TONAL AND LARYNGEAL CONTRASTS IN DIASPORA TIBETAN

Christopher Geissler

Yale University
christopher.geissler@yale.edu

ABSTRACT

The relationship between voice onset time (VOT) and F0 has been studied from the perspective of the effect of stop voicing contrasts on pitch and the interrelationship of phonation type and tone, but less has been reported about the effect of tone on VOT. This paper reports on an acoustic study of Diaspora Tibetan, which contrasts high and low tones as well as word-initial aspirated and unaspirated stops.

Results indicate that VOT is longer for high-tone words than low-tone words, but only for aspirated, not unaspirated stops. VOT difference may enhance the tone contrast, but does not appear to exist in a trading relation with pitch. Additionally, variable prevoicing, produced by all speakers, was only observed in unaspirated stops with low tone, not with high tone. These results suggest that phonological tone conditions VOT in Diaspora Tibetan.

Keywords: Tone, Aspiration, Tibetan, Consonant-Tone Interaction

1. INTRODUCTION

1.1. VOT and F0

Voice onset time (VOT) is a well-established acoustic parameter most closely associated with voicing and aspiration contrasts in stops. It also varies according to other parameters, such as place of articulation [9] and surrounding segmental context [8]. While languages commonly contrast long, short, and/or negative VOT, the precise values of these parameters also vary by language [3].

The relationship of VOT and pitch (F0) has been studied from a number of perspectives, including typological and experimental work on vocal fold tension, an articulation involved in producing both longer VOT and in higher F0; conversely, voicing is associated with lowered F0 [6,10]. Raising the larynx has also been linked both to voiceless stops and to raised pitch [15].

Past research has thus focused on the effect of common articulation on both F0 and voicing/ aspiration, or on the effect of laryngeal phonology and phonetics on F0. This paper, however, presents evidence of, in a sense, the opposite, where VOT is conditioned by tone. As shown in section 3.2, aspirated stops have longer VOT with high tone than with low tone, an unusual result that may serve as a secondary cue for tone. If degree of VOT difference were such a secondary cue, it might be expected to exist in a trading relation [14], though this hypothesis is not supported.

1.2. Language Background

Central Tibetan is typically described as contrasting two tones: high and low, where the high tone is level while the low tone exhibits a rise in pitch. The highest pitch achieved in both tones, while occurring at different points, is described by Duanmu [5] and Tournadre and Dorje [16] as reaching the same level. This suggests that the same tone gesture could be responsible for both tones, by being timed differently with respect to the oral gestures.

As in other Tibeto-Burman languages, tones and word-initial consonants have a close diachronic relationship in Tibetan, with the high and low tones arising from a reanalysis of historic voiceless and voiced consonants [7,11]. However, the language today is described as having voiceless aspirated and unaspirated stops co-occurring with both tones, and variable voicing of unaspirated stops with low tone [3,16].

Tibetan VOT and tone contrasts are most prominent at the beginnings of words. Stop aspiration and voicing contrasts are neutralized in non-initial positions, and the tone of a polysyllabic word is determined by the tone of its first syllable—the high-level or rise takes place over the duration of the word irrespective of the number of syllables in the word. Thus, this paper focuses on the measurement of tone in the first syllable of disyllabic words, and the VOT of word-initial stops.

2. METHODS

2.1. Speakers

Data was collected from 19 native speakers of Tibetan (8 women, ages 20-38, mean 24.2) living in the Kathmandu Valley, Nepal, recorded in 2016 as part of a larger study. Eleven were born and raised in Nepal, while the remaining eight were born in Central-Tibetan-speaking regions of the Tibet Autonomous Region but arrived in Nepal before the age of eleven.

2.2. Data Collection

Recordings were made on a Zoom H4N recorder at a 48kHz sampling rate with an Audio-Technica ATM73a headset microphone.

