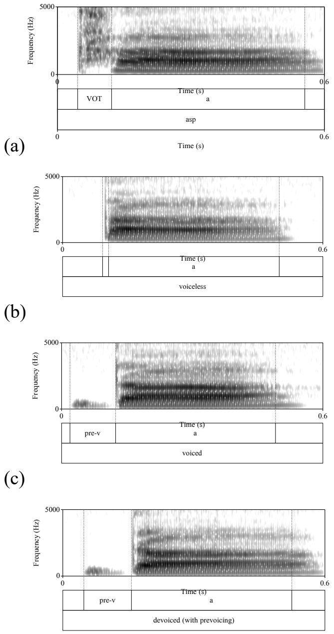
2.2. Recording

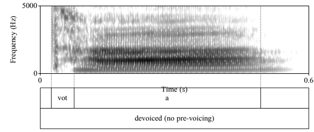
Within each recording session, each speaker read (1) typical syllables that appear in Drenjongke, (2) words in isolation, and (3) words in a frame sentence. This paper focuses on the analysis of syllabary readings, in which all Drenjongke consonants were pronounced with a following [a]. Here we focus on syllabary readings, as they control for lexical factors that may affect phonetic implementation patterns, and would tell us "pure" phonetic forms of Drenjongke. Having [a] also allows us to control for the intrinsic effects of vowel height on F0 and F1. The order of the syllabaries was randomized, and the speakers repeated the list five times. All the recording was made using a TASCAM recorder (DR100-MK). The stimuli were presented in the Tibetan script using Keynote on a Macintosh computer. The target of the current analysis included stop consonants from four places of articulation (bilabial, alveolar, retroflex, and velar), although the current analysis pools data from different place of articulation.

2.3. Acoustic analysis

Figure 1 provides some representative tokens of the four-types of laryngeal contrasts, which also serves to illustrate our measurement protocol. Aspirated consonants are realized with long lag VOT, (a); voiceless consonants are realized with short lag VOT, (b); voiced consonants are realized with prevoicing during closure (i.e. negative VOT), (c). Interestingly, devoiced consonants are variably realized with either prevoicing, (d), or positive VOT, (e). We measured the duration of these (negative and positive) VOT. In addition, since F0 and F1 are known to correlate with a laryngeal contrast (e.g. [6]), a 20 ms analysis window is created at the onset of the following [a]. Average F0 and F1 values were calculated within these analysis windows.

Figure 1: Representative alveolar tokens of the four-way laryngeal contrast. Devoiced consonants can be realized either with prevoicing or positive VOT.



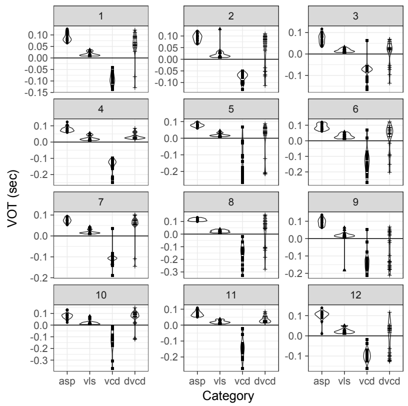


(e)

3. RESULTS

Figure 2 shows a violin-plot that shows the distribution of positive and negative VOT values for the four-way laryngeal contrast.

Figure 2: Distributions of negative and positive VOT by the four-way laryngeal contrast. Different panels show different speakers. The first two speakers are female.



Voiceless consonants have short-lag VOT. Aspirated consonants have long VOT. Voiced consonants usually have pre-voicing (i.e. negative VOT). There are a few exceptional tokens with positive VOT (e.g. Speakers 5 and 6), perhaps because these were read utterance-initially and it is hard to initiate voicing in this position [9]. Devoiced consonants are realized either with pre-voicing or positive VOT. Speakers 4 and 11 are the only speakers who do not show this variation. It seems to be the case that, given this variability, VOT alone cannot be used as a reliable acoustic cue for devoiced consonants in Drenjongke.

Figure 3 is a boxplot which shows the F0 values in the following [a] after the four different types of consonants. We observe that voiceless and aspirated consonants show higher F0 compared to voiced and devoiced consonants. This result accords well with van Driem's [7] characterization of these types of consonants: aspirated and voiceless consonants are “H-register” consonants and voiced and devoiced consonants are “L-register consonants”.

