Eccentric C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts

Christopher Geissler[1], Jason Shaw, Mark Tiede[2], Fang Hu[3]

Languages differ in how articulatory gestures are coordinated in time, including the relative timing between consonants and vowels, i.e., C-V timing. Previous work has observed that in lexical tone languages, such as Mandarin (Gao 2008), Thai (Karlin 2014), and Lhasa Tibetan (Hu 2016), the lag between consonant and vowel gestures (C-V lag) is longer than in non-tonal languages, such as English (Lofqvist and Gracco 1999). Additionally, C-V lag is longer for tonal syllables than non-tonal syllables in Mandarin (Zhang et al. 2019). In the present study, we focus on diaspora Tibetan, which furnishes a unique example of a language community where some speakers contrast tones and others do not, allowing comparison across speakers who do and do not contrast tone in the same language. We predicted that C-V lag would be longer for speakers that maintain the lexical tone contrast than for those that do not. The prediction follows from the empirical observations cited above as well as the model of tone as an articulatory gesture (Gao 2008; Niemann et al., 2011) organized with other gestures as competitive coupled oscillators (Browman & Goldstein 2000, Nam & Saltzman 2003).

Six speakers (four female) of Tibetan raised in India and Nepal and living in the United States participated in this study. Using Electromagnetic Articulography (EMA), we recorded the movement trajectories of the lips and tongue dorsum while speakers produced target words in a carrier phrase. Target items (N = 71) were one and two syllable words that varied in word-initial consonant (/m/, /p/, /p^h/) and tone (high-level and low-rising); first-syllable vowels drew from the set of back vowels in Tibetan: /u/, /o/, /a/. Speakers produced each item 4-10 times. Gestures for lip aperture and tongue dorsum retraction were identified as starting where 20% of peak velocity toward target was reached.

In order to identify which speakers produced a tone contrast, time-normalized F0 trajectories were analyzed for systematic differences by lexical tone category, using data from monosyllabic nasal-initial tokens (60 per speaker). Generalized Additive Mixed Models (*mgcv*, Wood 2011; in R:R Core Team 2011) were fit to the data of each speaker, predicting F0 based on terms for lexical tone category, a smooth for normalized timestep, a difference smooth by lexical tone category, and random smooths by word. The difference smooth for two participants (one female, one male) was not significant, indicating they did not produce a lexical tone contrast. The difference smooth for the other four participants was significant (Fig. 1).

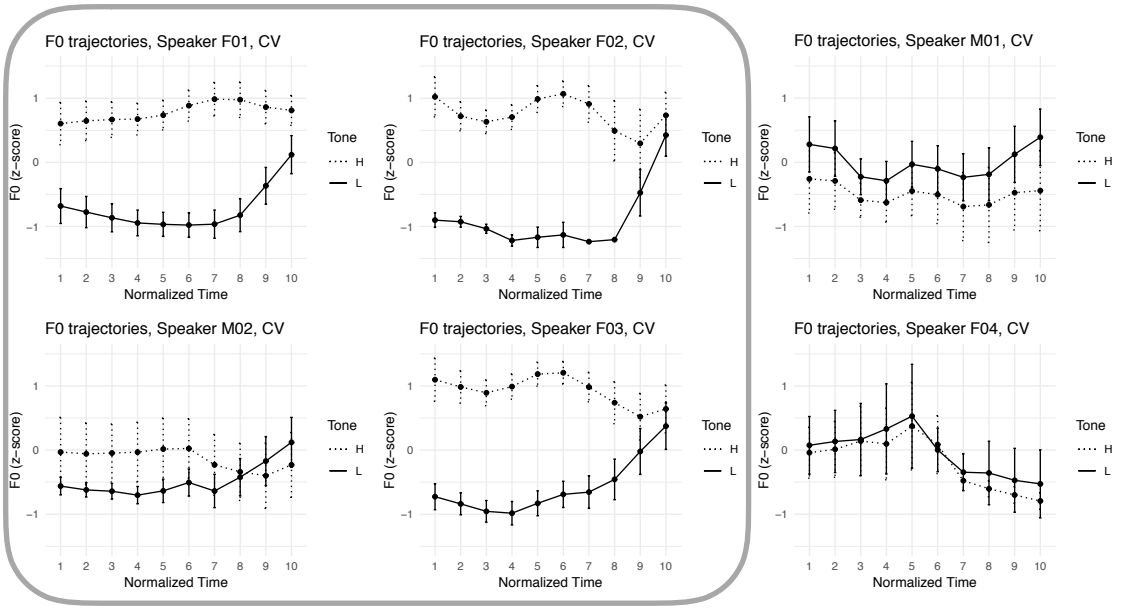
However, contrary to the predictions of the model of tone as an articulatory gesture, tonal and non-tonal speakers did not differ in C-V lag (Fig. 2). We used both C-V lag, defined as the difference between the gestural onset of the consonant gesture and the gestural onset of the vowel gesture, as well as a measure we call "C-V phasing," the ratio of the C-V lag divided by the duration of the C gesture (defined as gestural onset to attainment of target). In both cases, model comparison with linear mixed-effects models (*lme4*, Bates et al 2015) indicated that the effect of tone on C-V lag was no significant; C-V lag also did not differ across tones. Both groups of speakers did show a positive C-V lag, however, which is consistent with the