The 64 items used in this study were presented as a wordlist in the Tibetan orthography, for which speakers were asked to read each item twice. The items include a range of monosyllabic and disyllabic words with both high and low tones, word-initial onset consonants with a range of places of articulation and specification of aspiration, as well as diverse vowels.

However, while all possible combinations of both tones with word-initial onsets by aspiration category are represented, the numbers of these items are not balanced. 6 items were low-tone and unaspirated, 7 items were high-tone and aspirated, 8 items were low-tone and aspirated, but only one item was high-tone and unaspirated.

F0 and VOT measurements were made in Praat [2]. F0 measurements were taken at the midpoint and at ten time-normalized intervals across the first vowel of each of the 43 disyllabic words (16 high-tone, 27 low-tone). Disyllabic words were chosen because it is there that the high-level and low-rising contrast is most clearly realized. VOT was measured for the 22 stop-initial words from the beginning of the release burst to the onset of voicing.

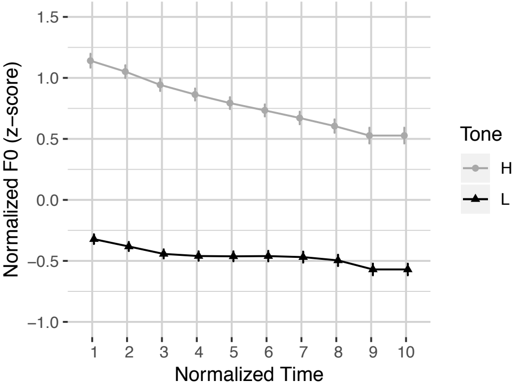
3. RESULTS

3.1. Status of pitch contrast

The z-score was calculated by speaker for each time interval across the first vowel of disyllabic words. Fig. 1 illustrates the difference in z-scored F0 tracks for all speakers, demonstrating the tonal contrast on first-syllable vowels.

In light of anecdotal concern that some speakers may exhibit a tone merger, we sought to determine the presence of a tone contrast for each speaker. The *lme4* [1] package in R [13] was used to fit a linear mixed-effects model to the normalized F0 data for each speaker. A baseline model included fixed effects of normalized time and a random effect of lexical item. This baseline model was compared with a second model featuring an interaction between tone and time. P-values were obtained by likelihood ratio tests of the second model against the baseline model, and all speakers showed that adding a fixed effect of tone significantly improved the model (p < .001).

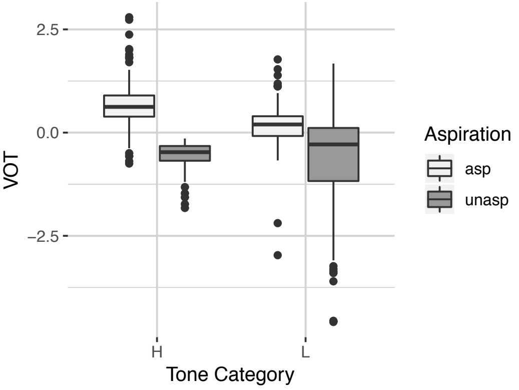
Figure 1: Time-Normalized z-score pitch trajectory in first-syllable vowel, across all speakers.



3.2. VOT contrast

Fig. 2 illustrates differences in word-initial VOT, z-scored by speaker, according to aspiration and tone categories. VOT was longer for aspirated stops with high tone (mean 72.7 ms, sd 22.6) than with low tone (mean 48.1 ms, sd 26.2), with less difference between unaspirated stops with high tone (mean 15.2 ms, sd 5.1) than with low tone (mean 9.4 ms, sd 64). With the exception of two outliers in the low-tone aspirated category, which we interpret as speech errors, all tokens produced with negative VOT were unaspirated and low-toned, representing 23% (n = 65) of unaspirated low-tone tokens. This is in line with previous descriptions of this category as voiced or variably voiced [3,16]. Prevoicing was highly variable: all speakers produced some tokens in this category with prevoicing, and no speakers produced all tokens in this category with prevoicing.