Figure 3: A boxplot representation of the F0 values in the following [a] vowel. The white circles represent the means in each condition.

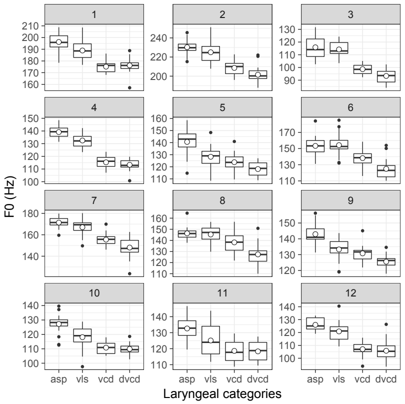
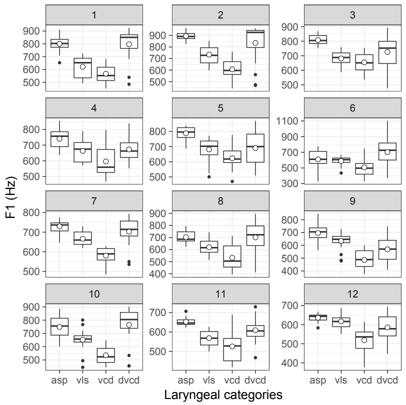


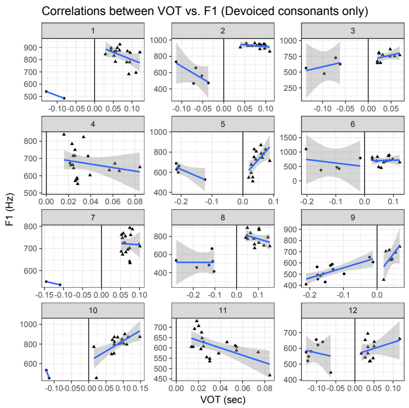
Figure 4 shows the F1 values of the following vowels. We observe that voiceless and voiced consonants show low F1, whereas aspirated and devoiced show high F1. Devoiced consonants, whose VOT profiles are rather variable (Figure 2), may instead be characterized as consonants with low F0 and high F1.

Figure 4: F1 values in the following [a] vowels.



One question that arises is whether high F1 values after devoiced consonants are intended (or controlled) or consequences of long VOT (which are exhibited by some tokens). Since F1 is correlated with the openness of the oral cavity [5], consonants with long VOT can show higher F1, because by the time F1 becomes measurable, the oral cavity is open more widely. To address this possibility, Figure 5 shows, for devoiced consonants, the correlation between F1 and VOT values, separately analysed by whether VOT values are negative or positive.

Figure 5: Correlation between F1 and VOT (devoiced consonants only).



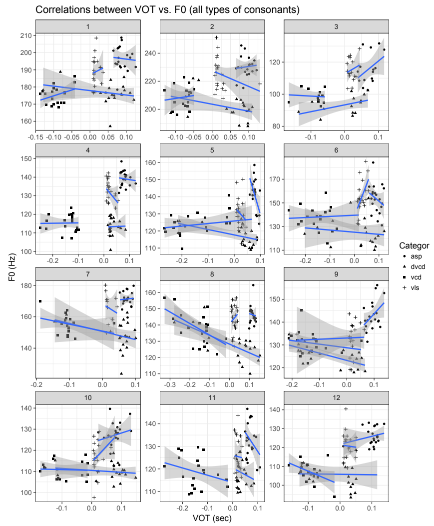
If high F1 is an automatic consequence of long VOT covering the opening phase of the oral cavity, there should be a positive correlation between F1 and VOT. However, this is true only for Speakers 5, 9 and 10. The rest of the speakers show either negative or no correlations, as summarized in Table 1. We thus suspect that high F1 of devoiced consonants is a consequence of an intended articulatory gesture; high F1 is particularly important to distinguish devoiced consonants from voiced consonants, both of which have low F0 in the following vowels.

Table 1: Pearson correlation coefficients between F1 values and VOT. Devoiced consonants with positive VOT values only.

