Exploring the effect of stress on gestural coordination

Yunting Gu[*]

Abstract. In this study, I examined stress in speech production within the framework of Articulatory Phonology. Specifically, I tested the hypothesis that stress could be analyzed as a prosodic gesture. Using articulatory data from an English corpus, I found that the CV lag–the gestural lag between a consonant and a vowel–of stressed syllables is significantly larger in terms of both duration and proportion than that of unstressed syllables. I also found that stressed consonant and vowel gestures are longer than unstressed ones. These findings seem to suggest that stress could be analyzed as a prosodic gesture. Moreover, my study reveals a source of variation in CV coordination, which can inform other kinematic studies.

Keywords. Articulatory Phonology; stress; gestural coordination; gestural timing; CV lag; gestural duration

1. Introduction. The research gap this study aims to address is that there is an insufficient understanding of the kinematic properties of stress (Byrd & Krivokapić 2021). Previously, it was claimed that a prosodic gesture may be attracted to stress (Byrd & Saltzman 2003) or shifts towards the stressed syllable (Katsika 2012; Byrd & Krivokapić 2021). However, there has not been much further research that followed up on these claims which connect stress with a prosodic gesture. To test the hypothesis that stress could be a prosodic gesture, I measured and compared the CV lag–the gestural lag between a consonant and a vowel–of stressed and unstressed syllables. I found that the CV lag of stressed syllables is significantly larger, in terms of both duration and proportion, than that of unstressed syllables. I also found that stressed gestures have longer duration than unstressed ones, and it is true for both consonants and vowels. The contribution of this study is threefold. First, it shows a novel kinematic correlation of stress–CV lag. Second, it reveals a source of variation in CV coordination, which can inform other kinematic studies. Thirdly, it supports the argument that stress could be analyzed as a prosodic gesture.

2. Stress in Articulatory Phonology. This study assumes the speech production model of Articulatory Phonology. There are a few previous studies that suggested stress could be a prosodic gesture. However, as far as I know, no previous English articulatory study has tested the effect of stress on gestural duration *and* inter-gestural coordination in terms of the manifestations of a prosodic gesture.

2.1. ARTICULATORY PHONOLOGY. According to Articulatory Phonology (Browman & Goldstein 1989, 1992), phonology is characterized in terms of gestures and the relations of gestures where a gesture is a basic unit and a relatively abstract concept. Under Articulatory Phonology, gestures are events that unfold during speech production, and these events consist of the formation and release of constrictions in the vocal tract. The consequences of gestures can be observed in the movement of speech articulators. A schematic illustration of a sample gesture can be found in Figure 1, where gestural onset (GON), target onset (TON), target offset (TOF), and gestural offset (GOF) are denoted from left to right (Gafos 2002).
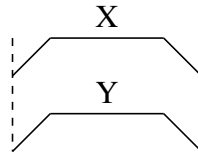
Figure 1. Sample gesture
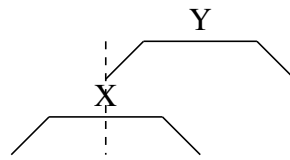

Figure 2. In-phase coordination

Figure 3. Anti-phase coordination

Within the Articulatory Phonology framework, the coupled oscillator model of syllable structure argues that syllabic structure is expressed articulatorily in differential timing relations (Nam & Saltzman 2003; Browman & Goldstein 2000; Iskarous & Pouplier 2022). Specifically, the classic view is that in-phase coordination (Figure 2) should apply to CV coordination, and anti-phase relationship (Figure 3) is applicable to CC and VC coordination.

2.2. THE PROSODIC GESTURE AND STRESS. The prosodic gesture model proposed by Byrd & Saltzman (2003) suggests that prosodic gestures "temporally stretch gestural activation trajectories" and prosodic gestures make the gestures in their activation domain longer, larger, and further apart. In Figure 4, for example, a prosodic gesture slows down the gestural coordination of gesture 1 and gesture 2 between the two dashed lines.
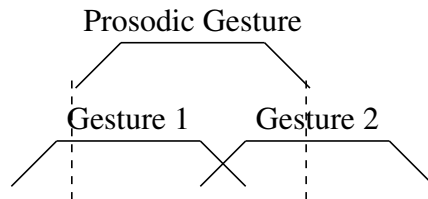

