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

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

[1]Yale University, USA
[2]Haskins Laboratories, USA
[3]Chinese Academy of Social Sciences, China

christopher.geissler@yale.edu, jason.shaw@yale.edu, tiede@haskins.yale.edu, hufang@cass.org.cn

Abstract

This paper investigates the relative timing of onset consonant and vowel gestures in Tibetan as spoken in the Tibetan diaspora. According to the coupled oscillator model of articulatory timing (Browman & Goldstein 2000, Nam & Saltzman 2003), the most readily-available coupling modes among gestures are in-phase (synchronous) or anti-phase (sequential) timing, with competition among these modes also giving rise to a stable timing pattern. The model predicts that other timing relations, i.e. "eccentric timing", are possible but not as readily available. Data gathered using electromagnetic articulography (EMA) shows relative C-V timing consistent with either competitive coupling or eccentric timing. Competitive coupling is a plausible explanation for CV syllables in a tone language (Gao 2008), but acoustic analysis showed that some speakers do not produce a pitch contrast corresponding to tone. In the apparent absence of a tone gesture, we conclude that these speakers exhibit eccentric C-V timing.

Keywords: speech prosody, articulation, tone, Tibetan, Articulatory Phonology, Electromagnetic Articulography

1. Introduction

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 (Löfqvist & Gracco 1999). Additionally, C-V lag is longer for tonal syllables than non-tonal syllables in Mandarin (Zhang et al. 2019). These observations support treating tone as an articulatory gesture (Gao 2008; Niemann et al. 2011) which can be organized with other gestures as competitively-coupled oscillators (Browman & Goldstein 2000, Nam & Saltzman 2003).

In the present study, we focus on Tibetan as spoken in the diaspora enclaves in South Asia and around the world, outside of traditional Tibetan-speaking regions. Speakers raised in these communities are exposed to a mixed linguistic input comprising speakers of tonal and non-tonal Tibetan dialects, along with extensive contact with other languages. Of these diaspora-raised speakers, some produce a tone contrast while others do not. This fact makes diaspora Tibetan speakers a unique case in which to test the effect of tone on articulatory timing.

1.1. Coupling relations

Pairwise coupling relations are a prominent way to account for the timing of gestures in Articulatory Phonology. By hypothesis, any relative phasing of gestures is possible, but two coupling modes are considered to be intrinsically stable: *in-phase* (synchronous) and *anti-phase* (sequential) (Saltzman & Byrd 2000). Language-specific timing patterns can arise from the interaction of these two coupling modes, as in the case of *competitive coupling.* Competitive coupling has been invoked to explain the partial overlap in onset consonant clusters (Nam & Saltzman 2003, Goldstein et al. 2000) as well as the difference in C-V timing between languages with and without lexical tone (Gao 2008). A schematic depiction of coupling graphs and their associated gestural scores is shown in **Figure 1**. The first two diagrams depict **(a)** in-phase coupling and **(b)** anti-phase coupling using common examples of a CV syllable and a CC sequence, respectively. In **Figure 1(c)**, competitive coupling is illustrated with anti-phase coupling of consonant and tone gestures and in-phase coupling of both with a vowel gesture.

While in-phase, anti-phase, and competitive coupling should be most readily available to speakers, cases of other timing relations are possible as well. Other timing relations are called *eccentric timing* (Goldstein 2011) and represent any timing values besides those already mentioned. Importantly, eccentric timing allows a pairwise timing between two gestures similar to that found in cases of competitive coupling, as illustrated in **Figure 1(d)**.

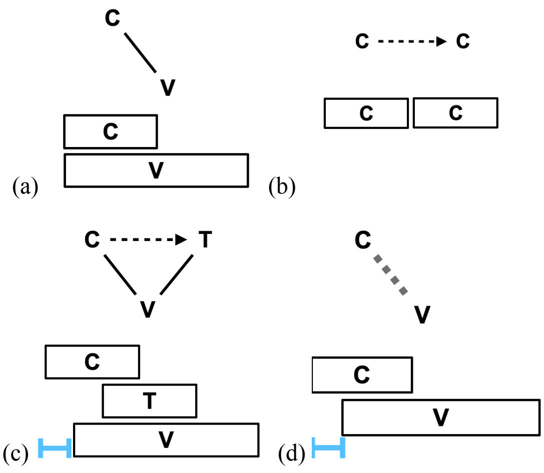

Figure 1: *(a) In-phase and (b) anti-phase coupling of articulatory gestures. (c) Competitive coupling of consonant, vowel, and tone gestures can result in similar C-V timing as (d) eccentric timing.*

1.2. Predictions

As a language with lexical tone, Tibetan may be expected to show similar C-V lag to other tone languages. Following Gao (2008), a tone gesture with competitive coupling would predict a C-V lag value of ~50 ms. However, Tibetan as spoken in diaspora presents a high degree of diversity, with some speakers lacking a tone contrast. Speakers who do not produce a tone contrast would be expected to exhibit in-phase C-V timing, and therefore ~0 ms C-V lag. Alternatively, with eccentric C-V timing, the C-V lag of non-tonal speakers could take on other values, including resembling that of tonal speakers.

