Acoustic analysis of onset voicing in Dzongkha obstruents

James Kirby[1], Gwendolyn Hyslop[2]
[1]University of Edinburgh, [2]University of Sydney
j.kirby@ed.ac.uk, gwendolyn.hyslop@sydney.edu.au

ABSTRACT

We present an acoustic analysis of cues to onset voicing in Dzongkha, the national language of Bhutan. Dzongkha is typically described as having a four-way laryngeal contrast between aspirated, unaspirated, prevoiced and devoiced obstruents. Previous descriptions suggest that this system may be changing, with the devoiced series either merging with the voiced series, or losing closure voicing but retaining contrastive pitch and/or voice quality. Based on data from 12 speakers, we find voiced and devoiced plosives are realised both with and without voicing lead. Tokens realized as phonetically voiced can be redundantly breathy; however, a low register tone always occurs on syllables headed by both voiced and devoiced obstruents, regardless of presence or absence of voicing lead. We discuss the implications of these findings for models of tonogenesis and historical sound change in the Tibeto-Burman context.

Keywords: *Voicing, tone, Tibeto-Burman*

1. INTRODUCTION

Dzongkha (dzo) is a Tibetic language of Bhutan, spoken by about 175,000 people as a native language. As the national language, it is also spoken by a majority of the over 700,000 inhabitants of Bhutan as a second language.

Dzongkha is usually described as contrasting voiceless aspirated, voiceless unaspirated, prevoiced, and *de*voiced obstruents [6, 11, 16, 15]. For example, consider the following minimal set:

(1) ta རྟ་ 'horse' (3) da མད་ 'arrow'
(2) tʰa ཐ་ 'letter *tha*' (4) d̥a ད་ 'now'

The fricative series, while lacking aspirates, also includes a devoiced member. The "devoiced" series is so-called by Tibetanists as it derives from a historically voiced obstruent series, as attested in Written Tibetan [5] and reflected in the Dzongkha orthography (see above). In many Tibetic languages, its reflex is now voiceless unaspirated. See [6, 16] for concise summaries of Dzongkha phonology.

There are some indications that the devoiced series may be disappearing in Dzongkha. Michailovsky and Mazaudon [11] state that non-native speakers frequently confuse the voiced and devoiced series, producing both as voiced. Conversely, Watters [15, 16] reports that the devoiced series remains distinct from both the voiced and voiceless series, but with acoustic realizations which are "a hybrid of the voiceless aspirated and unaspirated series" (2018:30), characterized by breathy voice quality and low pitch. To the best of our knowledge, these studies contain the only phonetic analyses of Dzongkha to date.

Along with its typologically unusual laryngeal contrast, Dzongkha has been analyzed as having a tonal register contrast that is linked to the voicing of the onsets: voiceless and aspirated obstruent onsets are associated with the high register while voiced and devoiced obstruent onsets are associated with the low register [6]. This tonal register contrast has its source in the historical laryngeal contrast. Complex onsets for which the second member was a sonorant have simplified to the sonorant only but with high tone, showing a contrast with previous simple sonorant onsets, now concomitant with a low tone. Among obstruents, however, there is strong correlation between onset type and f0 distribution, suggesting that tone is still incipient in this series [16].

In this paper, we investigate the acoustic cues to Dzongkha obstruents in onset position of monosyllabic words. In particular, we are interested in how obstruents differ in terms of VOT, f0, and measures of voice quality. Based on previous descriptions, we expect to find either a merger of the /voiced/ and /devoiced/ stops [11] or maintenance of the /devoiced/ series through at least some acoustic dimensions [15, 16].

2. METHODS

2.1. Participants

12 native speakers of Dzongkha (6 female) were recorded in Sydney, NSW. All participants were also fluent in English and spoke varying degrees of Nepali, Hindi, and/or Tshangla, in addition to other Bhutanese languages. Most speakers had at least one parent who was not a native speaker of Dzongkha. All except three were born in the Bhutanese capital of Thimphu and had lived in Australia for no more than 3 years. The one speaker born outside Thimphu was

from Paro, a region in western Bhutan; he was also the oldest speaker in our sample (age uncertain, but in his early to mid 40s). The remaining speakers aged from 15 to 27 (mean: 21).