Figure 4. Prosodic gesture

There is some evidence that suggests that stress could be analyzed as a prosodic gesture. First, Katsika (2012, 2016, 2018) found that in Greek the gestures of stressed syllables were longer and larger and that the prosodic gesture shifts towards the stressed syllables. Second, Saltzman et al. (2008) argued that stress could be modeled as a μ_T gesture, which is a generalization of the prosodic gesture, to account for the observation that stress elongates a single gesture. Moreover, the kinematic patterns in English suggested that gestures in syllables with greater

stress (nuclear accented) show less coarticulatory overlap (De Jong et al. 1993).

Despite these observations, however, little research has followed up on these studies. Moreover, as far as I know, no previous study has probed stress as a prosodic gesture by testing the two effects of a) increasing gestural lag and b) gestural lengthening using English articulatory data.

3. Hypothesis. As mentioned before, a prosodic gesture makes its affected gestures longer and further apart (Byrd & Saltzman 2003). To evaluate the hypothesis that stress could be analyzed as a prosodic gesture, two sub-hypotheses as in (1) are tested in the study.

(1) a. The CV lag in stressed syllables is larger than that in unstressed syllables.

   b. The C and V gestures in stressed syllables are longer in duration than their corresponding gestures in unstressed syllables.

Sub-hypothesis (1a) concerns the gestural timing difference between a consonant and a vowel – CV lag, which can be computed by subtracting a C timestamp from a V timestamp. In this current study, each CV syllable has four absolute CV lag measurements: CV lag based on 1) gestural onset (GON), 2) target onset (TON), 3) target offset (TOF), and 4) gestural offset (GOF). For example, as indicated by the black dashed line in Figure 5, the timestamp of the gestural onset (GON) of a C is subtracted from that of the gestural onset (GON) of a V to get the CV lag based on gestural onset. Similarly, the timestamp of the target onset (TON) of a C is subtracted from that of the target onset (TON) of a V as in the blue dashed line, which denotes the CV lag based on target onset.

To control for the effect of speech rate and the fact that stressed syllables are longer, normalized CV lags are calculated by dividing absolute CV lags by the duration of a syllable. The duration of a syllable is computed by calculating the difference between the maximal and the minimal timestamps among the 8 timestamps in a CV syllable – 4 from C and 4 from V. In Figure 5, for instance, the syllable duration is indicated by the red dashed line. Moreover, dividing the CV lag based on gestural onset (i.e., black dashed line) by the syllable duration (i.e., red dashed line) gets the normalized CV lag based on gestural onset, one of the 4 normalized gestural lag measurements. Therefore, in the current study there are 8 CV lag measurements – 4 absolute and 4 normalized – for each CV syllable. Each CV lag measurement of stressed syllables is then compared to the corresponding CV lag measurement of the unstressed syllables to evaluate sub-hypothesis (1a).
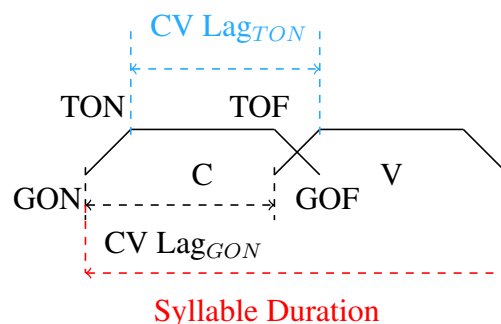

Figure 5. CV lag computation

Sub-hypothesis (1b) is tested by comparing the gestural duration of stressed and unstressed syllables. Specifically, the gestural duration is computed by subtracting the gestural onset timestamp from the gestural offset timestamp of a C or V gesture. While vowel gestural durations in stressed and unstressed syllables are compared to each other, consonant gestural durations are analyzed separately.