Speakers	*r*	Speakers	*r*
1	-0.46	7	-0.03
2	-0.31	8	-0.34
3	0.01	9	0.64
4	-0.23	10	0.64
5	0.59	11	-0.58
6	0.05	12	0.32

One remaining question is how the high F1 values are achieved in this language. One possibility is jaw movement: the Drenjongke speakers may be opening their mouth more quickly after devoiced consonants, as English speakers do for voiceless consonants [8]. Alternatively, devoiced consonants may be accompanied by pharyngeal constriction: while constriction in the oral cavity generally lowers F1 [5], Al-Tamimi (2017) [1] shows that vowels in pharyngealized context show higher F1 in Jordianian and Moroccan Arabic.

Figure 5: Correlation between F0 and VOT (all categories).



Finally, our data allows us to address one theoretical question that is currently debated [2,3]: whether effects of onset consonants on the F0 of the following vowels (Figure 3) is based on phonetic categories or continuous VOT categories. This question is important as it bears on the question of how automatic/controlled F0 perturbation is. It could be the case that some aerodynamic and/or articulatory factors associated with aspiration/voicelessness can automatically result in higher F0 (see e.g. [4]). On the other hand, speakers may have distinct F0 target for different types of consonants ([6]). To address this question, Figure 6 plots the correlations between VOT and F0 of the following vowels: the regression lines are calculated within each laryngeal category. We observe that the correlations generally do not exist or are negative. The only clear positive correlations are observed for voiceless consonants for Speaker 6 and aspirated consonants for Speaker 9. We conclude from this data, a la [2,3], that it is the phonological category, rather than raw phonetic values, that determine the F0 values of the following vowel. It further implies that differences in F0 after different laryngeal categories are consequences of intended articulatory gestures rather than automatic consequence of laryngeal configurations.

4. CONCLUSION

The current paper set out to explore how the four-way laryngeal contrast in Drenjongke is acoustically realized. This was important because (1) there has not been instrumental studies examining how this contrast is acoustically realized, and (2) a four-way laryngeal contrast is cross-linguistically rare. Our finding is summarized in Table 2:

Table 2: A summary of how the four-way laryngeal contrast is distinguished acoustically in Drenjongke.

	aspirated	voiceless	voiced	devoiced
VOT	long	short	negative	variable
F0	high	high	low	low
F1	high	low	low	high

We admit that this may not be the complete picture of the laryngeal contrast in Drenjongke. Since our analysis is based on CV-tokens, we were unable to measure other acoustic correlates that are known to signal laryngeal contrasts, such as preceding vowel duration and consonant duration [6]. Our future work will aim to examine these acoustic correlates by analyzing VCV-tokens. With this limitation in mind, however, our data allows us to conclude that Drenjongke speakers use at least three acoustic dimensions—VOT, F0 and F1—to distinguish the four-way laryngeal contrast.

ACKNOWLEDGEMENTS

This work is supported by Strategic Japanese-Swiss Science and Technology Programme of JSPS and SNSF. A significantly expanded version of this paper is appearing in the *Journal of the Phonetic Society of Japan*.

5. REFERENCES

[1] Al-Tamimi, J. 2017. Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology*, 8(1):1-40.

[2] Clayards, M. 2018. Individual talker and token variability in multiple cues to stop voicing. *Phonetica*, 75(1):1-28.

[3] Dmitrieva, O., Llanos, F., Shultz, A. A., Francis, A. L. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, 49: 77-95.

[4] Hombert, J-M., Ohala, J., Ewan, W.G. 1979. Phonetic explanations for the development of tones. *Language* 55, 37–58.

[5] Johnson, K. 2003. *Acoustic and Auditory Phonetics: 2nd Edition*. Blackwell, Malden and Oxford.

[6] Kingston, J., Diehl, R. 1994. Phonetic knowledge. *Language*, 70:419-454.

[7] van Driem, G. 2016. The phonology of Dränjoke: Experimental development of Roman Dränjoke and Phonological Dränjoke. Manuscript. University of Bern.

[8] van Summers.W. 1987. Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, 82:847-863.

[9] Westbury, J. R., Keating, P. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics,* 22:145-166.