2. Methods

Six native speakers (four female, two male) of Tibetan participated in this study. All were raised in Tibetan diaspora communities in India and Nepal and were living in the United States.

Target items (N = 71) were one- and two- syllable words that varied in word-initial consonant (/m/, /p/, /pʰ/) 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. Target words were presented using Tibetan orthography in the carrier phrase /t͡sʰík t̪ĭ ____ t̪ŭk/ 'This word is ____.' This phrase provided a front vowel preceding the target item, allowing the back vowel of the target item to be identified with tongue dorsum retraction.

Recordings were made using the NDI Wave EMA system using sensors aligned with each speaker's midsagittal plane. Consonant gestures were identified using lip aperture, the Euclidean distance between sensors on the upper and lower lips, and vowel gestures using movement in the longitudinal (anterior-posterior) dimension of a sensor placed on the tongue dorsum. Articulatory gestures were identified using *Mview* (Tiede 2005): the start time of gestures was identified as the point at which 20% of the maximum velocity toward the target was attained, and the attainment of gestural target was identified as the point at which velocity had reduced to 20% of the maximum velocity toward target. C-V lag was calculated as the difference in start time between consonant and vowel gestures.

In order to identify which speakers produced a tone contrast, time-normalized F0 trajectories were analyzed for systematic differences by lexical tone category. Data for this analysis came from monosyllabic nasal-initial tokens (60 per speaker). The target syllables were identified as beginning with the acoustic start of the nasal consonant and ending with the end of a visible second formant at the onset of the following stop closure. F0 was measured at ten evenly-spaced intervals using a script (DiCanio 2012) in Praat (Boersma & Weenink 2018)

3. Results

3.1. Tone contrast status

Four of the six speakers were found to produce a significant difference in F0 trajectories on nasal-initial syllables, which we interpret as evidence for a tone contrast in production. 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. Raw pitch trajectories are presented in **Figure 2**.

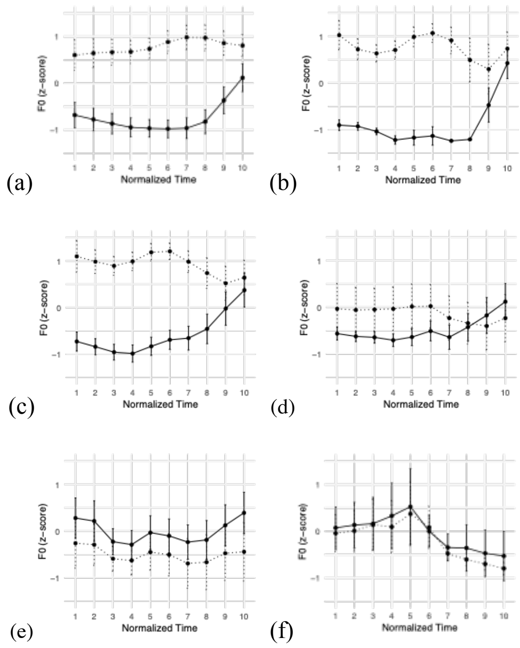

Figure 2: *Time-normalized F0 trajectories on /mV/ syllables. F0 values are z-scored by speaker. Model comparison with GAMMs found significant differences by tone in (a)-(d) but not for (e)-(f)*

3.2. C-V lag

Overall C-V lag values were similar to those observed in other languages with lexical tone, centering around 50 ms, and these values did not differ by tone. However, C-V lag values were similar for speakers with and without tone contrasts, as established in 3.1. C-V lag values according to tone contrast status are presented in **Figure 3**.

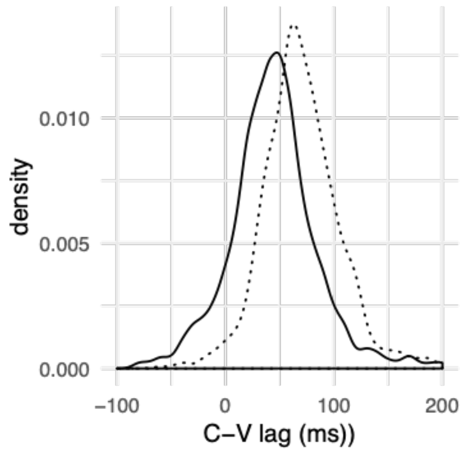

Figure 3: *C-V lag among speakers with tone contrast (solid line) and without tone contrast (dotted line).*

Furthermore, we found that C-V lag is positively correlated with the duration of the onset consonant gesture. This held true for both speakers with and without the tone production contrast, as shown in **Figure 4**. This pattern is not expected if the gestures are in-phase (Shaw, Durvasula, Kochetov, 2019).