4. Methods.

4.1. THE CORPUS. To test the hypothesis that stress could be a prosodic gesture, I analyzed the kinematic data from the Wisconsin X-ray Microbeam Database (Westbury et al. 1990). To collect data for this corpus, microphones were used to record acoustic signals and several pellets were placed on each speaker's head. Figure 6 shows the positions of the pellets schematically. To obtain reference points indicated by Ref in Figure 6, three pellets were attached to the speaker's head: one on the bridge of the nose, the second on buccal surface of the maxillary incisors, and the third either on the nosebridge lower than the first or an arm projecting from a snug-fitting pair of eyeglass frames. To extract information about tongue movement, four pellets, which are denoted by T1 to T4 in Figure 6, were attached along the longitudinal sulcus of each speaker's tongue. T1 was placed 10 mm posterior to the tongue tip, and T4 was placed about 60 mm posterior to the tongue tip, depending on each speaker's tolerance. Positions of T2 and T3 were chosen so that the four tongue pellets were equally distanced. As for labial articulation, one pellet each was attached to the upper lip (UL) and lower lip (LL). The data collected by each pellet was indicated by the same label in Matlab by using the *lp_findgest* algorithm of the *mview* package where each pellet's movement was indicated by one row of curves (Tiede 2005) as in Figure 7.
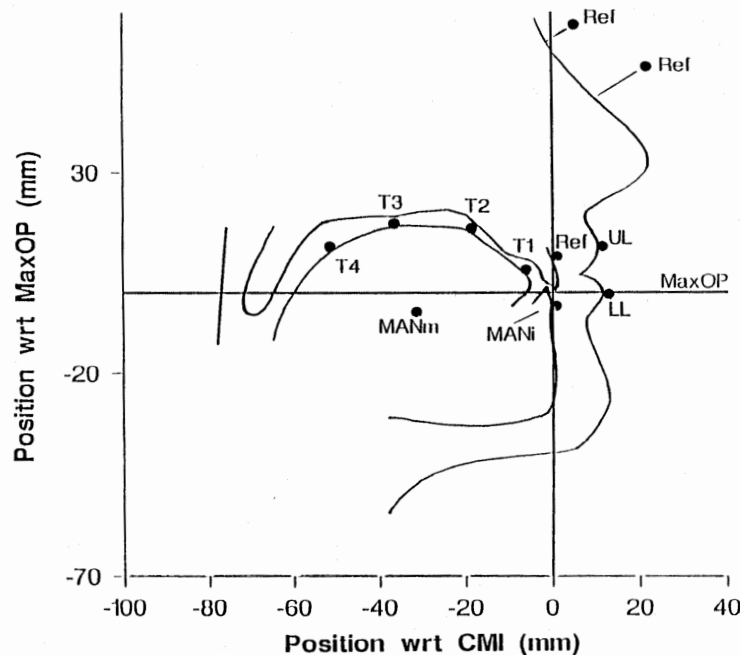

Figure 6. Approximate pellet placement locations. This figure is reproduced from Figure 5.2 of the Wisconsin X-ray Microbeam Database manual (Westbury et al. 1990).

4.2. STIMULI. The stimuli of the experiment, namely word pairs of stressed and unstressed CV syllables, are shown in Table 1. While most stimuli are produced in word list reading tasks, the stimulus ⟨banana⟩ comes from paragraph reading and the stimulus ⟨combine⟩ occurs in a sentence reading task as in *He always answers, 'Banana oil!'* and *Combine all the ingredients in a large bowl.*

The consonant and vowel gestural measurements are shown in the last two columns of Table 1. The measurement of C and V were chosen based on previous literature (Gao 2008; Zhang et al. 2019; Hall 2010) and an understanding of the articulatory events involved. For example, [n] involves tongue tip alveolar closure gestures (Hall 2010), so T1 which stands for tongue tip was measured for the consonant [n]. Similarly, the dental consonant in ⟨thi⟩ involves tongue tip and T1 was measured. The feature labial in [m] corresponds to the use of the lip tract variables (Gao 2008; Zhang et al. 2019; Hall 2010), and lower lip (LL) was measured for [m]. Furthermore, the velar stop [k] was measured by T4 which represents tongue root, and the vowel gestures of the stimuli were measured by T2 or T3, which stands for tongue blade.

