Is lexical tone like a consonant? An ultrasound investigation of Bangkok Thai

Sarah Ellen Johnson and Ryan K. Shosted
Department of Linguistics, University of Illinois at Urbana-Champaign

We consider lexical tone as a dynamically organized gesture in order to investigate the temporal relationship between supralaryngeal and laryngeal articulators. The current study assesses the recent proposal that lexical tone is coordinated in a similar manner as a consonant in the Articulatory Phonology (AP) framework (Gao, 2008). According to this model lexical tone is temporally coordinated with onset consonants around the articulatory start of the vowel, in a similar manner as the so-called “c-center” relationship exhibited by consonant clusters in syllable onset (Marin & Pouplier, 2010). Gao claims to observe a similar pattern where the temporal center of articulation of Mandarin tones and consonants occurs around the start of the vowel, as if lexical tone were treated like an extra onset consonant in terms of timing.

This proposal is in unique contrast to autosegmental approaches that situate tone on a separate organizational tier than consonants and vowels. Indeed, suprasegmentals and segmentals have often been described as the “tune” and the “text”, respectively. However, when we consider this AP model of tone in light of consonant-based tonogenesis theories, the proposed articulatory relationship between consonants and tone does not seem like an arbitrary one. In Haudricourt’s (1954) analysis of tonogenesis, surrounding consonants’ laryngeal settings influenced the f0 height and contours of lexical tone in a previously toneless language, Vietnamese. Is the c-center coordination Gao observes with lexical tone and onset consonants a product of structural phonological similarity, support for which we observe in the tendency for consonant-based laryngeal setting to spread to or initiate the development of lexical tone? By reassessing our models of consonant-tone coordination, we will gain a deeper understanding of how the laryngeal and supralaryngeal articulators coordinate to produce speech. This may even have repercussions on methods in L2 pronunciation teaching or speech therapy for varieties of dysphonia by shedding light on the question: Should tone be taught as “tune” to sing over “text”, as synchronously integrated articulations, or some combination of these approaches?

Gao’s (2008) model of tone deserves further kinematic analysis within the AP framework. To do so, we adapt measures used in past study of syllable structure with ultrasound and acoustic methods for Bangkok Thai speech, a language that possesses both consonant clusters and lexical tone. We consider two timing phenomena associated with consonant production and assess whether lexical tone is produced in similar ways. These phenomena include gesture shift in a competitively coupled onset and stability (less variability) of the c-center (Marin & Pouplier, 2010). If lexical tone were coupled with consonants in the syllable onset, we would expect the first tonal gesture to shift later in time as extra consonants are added to the onset. We would also expect a c-center measured to include the tonal gesture to manifest more stability as compared to a c-center with only consonants.

We analyze the speech dynamics of three speakers of Bangkok Thai using kinematic and f0 velocity profiles. We accomplish this by performing optical flow analysis of ultrasound video (Horn & Schunck, 1981; Moisik & Esling, 2010) (Figure 1). By doing so, we demonstrate that the ultrasound optical flow method can be used to produce a continuous signal of supralaryngeal articulation while obviating the intrusive nature of intraoral devices like electromagnetic articulography (EMMA).

Speakers produced sentences containing alternations of the falling and rising tones in syllables differing by onset constituency: singleton or cluster combinations of a sonorous and
plosive consonants. We measured the time of achievement of maximum positive and negative velocity in the falling and rising tones’ f0, respectively. Finally, we analyzed the nature of f0 timing based on syllable onset type.

Results show that lexical tone is not produced in ways predicted by current consonant coordination models in AP. Lexical tone gestures do not follow consonant-like patterns of gestural shift with added onset consonants; and a center that includes lexical tone is not more stable than one including only consonants. Tone is not similar to a consonant, at least in terms of temporal coordination.

This result may best be interpreted in light of studies by House (1985) that find that the dynamic f0 region of rising and falling tones are crucial in perception. The rapid spectral change between a sonorous onset consonant and the following vowel is not conducive for perception; House found that listeners could not perceive a dynamic f0 change in such environments. This finding calls into question the premise of Gao’s (2008) proposal that lexical tone is competitively coupled at the syllable onset. It seems unlikely that a speaker would synchronize an f0 gesture with the vowel start, where any rapid spectral change would inhibit its perception.

What does this mean for a phonological model of tone that seeks to describe gestural coordination of the supralaryngeal and laryngeal articulators? In the case of lexical tone, we may be better served by an articulatory model that is perceptually driven. In many cases of tonogenesis, the laryngeal settings of surrounding plosive consonants incited the development of syllabic tone or the splitting of one tone category into two. However, this influence of consonants’ laryngeal features does not necessarily entail a phonological sameness between categories. Indeed, these categories require strikingly different configurations for viable perception of their articulations. For example, voiceless plosives require a lack of vocal fold vibration such that brief silence is a primary perceptual cue. Furthermore, researchers have demonstrated that lexical tone requires a substantial vocalic period in order for tonal contours to be produced fully and phonologized (Zhang, 2002). Therefore, when modeling lexical tone, rather than relying solely upon abstractly conceived coupling relations between consonants and tones, we should instead consider how speakers’ articulatory strategies sculpt the perceptual landscape to produce tractable and distinct categories in a timely manner.