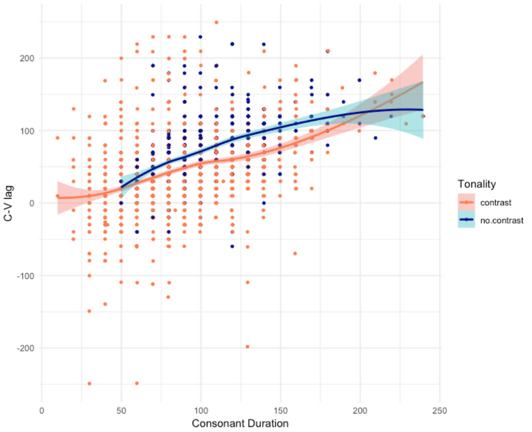

Figure 4: *Positive correlation of consonant gesture duration on C-V lag.*

These observations—the lack of difference across speakers by tonality, and the correlation of C duration on C-V lag—were substantiated using linear mixed-effects models (lme4; Bates et al. 2015). We performed a linear mixed-effects model analysis of the relationship between consonant gesture duration and C-V lag. A baseline model included fixed effect of onset consonant [m p pʰ] and random effects of speaker and lexical item. This was compared to a second model that also included a fixed effect of consonant duration. Comparison was also made to a third model that included a fixed effect of tonality: whether or not a speaker produced a tone contrast. As shown in **Table 1**, the inclusion of consonant duration led to an improvement over the baseline model, but the further inclusion of tonality did not further improve the model.

model	AIC	BIC	loglik	Chi-sq	p-value
baseline	31776	31810	-15882		
baseline + C dur	31764	31804	-15875	13.9403	0.0002
baseline + C dur + tonality	31764	31810	-15874	1.7267	0.1888

Table 1: *Comparison of linear mixed-effects models shows improvement with consonant duration, but no improvement with tone contrast status of speakers.*

4. Discussion and conclusion

Observed C-V lag values were around 50 ms and were positively correlated with consonant gesture duration, consistent with both competitive coupling and eccentric timing. However, speakers lacking a tone contrast exhibited similar C-V lag to speakers with a contrast, which we interpret as evidence that non-tonal speakers, at least, use eccentric timing. This is because these speakers lack a tone gesture, and so do not have another gesture with which to make competitive coupling possible. We further rule out the possibility of C-V timing being anti-phase because the C-V lag values are substantially less than consonant duration (even defined as just the closing phase of the consonant gesture).

We thus conclude that while tone-contrasting Tibetan speakers may use competitive coupling, the non-tonal speakers in our sample used eccentric timing. All the speakers in this study learned a pattern of C-V timing common to their community, and so some speakers must have accomplished this using eccentric timing (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 example of eccentric timing that mimics competitive coupling.

5. Acknowledgements

We would like to thank the speakers for participating in this study, Yale University and the National Science Foundation Graduate Research Fellowships Program for funding support, and Muye Zhang for assistance with EMA recording. We are also grateful for feedback from audiences at Yale Linguistics, LSA 2020, and ISSP 2020.

6. References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer software]. Version 6.0. 43.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de La Communication Parlée*, *5*, 25–34.

DiCanio, C. (2012). *Pitch Dynamics Script for Praat, version 6*. https://www.acsu.buffalo.edu/~cdicanio/scripts/Pitch_-Dynamics_6.praat

Gao, M. (2008). Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Unpublished Doctoral Dissertation (Linguistics), Yale University, CT*.

Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In Proceedings of the XVIth international congress of phonetic sciences (pp. 241-244).

Goldstein, L. (2011). *Back to the Past Tense in English* (R. Gutiérrez-Bravo, L. Mikkelsen, & E. Potsdam, Eds.; pp. 69–88). Linguistics Research Center, UC-Santa Cruz Department of Linguistics.

Hu, F. (2016). Tones are not abstract autosegmentals. *Speech Prosody*, 302–306.

Karlin, R. (2014). The articulatory TBU: gestural coordination of tone in Thai. *Cornell Working Papers in Linguistics*.

Löfqvist, A., & Gracco, V. L. (1999). Interarticulator programming in VCV sequences: Lip and tongue movements. *The Journal of the Acoustical Society of America*, *105*(3), 1864–1876.

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, *14*(3), 380–407.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of Syllable Structure: A Coupled Oscillator Model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to Phonological Complexity*. Walter de Gruyter. https://doi.org/10.1515/9783110223958

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. *Proceedings of the 15th International Congress of the Phonetic Sciences*.

Niemann, H., Mücke, D., Nam, H., Goldstein, L., & Grice, M. (2011). Tones as Gestures: The Case of Italian and German. *Proceedings of the 17th International Congress of the Phonetic Sciences*, 1486–1489.

R Core Team. (2013). R: A language and environment for statistical computing.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*(4), 499–526.

Shaw, J. A., Durvasula, K., & Kochetov, A. 2019. The temporal basis of complex segments. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019 (pp. 676-680). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Tiede, Mark. (2005). Mview: Software for visualization and analysis of concurrently recorded movement data.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 3–36.

Zhang, M., Geissler, C., & Shaw, J. (2019). Gestural representations of tone in Mandarin: Evidence from timing alternations. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019 (pp. 1803-1807). Canberra, Australia: Australasian Speech Science and Technology Association Inc.