Index	Stressed	Unstressed	C-Measurement	V-Measurement
(1)	ba**na**na [næ]	bana**na** [nə]	[n]-T1	[æ/ə] -T2/T3
(2)	**mo**ment [mo]	al**mo**st [mo]	[m]-LL	[o]-T2/T3
(3)	**thi**ngs [θɪ]	no**thi**ngs [θɪ]	[θ]-T1	[ɪ]-T2/T3
(4)	be**co**me [kʌ]	**co**mbine [kə]	[k]-T4	[ʌ/ə]-T2/T3

Table 1. Stimuli. T1 stands for tongue tip, T2 or T3 tongue blade, T4 tongue root, and LL lower lip. All stimuli except for ⟨banana⟩ and ⟨combine⟩ come from word list reading.

4.3. DATA ANNOTATION AND ANALYSIS. The articulatory trajectories of the stimuli were annotated in Matlab using the default settings of the *lp_findgest* algorithm of the *mview* package, which means that gestural onset, gestural offset, target onset, and target offset used the 20 percent threshold (Tiede 2005). Based on the information of the acoustics as well as the articulatory movement trajectories, I annotated the consonant gesture and the vowel gesture of each token. For instance, Figure 7 shows the annotation of ⟨na⟩ in ⟨banana⟩ – where only relevant rows T1 (in blue) and T2 (in purple) are displaced for clarity. In one T1 gesture, gestural onset (GON), target onset (TON), target offset (TOF), and gestural offset (GOF) were denoted by white texts, and the timestamp information of these four landmarks was recorded for each gesture.

To evaluate the effect of stress on inter-gestural coordination, plots were generated by the *tidyverse* package (Wickham et al. 2019) and mixed-effects modeling with the random intercept of participants was conducted by the *lme4* (Bates et al. 2014) and *lmerTest* (Kuznetsova et al. 2017) packages in R (R Core Team 2017).

5. Results. Analyzing the data shows that 1) CV lag in stressed syllables is longer than that in unstressed syllables; 2) the duration of a single gesture in stressed syllables is larger than that in unstressed syllables. These suggest the hypothesis that stress could be a prosodic gesture has been supported by the current study.

5.1. CV LAG COMPARISON. The timing difference between the C and V gestures in stressed and unstressed syllables was compared descriptively and statistically. The descriptive plots in Figure 8 shows that the lag between a consonant and a vowel increases with stress for gestural
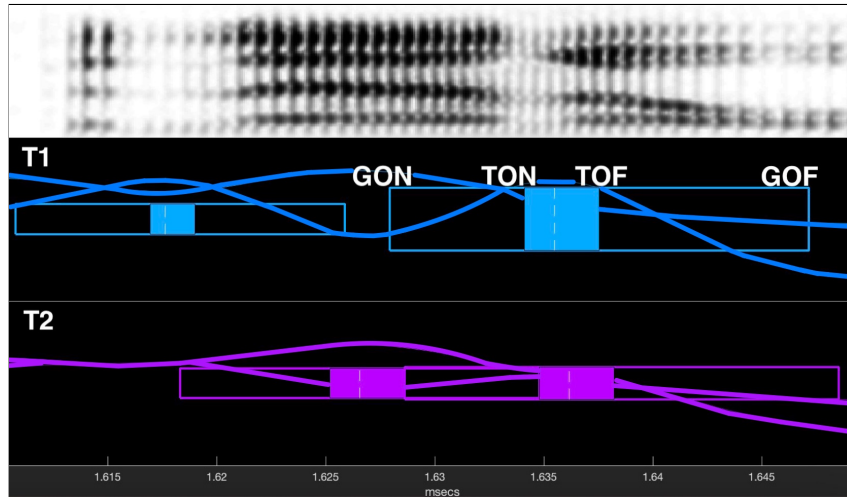
Figure 7. Sample annotation of ⟨banana⟩ from speaker JW16. One gesture is added with land-mark labels of gestural onset (GON), target onset (TON), target offset (TOF), and gestural offset (TOF). The bottom of the figure indicates timestamp information so that four timestamps were recorded for each gesture. Note that in this sample annotation for ⟨banana⟩ where the former ⟨na⟩ syllable is stressed and the latter ⟨na⟩ syllable is unstressed, the CV lag of the stressed syllable is larger as hypothesized.

