From Gestural Landmarks to Analysis-by-Synthesis: Tone-driven Gestural Timing in Tibetan

Christopher Geissler

Heinrich Heine University Düsseldorf

This study investigates how tone affects the timing of oral articulatory gestures. This is accomplished through a combination of data gathered from electromagnetic articulography (EMA) and synthesis of articulatory trajectories.

According to the the Coupled Oscillator Model of articulatory timing [1], [2], an articulatory gesture is modeled as an oscillating system. Gesture oscillators are hypothesized to be coordinated in pairwise relationships with one of two stable coupling modes: *in-phase*, referring to synchronous timing as in C-V pairs, and *anti-phase*, referring to sequential timing as in C-C, V-V, and V-C pairs. Crucially, a complex onset (CCV) consists of both in-phase C-V and anti-phase C-C couplings, with the competition between them resulting in an intermediate phase relationship. This has been invoked to explain the well-documented "C-center effect" of partial overlap in complex onsets ([1].

Lexical tone has been incorporated into this framework with tone gestures, coordinated in-phase to consonants and anti-phase to vowels. This was proposed as an explanation for non-simultaneous C-V timing in Mandarin [3], and has been invoked to describe similar observations in Thai [4] and Tibetan [5], [6]. However, these studies suffer from methodological limitations: they use a standard heuristic for identifying gesture start times (usually, that gestures begin at 20% of the peak velocity towards target), and subtracting the start times of C and V gestures to derive "C-V lag".

A more robust approach would compare entire articulatory trajectories from experimental data with simulated trajectories produced in accordance with the coupled oscillator model.

Using this technique, as implemented by the Task Dynamics Application (TADA) [7], three types of stimuli were simulated, drawing on hypotheses from [8], and all using the syllable [ma]. The first used in-phase coupling, as would be expected for a CV syllable without tone. The second was timed with the partial overlap of cluster-like timing, as predicted for a tonal syllable Third used anti-phase coupling, an alternative arrangement that would generate a longer C-V lag. These trajectories were compared with real data from one of the speakers recorded using EMA in [8], producing the syllable [má] in the carrier phrase [tsìk tì __ tùk] 'this word is __'. For both simulated and real data, two articulatory trajectories were of interest: *lip aperture*, the distance between upper and lower lip sensors; and *tongue dorsum*, the position of this sensor on the horizontal axis. The [m] included a lip aperture closing gesture, and the [a] included a tongue dorsum retraction gesture. These sensors were used to identify a window in the experimental data: the first local maximum lip aperture during the previous (acoustic) segment, and the first local maximum in tongue dorsum fronting in the following stop closure. These were chosen as clear points before the onset of the lip closure, and after the end of the tongue dorsum retraction.

As shown in Figure 1, the lip aperture gesture in the observed data exhibits timing intermediate between that of the in-phase and competitive-coupling simulations. Unlike the above-cited studies, which reduced gestures to single points corresponding to heuristically-identified gestural starts, this result suggests that competitive coupling may not be sufficient to explain C-V coordination in Tibetan. Synthesis of articulatory trajectories by manipulating additional parameters offers a compelling method to address the problem.

doi:10.11576/pundp2022-1022

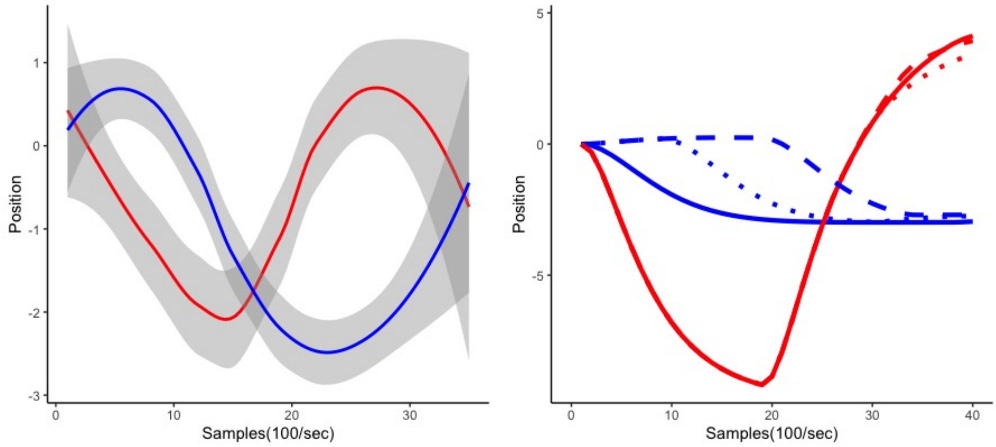

Figure 1: Articulatory trajectories over time. Tongue dorsum in blue, lip aperture in red. Left: [ma] in carrier phrase. Right: simulated [ma]; solid = in-phase, dotted = competitive coupling, dashed = anti-phase.

References

[1] Catherine P. Browman and Louis Goldstein. Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée*, (5):25–34, 2000.

[2] Hosung Nam and Elliot Saltzman. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of the Phonetic Sciences*, 2003.

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

[4] Robin Karlin. The articulatory TBU: gestural coordination of tone in Thai. In *Cornell Working Papers in Linguistics*, 2014.

[5] Fang Hu. Tones are not abstract autosegmentals. In *Speech Prosody*, pages 302–306, 2016.

[6] Christopher Geissler, Jason Shaw, Fang Hu, and Mark Tiede. Eccentric CV timing across speakers of diaspora Tibetan with and without lexical tone contrasts. In *Proceedings of the 12th International Seminar on Speech Production.*, 2021.

[7] Elliot Saltzman, Hosung Nam, Jelena Krivokapic, and Louis Goldstein. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Speech Prosody*, page 10, 2008.

[8] Christopher Geissler. *Temporal articulatory stability, phonological variation, and lexical contrast preservation in diaspora Tibetan*. PhD thesis, Yale University, 2021.

doi:10.11576/pundp2022-1022