2.1. Materials

Participants were recorded producing a list of 126 lexical items in isolation, with orthographic onsets /b̥ b p pʰ d̥ d t tʰ ɖ̥ ɖ ʈ ʈʰ g̊ g k kʰ dʒ̊ dʒ tʃ tʃʰ z̬ z s ʒ̊ ʒ ʃ dz ts tsʰ l̬ l m n/. There were 3-5 unique items for each onset, with varying vowel qualities; low vowels in open/sonorant-final monosyllables were selected wherever possible. For the sonorants /l m n/, the wordlist included an equal number of items from both the high and low tonal registers. Recordings were made in a quiet room at a private residence in Sydney with a Beyerdynamic 55.18 Mk II microphone connected to a Marantz PMD-661 digital recorder.

2.2. Data processing and analysis

Recordings were annotated in Praat [1] for onset and release of closure, duration of rime, and onset of voicing. Obstruents were also coded for PHONETIC VOICING, indicating whether a de/voiced segment was realized with any amount of voicing lead.

Spectral measures were extracted using PraatSauce [9]. f0, F1-F3, and harmonic amplitudes (H1, H2, A1-A3) were measured at every 1 ms in the target rime. Harmonic amplitudes were corrected for the influence of vocal tract resonances (formant frequencies and bandwidths) following [8].

Due to space restrictions, here we focus on the plosives /b̥ b p pʰ d̥ d t tʰ ɖ̥ ɖ ʈ ʈʰ g̊ g k kʰ/, fricatives /z̬ z s ʒ̊ ʒ ʃ/ and sonorants /l̬ l m n/. For brevity, we will use /D/, /D̥/, /T/, /Tʰ/ to refer to phonologically /voiced/, /devoiced/, /voiceless/, and /aspirated/ plosives, and [D], [T], and [Tʰ] to refer to their phonetic realizations in terms of voicing lead/lag.

2.3. Results

2.3.1. Phonetic realization of devoiced segments

VOT distributions for the four plosive places of articulation are shown in Figure 1, with means and standard deviations in Table 1. Both /D/ and /D̥/ are realized with and without closure voicing. The distribution varies with speaker: some speakers primarily realize both /D/ and /D̥/ as [D], others primarily as [T], and some a mixture. Overall, 35% of /devoiced/ stops were [voiced] (range: 7-93%) as were 60% of the /voiced/ stops (range: 0-82%). The percentage of [voiced] realizations in each context is strongly correlated (r=0.7), indicating that most of the speakers preferentially produce both /D̥/ and /T/ either as [D] or as [T].

Figure 1: VOT for plosives by voicing and place of articulation, over all speakers and items.

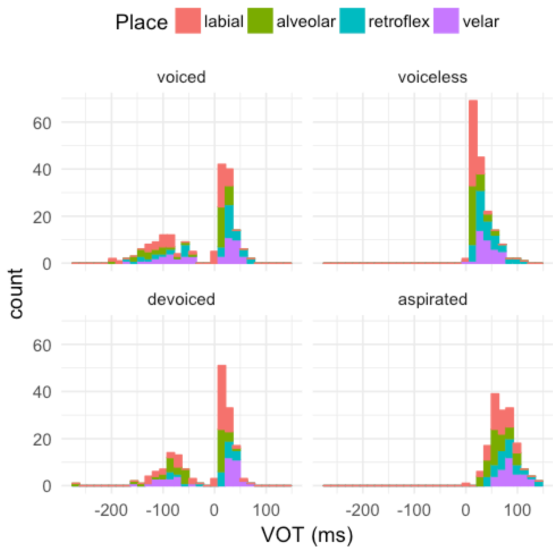


Table 1: Means & standard deviations of VOT (in ms) for Dzongkha plosives, averaged over speakers and place of articulation.

	Mean	SD
/D/ → [D]	-102	38
/D/ → [T]	25	15
/D̥/ → [D]	-89	37
/D̥/ → [T]	25	13
/T/	30	20
/Tʰ/	72	23

A mixed model was fit to the data for non-aspirated plosives with fixed factors VOICING, PLACE, their interaction, PHONETIC VOICING, and the interaction of VOICING and PHONETIC VOICING, along with random intercepts for speaker and item, speaker-specific slopes for VOICING and PHONETIC VOICING, and item-specific slopes for PHONETIC VOICING. Pairwise comparisons of the estimated marginal means showed no significant differences in the phonetic realizations: voicing lead times of plosives realized as [D] were indistinguishable regardless of whether they were orthographic /D/ or /D̥/ onsets; when realized as [T], both were indistinguishable from /T/ onsets (and from one another).