Figure 2: Word-initial voice onset time (VOT) z-scored by speaker, across categories of tone and aspiration.



To test the significance of the observed differences in VOT by tone, we fit linear mixed-effects models to the speaker-normalized VOT data. A baseline model included random effects of lexical item and speaker and a two-level fixed effect for prevoicing (with a reference level of positive VOT). This was compared to a second model with a four-level fixed effect of tone and aspiration, representing the four combinations shown in Fig. 2; the low-tone aspirated category was chosen as the reference level because of its apparent intermediate value and central position in the contrast system. The model comparison using ANOVA is summarized in Table 1. Results indicate that the addition of tone-aspiration category significantly improved fit over the baseline model.

Table 1: Comparison between models predicting VOT: a baseline model and one with a four-level effect of tone/aspiration category

Model	Df	AIC	logLik	*p* value
Baseline	5	973.7	-481.9	
+ tone/asp	8	957.4	-470.7	>.001*

The coefficients of this model are presented in Table 2. As compared to the VOT of the low-tone aspirated category (represented by the intercept, which is not significantly different from zero), this table shows that the low-tone unaspirated category is not significantly different, the high-tone unaspirated category has lower VOT, and the high-tone aspirated category has higher VOT. The effect of negative-VOT tokens as compared to positive VOT tokens is also included, identifying the subset of low-tone unaspirated tokens with negative VOT.

Table 2: Fixed effects of the better-fitting model in Table 1. Low-tone aspirated is the reference level, with the next three rows representing other combinations of tone and aspiration. “Negative VOT” is a different factor included to identify the effect of prevoicing.

Fixed effect	Estimate	Std. error	*p* value
(Intercept)	0.193	0.146	0.1999
L,unaspirated	-0.315	0.182	0.1010
H,unaspirated	-0.813	0.348	.0315*
H,aspirated	0.472	0.181	.0181*
Negative VOT	-2.271	0.064	>.001*

This conditioning of VOT in (only) aspirated stops by tone is an unusual result, and lacks a clear physiological explanation or motivation. It has long been known that multiple cues can contribute to a contrast, often in a trading relation [14].

In order to test for a possible trading relation between VOT and F0, we looked for a correlation between the size of VOT contrast among high-tone words with the size of the F0 contrast. For each speaker, we measured the difference in speaker-normalized VOT in high- and low-tone aspirated stops. We also measured the difference between speaker-normalized F0 at the midpoint of first-syllable vowels with both tones; the midpoint was chosen because the correlation of interest was with tone target, not local coarticulatory effects. However, a Pearson Product-Moment Correlation found no significant correlation ($r(17) = -0.32$, $p = .175$) between the magnitude of F0 and VOT differences.

4. DISCUSSION

The present study has confirmed the robust nature of the tonal contrast Diaspora speakers of Central Tibetan, and investigated the relationship between tone, voicing, and aspiration.

As discussed in section 3.1, tonal contrast serves to distinguish the high- and low-toned words, so the overlapping VOT ranges discussed in section 3.2 do not obscure this contrast. However, the lack of VOT difference between aspirated and unaspirated stops with low tone—an apparent (partial) merger—indicates that the variable prevoicing on low-tone unaspirated words may play an important role in maintaining this contrast.

In addition to prevoicing, tone also conditions VOT among aspirated stops: high-tone aspirated stops have longer VOT than their low-tone counterparts. The direction of this association should not come as a surprise, given the natural association between aspiration and raised pitch through vocal fold tension, for example [6]. Additionally, while F0 (at first-vowel midpoint) and VOT do not appear to exist in a trading relation, the condition could serve to enhance contrast. While the aspirated/unaspirated contrast among low tones is enhanced by variable prevoicing, this option is not available for high-tone unaspirated stops. Therefore, to maximize the contrast between high-tone stops, as well as between aspirated stops, the lengthening of high-tone aspirates is a reasonable strategy.

