Effects of co-speech gesture on magnitude and stability of oral gestures

Karee Garvin[1], Walter Dych[2], Eliana Spradling[1], Clarissa Briasco-Stewart[1], and Kathryn Franich[1]

[1]*Harvard University,* [2]*University of Delaware*

Introduction: Stress influences the spatio-temporal properties of speech, with more extreme targets and longer durations for speech gestures under stress [1,2,3,4]. Co-speech gestures tend to synchronize with stressed/metrically-prominent syllables cross-linguistically [5,6], but the influence of co-speech gestures on speech is not well-studied. Research on interlimb coordination demonstrates that synchronous coordination results in higher gesture amplitude and greater timing stability of arm movements [7]. Here, we investigate whether the presence of coordinated co-speech manual beat gestures has a similar effect on the magnitude and stability of speech articulatory gestures. Our results show that the presence of beat gestures corresponds with greater displacement and velocity of tongue movements for vocalic gestures in target words and that beat gestures increased intergestural stability of jaw and tongue movements. We conclude that co-speech gestures serve as a stabilizing and enhancing force on speech production, beyond what is observed for stress.

Method: EMA and video data from six English speakers were collected for stimuli consisting of CVbV sequences controlling for vowel quality, /i, o, a~ə/, initial consonant, /s, p, l/, and stress, /pábə/ or /pəbá/, produced in the carrier phrase *I saw the CVbV today*. Participants were asked to say all sentences as if relaying exciting news to a friend. Participants produced half of the blocks with a co-speech manual beat gesture timed with the target word with order of blocks randomized by participant. A total of 216 tokens/subject were produced (9 word shapes x 2 stress conds x 2 gesture conds x 6 reps). EMA sensors tracked the Tongue Tip (TT), Tongue Blade (TB), Jaw (JW), upper and lower lips (UL, LL). Kinematic co-speech gesture data from participants' right wrist was extracted from video data using MediaPipe [8].

Results We found a strong correlation between stressed syllable pitch accent peak f0 and co-speech gesture apex timing ($r(1501)=.96$, $p<.001$), consistent with [5,6]. GAMM analyses revealed significantly more vertical displacement and higher velocity for TT and TB in the co-speech gesture condition, though this was not significant for TB in the initial stress condition (Figs. 1,2), perhaps due to a ceiling effect of positional prominence in word-initial syllables. There was also a later onset of closure gesture following the stressed vowel for TT, as indicated by the timing of zero crossing for TT velocity in Fig. 1, supported by a significant acoustic difference in stressed vowel duration (stress*gesture β =-0.008, t=13.842, p< 0.001). The significant displacement effects were localized to the timing of the apex. Furthermore, by-vowel results in the GAMM analysis did not reflect patterns of hyperarticulation (i.e., more peripheral jaw and tongue trajectories across vowels), as predicted by [9] on stress/accent-based enhancement. We also found significantly lower lag between target achievement of TT and JW in the co-speech gesture condition (Fig. 3).

Discussion This study provides novel evidence that gesture exerts an influence on the spatiotemporal properties of articulatory gestures, with more extreme (lower) tongue and jaw displacement in the co-speech gesture condition. The effects are beyond the effects of stress, where vowel-specific results indicated a lack of more *peripheral* articulations of vowels in the gesture condition, suggesting that coordination has effects on articulation that are distinct from stress-based articulatory enhancement [10]. Likewise, the localization of these effects at the site of the gesture apex suggests that coordination with the gesture, and not another prosodic feature accompanying gesture, determines the effects. Speech-gesture synergies therefore constitute an important variable to consider when modeling the effects of prosodic prominence on articulatory patterns. Together, our results provide key insights on the mechanisms that underlie prosodic enhancement in naturalistic speech.