To estimate the effects of place of articulation on VOT, we also fit a model with fixed terms PLACE and PHONETIC VOICING, random slopes for speaker and item, and speaker-specific slopes for PHONETIC VOICING. No differences were found in voicing lead times for [D] realizations, but the expected effects of PLACE emerged for [T] and [Tʰ], with retroflex and velar having longer VOTs (~40 ms voiceless, ~80 ms aspirated) than bilabial and alveolar plosives (~20 ms voiceless, ~60 ms aspirated), cf. [4].

Fricatives in Dzongkha can also be voiced (/Z/) or devoiced (/Z̥/). Like plosives, the presence of voicing during the constriction was variable for nearly all speakers in our sample. However, as seen in Table 2, the frication durations for orthographically /voiceless/ fricatives are rather longer than those of /voiced/ or /devoiced/.

Table 2: Means & standard deviations of frication duration (in ms) for Dzongkha fricatives, averaged over speakers and place of articulation.

	Mean	SD
/Z/ → [Z]	158	40
/Z/ → [S]	165	41
/Z̥/ → [Z]	146	39
/Z̥/ → [S]	147	32
/S/ → [S]	185	41

2.3.3. *f0*

Figure 2 shows the realization of f0 (in semitones) over the first half of the rime by phonological voicing specification (here, plosives and fricatives are combined). Note that for sonorants, "voiceless" and "voiced" refer to high and low tonal register, respectively, and that /voiceless/ and /aspirated/ segments are plotted in the same quadrant.

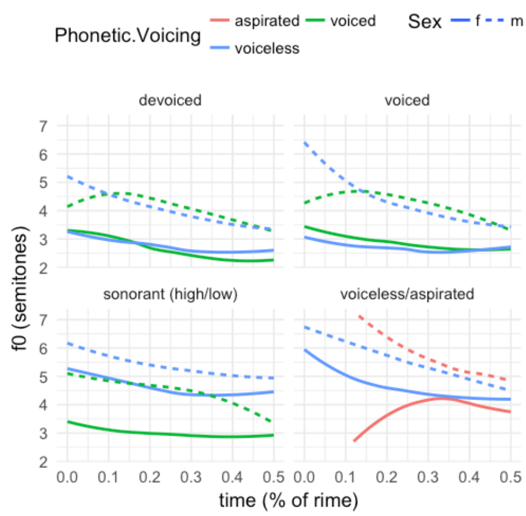
For females, we see the expected effect of tonal register: /voiced/ and /devoiced/ obstruents are followed by lower f0 (< 4 ST), /voiceless/ by high f0 (> 4 ST). However, there do not appear to be any differences between /devoiced/ and /voiced/ stops as a function of their phonetic realization. For /voiceless/ stops, there appears to be an onset f0 difference conditioned by aspiration, with [aspirated] stops showing lower onset f0, similar to some Chinese languages [3, 17]; further inspection (not shown here) indicates this effect is driven by just a few of the female speakers.

For males, the effect of tonal register on f0 in sonorants is visually much smaller. Both /voiced/ and /devoiced/ obstruents are followed by similar f0 trajectories, with [voiceless] realizations showing slight onset f0 perturbation effects. For males, f0 following both /voiceless/ and /aspirated/ segments is largely similar.

To estimate the size of the f0 differences, we fit a mixed model to mean f0 over the first 25% of the vowel with fixed factors SEX, VOICING, PHONETIC VOICING, and all two- and three-way interactions. We also included random intercepts for speaker and word, and speaker-specific slopes for VOICING and PHONETIC VOICING; this was the maximal model which converged. Pairwise comparisons of the marginal means estimate the difference between high and low register sonorants to be around 2.2 ST for females and 1.4 ST for males. Within orthographic/phonological voicing category, no other pairwise differences were significant.

Comparing the effects of *phonological* voicing across *phonetic* realizations, /D̥/ → [T] onsets were significantly lower than /T/ → [T] onsets by around 2 ST for females and 1.5 ST for males; differences of similar magnitude are found between /D/ → [T] and /T/ → [T]. There were no significant differences in the realization of pitch between the /D̥/ → [T] and /D̥/ → [D] onsets for either group. This shows that the phonetic realization of the onset does not generally affect f0 of the following vowel, consistent with tonal register being lexically specified.