However, the contrast enhancement between the aspirated stops of different tones comes at the expense of another contrast, that between low-tone aspirated and unaspirated stops. The VOT values of these stops are effectively merged, since the low-tone unaspirated stops exhibiting prevoicing constitute a minority of tokens even in the very careful speech style elicitation context used in this study. That the prevoicing has remained confined to the low-tone unaspirated category indicates that the aspirated-unaspirated contrast remains active for speakers despite the frequent ambiguous VOT durations. In light of this, why would speakers not produce low-tone aspirated stops with longer VOT, in order to distinguish them from the majority of low-tone unaspirated tokens that lack prevoicing?

These results parallel those of McCrea and Morris [12], who found an effect of F0 on VOT for English voiceless (aspirated) stops but not "voiced" (unaspirated) stops. However, they found shorter, not longer, VOT with higher pitch, and of course English does not have lexical tone as Tibetan does. But they speculate that stops with long VOT may be more permissive of phonetic effects that condition VOT given their wider temporal range; speakers may control short-VOT stops, with their smaller ranges, to minimize such effects. They further speculate that the physiology of short-lag stops may reach a maximum VOT value which, if exceeded, would lead to the different glottal state used that results in aspiration noise.

Of these two explanations, the former—tighter speaker control of short-VOT stops—is less plausible for Tibetan given the wide range, both above and below the median, of the VOT values in the low-tone unaspirated category. With such a variation in VOT, adequate range should exist for speakers to allow variation caused by any phonetic effects of tone. That these are not observed in unaspirated stops suggests this explanation is not adequate for the Tibetan data.

Alternatively, there could be a significant difference in the phonetics of aspirated vs. unaspirated stops beyond VOT alone. Whatever factor is present for aspirated stops could thus allow for greater interaction with F0 than is possible with unaspirated stops. However, investigating this would require more extensive research into the articulation these stops in terms of vocal fold stiffness, larynx raising, and other factors.

5. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1-48.

[2] Boersma, P., 2002. Praat, a system for doing phonetics by computer. *Glot international* 5.

[3] Chang, K., Shefts, B., 1964. *A Manual of Spoken Tibetan (Lhasa Dialect)*.

[4] Cho, T. and Ladefoged, P., 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27, 207-229.

[5] Duanmu, S., 1992. An autosegmental analysis of tone in four Tibetan languages. *Linguistics of the Tibeto-Burman area* 15, 1-27.

[6] Halle, M., 1971. A note on laryngeal features, QPR Research Laboratory of Electronics. MIT 101, 198-213.

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

[8] Klatt, D.H., 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Am.* 59, 1208-1221.

[9] Lisker, L. and Abramson, A.S., 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.

[10] Löfqvist, A., Baer, T., McGarr, N.S. and Story, R.S., 1989. The cricothyroid muscle in voicing control. *J. Acoust. Soc. Am.* 85, 1314-1321.

[11] Matisoff, J.A., 1999. Tibeto-Burman tonology in an areal context. In P*roceedings of the symposium "Crosslinguistic studies of tonal phenomena: Tonogenesis, Japanese Accentology, and Other Topics.* 3-31. Tokyo: Tokyo University of Foreign Studies

[12] McCrea, C.R. and Morris, R.J., 2005. The effects of fundamental frequency level on voice onset time in normal adult male speakers. *Journal of Speech, Language, and Hearing Research* 48, 1013-1024.

[13] R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL www. Rproject.org

[14] Repp, B.H., 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological bulletin* 92, 81.

[15] Stevens, K.N., 1977. Physics of laryngeal behavior and larynx modes. *Phonetica* 34, 264-279.

[16] Tournadre, N. and Rdo-rje (Gsaṇ-bdag.), 2003. *Manual of standard Tibetan: language and civilization*. Snow Lion Publications.