onsets. This pattern is true for all 6 other gestural measurements including absolute and normalized gestural offset, target onset, and target offset.[1] To test the statistical significance of the observations, mixed-effect models with the random intercept of participants were fitted. All measurements exhibit statistical significance as exemplified by the models for gestural onsets shown in Table 2.
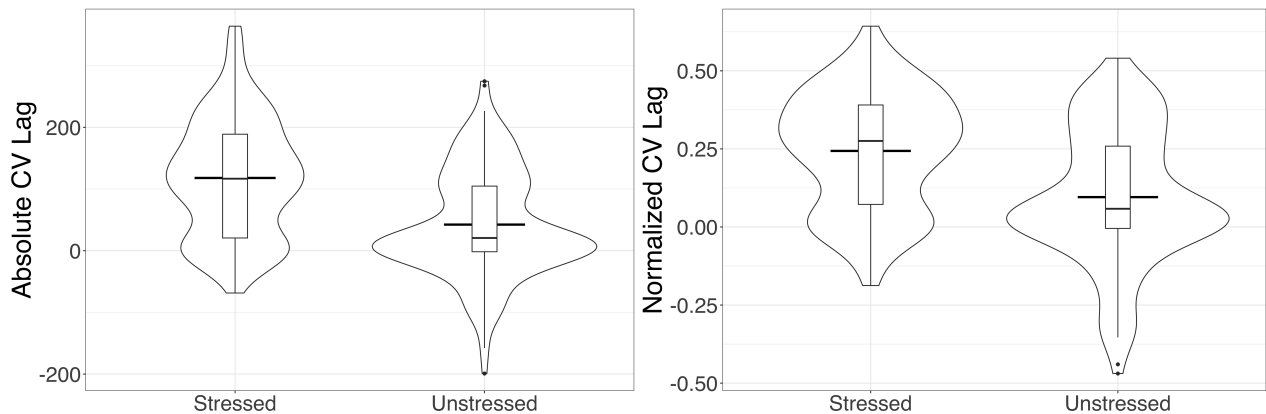

Figure 8. CV lag based on gestural onsets increases with stress

To understand the data in more depth, the whole dataset was also separated by word pairs and sample descriptive results can be found in Figure 9. For normalized CV lag based on gestural

---

[1] Other plots are not presented here in the interest of space.

CV Lag Measurement	Estimate	Std. Error	df	t value	Pr(>|t|)
Absolute Gestural Onset					
(Intercept)	117.88	7.01	329.00	16.81	<0.001
Condition Unstressed	-75.59	9.96	329.00	-7.59	<0.001
Normalized Gestural Onset					
(Intercept)	0.24	0.02	329.00	15.57	<0.001
Condition Unstressed	-0.15	0.02	329.00	-6.66	<0.001

Table 2. Mixed effects model results for CV lag based on gestural onsets

onsets, all four word pairs exhibit the same expected pattern that CV lag in stressed syllables are longer.
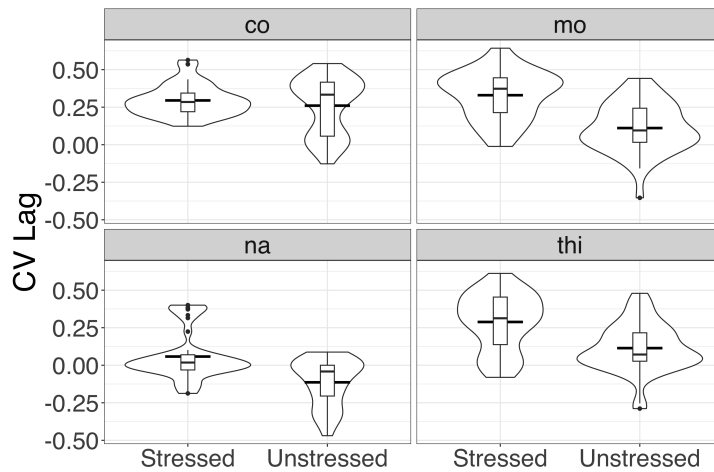

Figure 9. Normalized CV lag based on gestural onset by word pairs

5.2. GESTURAL DURATION COMPARISON. Comparing single gestural duration in stressed and unstressed syllables yields the following results shown in Figure 10 and Table 3. Both the consonant and vowel gestural duration in stressed syllables are statistically significantly larger than those in unstressed syllables. Therefore, just like a prosodic gesture, stress lengthens gestures.