Figure 2: f0 (in semitones) averaged over speakers, places, and manners (plosive/fricative) by phonological (orthographic) voicing specification.



2.3.3. *Voice quality*

We also examined our data for variation in voice quality, as measured by differences in spectral slope [7, 16]. If /devoiced/ obstruents involve a lax vocal fold setting, we should see higher spectral slopes at the onsets of vowels following /devoiced/ as opposed to /voiced/ or /voiceless/ obstruents (or sonorants).

Figures 3 and 4 show the time course of H1*-H2* and H1*-A1* over the first half of the syllable rime by voicing category, phonetic realization, and speaker sex. For H1*-H2*, the expected differences are apparent (females having greater spectral slopes than males). For H1*-A1*, onset breathiness appears to be a function of *phonetic*, rather than phonological voicing: the onset of vowels following both /D̥/ →

[D] and /D/ → [D] are breathier than those following /D̥/ → [T] or /D/ → [T], or /T/ → [T]. Mixed models similar to those for f0 (not reported here) corroborate the visual assessment. Results for F1 (not shown here) follow a similar pattern (F1 higher after [D] and [T^h]).

Figure 3: H1*-H2* (in dB) averaged over speakers, places, and manners (plosive/fricative) by phonological (orthographic) voicing specification.

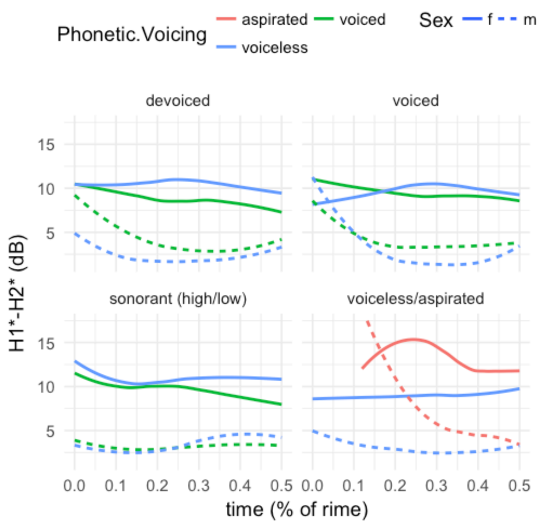
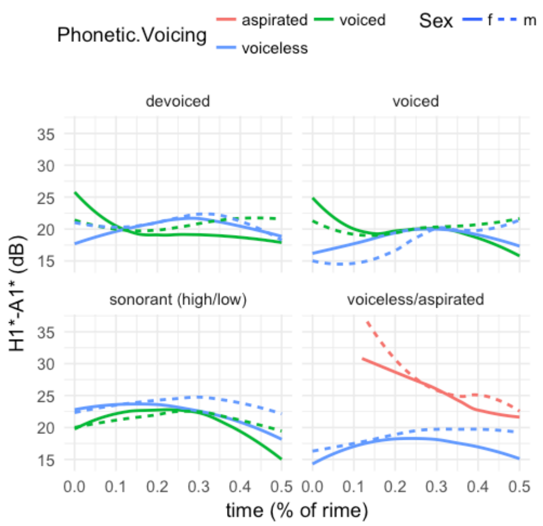


Figure 4: H1*-A1* (in dB) averaged over speakers, places, and manners (plosive/fricative) by phonological (orthographic) voicing specification.



4. DISCUSSION

Unlike the speakers studied by Watters [16], VOT does not distinguish /devoiced/ and /voiceless/ plosives for our speaker sample and corpus. For speakers or tokens where /voiced/ or /devoiced/ are realized as [voiceless], the VOT distributions do not differ significantly. Similarly, when /devoiced/ plosives are realized as [voiced], they are not distinct from /voiced/ realized as [voiced]. Furthermore, both /devoiced/ and /voiced/ plosives are associated with low f0 on the following vowel, regardless of whether or not closure voicing is realized. This indicates that, despite the variability in (or confusion regarding) how the onsets themselves should be pronounced, speakers are clear as to which tonal register these lexical items belong to.