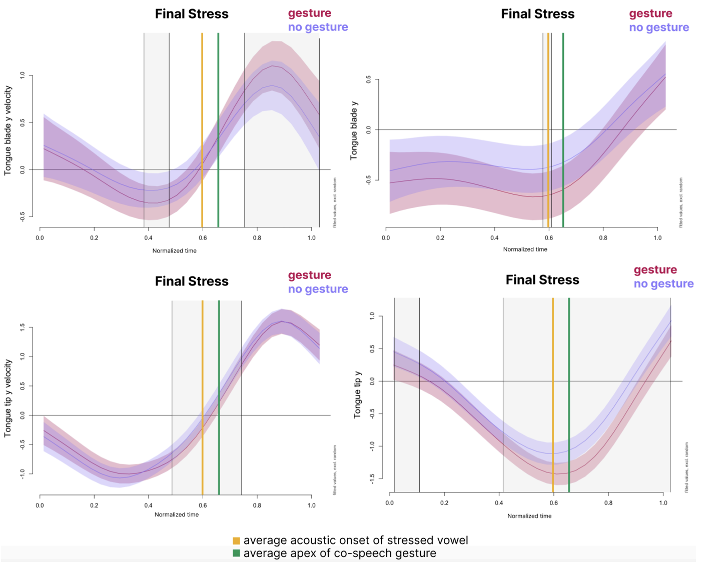

Figure 1: GAMM of vertical displacement and velocity for TB and TT in final stress cond., comparing gesture conditions. Shading indicates portions of the trajectory that are significantly different. Yellow line indicates avg. acoustic onset of stressed vowel; green indicates avg. apex time of co-speech gesture. x-axis time normalized by target word. bam(X~gesture + (time) + (time by task) + (task by subject)).

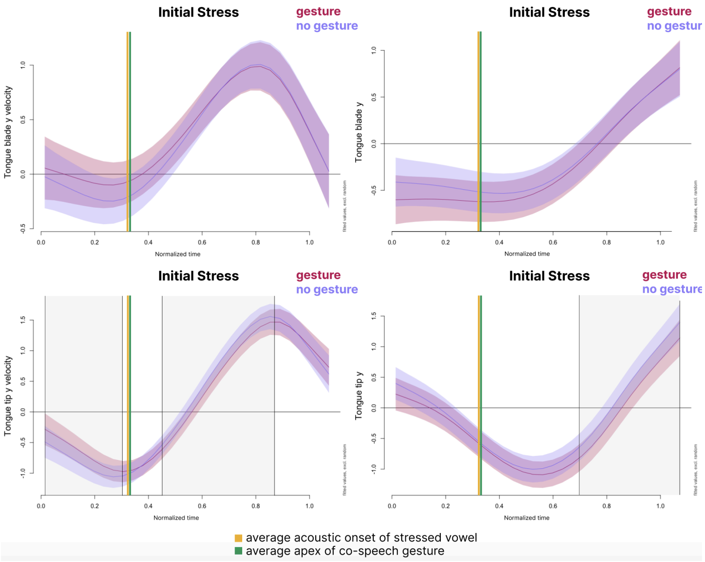

Figure 2: GAMM of vertical displacement and velocity for TB and TT in initial stress cond., comparing gesture conditions. Shading indicates portions of trajectory that are significantly different. Yellow line indicates avg. acoustic onset of the stressed vowel; green indicates avg. apex time of co-speech gesture. x-axis time normalized by target word. bam(X~gesture + (time) + (time by task) + (task by subject)).

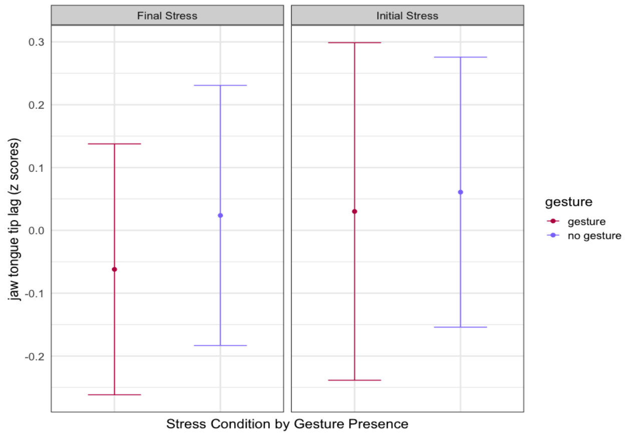

Figure 3: Lag in timing of TT max. displacement and JW max. displacement in stressed syllables. lmer(zlag~stress*gesture+(1+gesture|Subject)). stress*gesture: β=-0.06, t=-5.01, p< 0.00

References: [1] M. Beckman et al. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. [2] M. Beckman and J. Edwards. 1994. Articulatory evidence for differentiating stress categories. [3] T. Cho. 2006. Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. [4] J. Harrington et al., Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. [5] J. Krivokapic et al. 2016. Speech and manual gesture coordination in a pointing task. [6] N. Esteve-Gibert et al. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. [7] Schwartz, M. et al. 1995. Superimposition in interlimb coordination [8] Lugaresi et al. 2019. MediaPipe: A Framework for Building Perception Pipelines. [9] K. de Jong et al. 1993. The interplay between prosodic structure and coarticulation. [10] K. Franich, 2022. How we speak when we speak to a beat: The influence of temporal coupling on phonetic enhancement.