Measurement	Estimate	Std. Error	df	t value	Pr(>|t|)
C Duration					
(Intercept)	269.19	5.03	330.00	53.52	<0.001
Condition Unstressed	-32.82	7.16	330.00	-4.59	<0.001
V Duration					
(Intercept)	315.77	6.73	122.78	46.89	<0.001
Condition Unstressed	-25.24	9.01	284.43	-2.80	0.01

Table 3. Mixed effects model results for gestural duration

If the dataset is separated by word pairs, the consonant duration comparisons in stressed
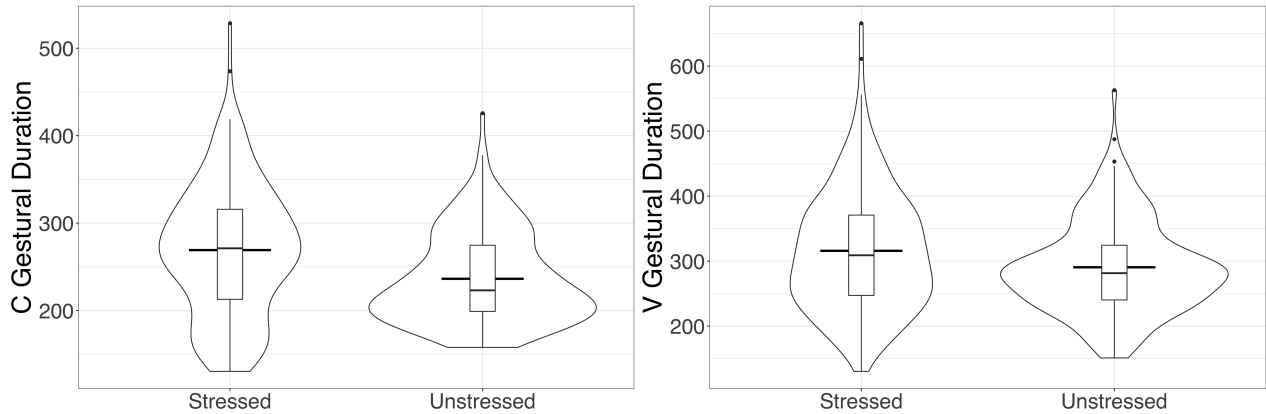
Figure 10. Gestural duration increases with stress

and unstressed syllables (Figure 11) are not equal for different word pairs in question. Specifically, while stressed consonants are longer for the ⟨mo⟩ pair ⟨moment, almost⟩ and the ⟨thi⟩ pair ⟨things, nothing⟩, the other two pairs do not seem to show this pattern.[2] As for the vowel duration comparison shown in Figure 12, stressed vowels are likely to be longer for all word pairs, though the ⟨co⟩ pair ⟨become, combine⟩ and the ⟨mo⟩ pair ⟨moment, almost⟩ show the patterns in a slightly more obvious way. The reason the ⟨mo⟩ pair shows the duration patterns in a more obvious way may be that the bilabial nasal [m] and vowel use different articulators – lip and tongue respectively. Further investigations are necessary to determine whether there is a meaningful durational difference by word pair and why.
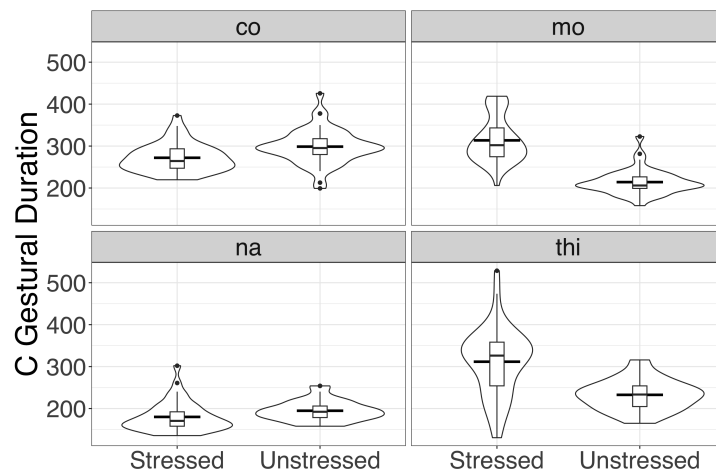