Watters argues that breathiness and pitch are part of a bundle of features that distinguish /devoiced/ from /voiceless/ obstruents in Dzongkha. In our data, if breathiness is indicative of anything, it is of *phonetic* voicing: H1*-A1* (though not always H1*-H2*) is greater following /voiced/ and /devoiced/ stops when they are realized as [D], but not when as [T] (cf. [2, 12]). This suggests that breathiness and pitch are largely independent, pitch being primarily a function of the tonal register associated with a (monosyllabic) word, with voice quality either a phonetic consequence of the way speakers implement closure voicing, and/or an enhancement to improve the perceptibility of voicing when it is realized.

While the present data are intriguing, they should be interpreted with caution. In addition to Dzongkha and English, nearly all of our participants also spoke some Tshangla, Nepali, and/or Hindi, in addition to other local languages. This is a common situation in Bhutan, but nevertheless raises questions about code-switching and transfer effects. We also need to test with a wider range of vowel qualities; the voice quality results in particular could be the result of imbalances in our sample in terms of vowel qualities (although in theory, the use of corrected spectral balance measures should control for this).

In Dzongkha as elsewhere in Tibeto-Burman, it is often supposed that the process of tonogenesis in the obstruent series began with a loss of onset voicing, resulting in a compensatory lowering of pitch via breathy phonation [10, 13, 14, 16]. As nearly all the speakers in our sample are quite young, they may be somewhat further along in this process: having apparently associated low f0 with voicing, voice quality may no longer be perceptually influential, and simply appears as a side effect of phonetic voicing of obstruents. The significant confusion about whether /voiced/ and /devoiced/ obstruents involve vocal fold vibration indicates that these categories may be in the process of merging, but the ultimate direction the merger will take is not yet clear.

5. ACKNOWLEDGMENTS

Funded in part by The University of Sydney. Thanks to Michelle Yuan, Matthew Sung, and Leah McPherson for assisting with the data preparation and annotation. We are also grateful to Gyambo Sithay for his assistance in Sydney.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2015. *Praat: doing phonetics by computer*.

[2] Brunelle, M., Tạ Thành Tấn, Kirby, J., Đinh Lư Giang. 2019. Obstruent devoicing and registrogenesis in Chru. *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne.

[3] Chen, Y. 2011. How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics* 39(4), 612–625.

[4] Cho, T., Ladefoged, P. 1999. Variation and universals in VOT: evidence from 17 endangered languages. *Journal of Phonetics* 27, 207–229.

[5] Delancy, S. 2017. Classical Tibetan. In: LaPolla, R., Thurgood, G. (Eds.), *The Sino-Tibetan Languages*. London: Routledge, 369–384.

[6] Driem, G. van. 1998. *Dzongkha*. Leiden: CNWS.

[7] Gordon, M., Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29, 383–406.

[8] Iseli, M., Shue, Y.-L., Alwan, A. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America* 121(4), 2283–2295.

[9] Kirby, J. 2018. *praatsauce: Praat-based tools for spectral analysis*. Retrieved from https://github.com/kirbyj/praatsauce

[10] Mazaudon, M. 2012. Paths to tone in the Tamang branch of Tibeto-Birman (Nepal). In: De Vogelaer, G., Seiler, G. (Eds.), *The dialect laboratory: dialects as a testing ground for theories of language change*. Amsterdam: John Benjamins, 139–178.

[11] Michailovsky, B., Mazaudon, M. 1988. Lost syllables and tone contour in Dzongkha (Bhutan). In: *Prosodic analysis and Asian linguistics: to honour R. K. Sprigg*. Pacific Linguistics, 115–136.

[12] Pittayaporn, P., Kirby, J. 2017. Laryngeal contrasts in the Tai dialect of Cao Bằng. *Journal of the International Phonetic Association* 47(1), 65–85.

[13] Suzuki, H. 2015. New perspectives on the suprasegmentals in mBrugchu Tibetan: an introduction to the tonogenesis triggered by breathy voice. *Bulletin of Chinese Linguistics* 117–139.

[14] Thurgood, G. 2002. Vietnamese tone: revising the model and the analysis. *Diachronica* 19(2), 333–363.

[15] Watters, S. A. 2002. Sounds and tones of five Tibetan languages of the Himalayan region. *Linguistics of the Tibeto-Burman Area* 25(1), 1–65.

[16] Watters, S. A. 2018. *A grammar of spoken Dzongkha (dzo): phonology, words, and simple clauses* (PhD dissertation). Rice University, Houston, TX.

[17] Xu, C. X., Xu, Y. 2003. Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association* 33(2), 165–181.