---

[1]Yale University; [2]Haskins Laboratories; [3]Chinese Academy of Social Sciences

predictions for tonal speakers of the competitive gestural coupling account; in other words, both tonal and non-tonal speakers exhibited C-V lag like that predicted for tonal speakers.

We interpret the lack of difference in C-V lag across tonal and non-tonal speakers as evidence for eccentric C-V timing by non-tonal speakers. While the model of tone gestures and competitive coupling does explain the C-V lag observed in tonal speakers, no tone gesture is present for the non-tonal speakers. Instead, it appears that all the Tibetan speakers in this study learned a pattern of C-V timing common to their community: tonal speakers may have used a tone gesture with competitive coupling, but the non-tonal speakers would have to learn these as eccentric coupling relations (Marin & Pouplier 2010). Such eccentric coupling is predicted to be possible (e.g. Goldstein 2011), but less likely than in-phase and anti-phase coupling, and the present study demonstrates an instructive example of eccentric timing that mimics competitive coupling.



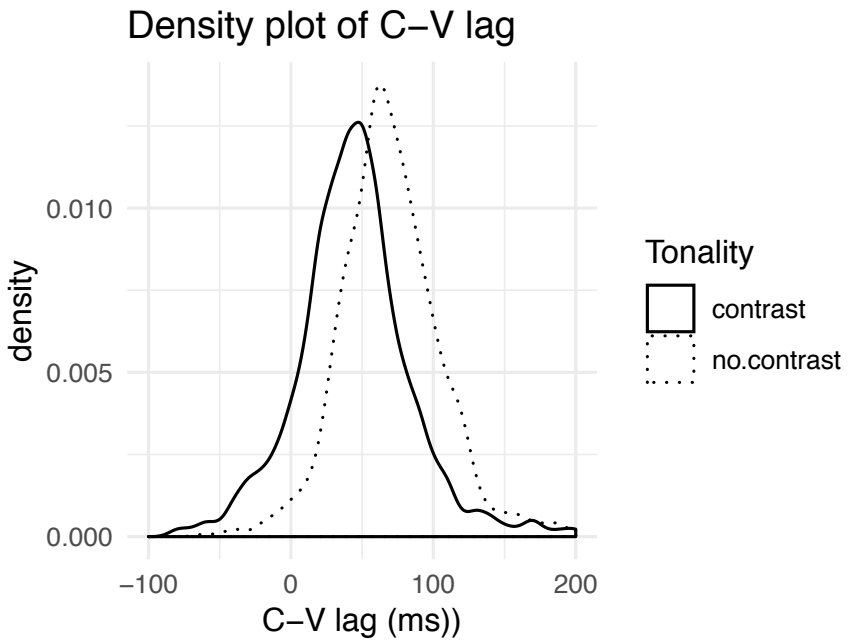


Fig. 1 (above): F0 trajectories for each speaker. The four speakers at left (circled) were found to contrast tone. Z-scored time-normalized F0 is plotted, with error bars at 95% confidence intervals.

Fig. 2 (left): Density plot of C-V lag (ms). The four tone-contrasting speakers and two non-contrasting speakers are indicated, with positive C-V lag values for both groups.