Figure 11. C duration comparison by word pairs

[2] Syllable position may play a role here since the ⟨mo⟩ and ⟨thi⟩ pairs have stressed syllables word initially and unstressed syllables word-medially, while the other pairs have stressed syllables word-medially and unstressed syllables at word boundaries. Also, it is true that it could be a random variation.
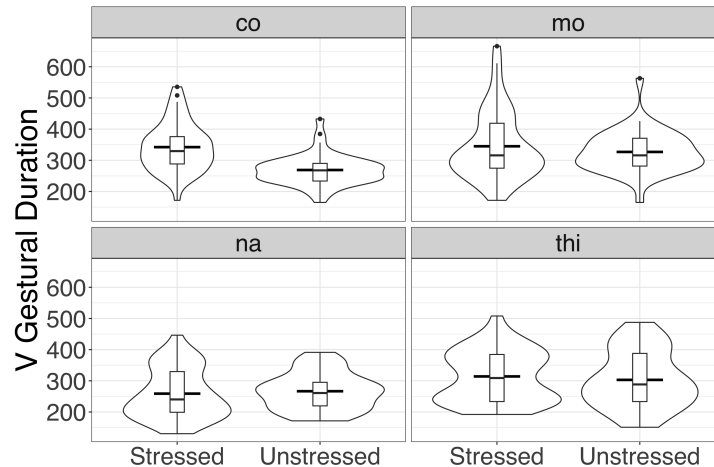
Figure 12. V duration comparison by word pairs

6. Discussion. The descriptive and statistical analysis clearly showed that the CV lag in stressed syllables is significantly larger than the CV lag in unstressed syllables. Moreover, the vowel and consonant gestures in stressed syllables are longer in duration than those in unstressed syllables. These findings suggest that stress itself can be analyzed as a prosodic gesture–since a prosodic gesture slows down the internal clock used for articulatory planning (Byrd & Saltzman 2003). Since previous studies have claimed stress *attracts* prosodic gestures, future studies are necessary to make a nuanced distinction between the claim stress is a prosodic gesture and that stress attracts prosodic gestures.

One implication of the current study is relevant to CV coordination, which serves as one of the fundamentals in Articulatory Phonology. Since stress can introduce variation in CV coordination, the variable of stress needs to be carefully controlled and considered in articulatory studies to avoid experimental confounds or misinterpretation. For example, Zhang et al. (2019) analyzed the kinematic data in Mandarin and found that the CV lag for the full-tone condition was significantly greater than the lag in the toneless condition. This seemed to suggest that the tone gesture has a sequential relationship with CV gestures. Given the effect of stress on CV alignment and the fact that Mandarin toneless syllables are always weakened and unstressed (Chao 1965; Lin 2000; Yip 2002; Lee 2003), the results in Zhang et al. (2019) can be accounted for purely in terms of the stress difference in the stimuli, without invoking a need for distinguishing different alignments based on tone.

The major shortcoming of this current study is that its stimulus choice is limited by the available corpus and set experimental designs. Future studies could improve by using stimuli that have controlled word position, syllable position, and vowel quality.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/gcrnkw.
Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6(2). 201–251. https://doi.org/bj2qt9.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49(3-4). 155–180. https://doi.org/10.1159/000261913.

Browman, Catherine P & Louis Goldstein. 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.

Byrd, Dani & Jelena Krivokapić. 2021. Cracking prosody in articulatory phonology. *Annual Review of Linguistics* 7. 31–53. https://doi.org/10.1146/annurev-linguistics-030920-050033.

Byrd, Dani & Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31(2). 149–180. https://doi.org/fsp7br.

Chao, Yuen Ren. 1965. *A grammar of spoken Chinese.* Berkeley: University of California Press.

De Jong, Kenneth, Mary E. Beckman & Jan Edwards. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2-3). 197–212. https://doi.org/ggdc52.

Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20(2). 269–337. https://doi.org/10.1023/A:1014942312445.

Gao, Man. 2008. *Mandarin tones: An articulatory phonology account*. New Haven, CT: Yale University dissertation.

Hall, Nancy. 2010. Articulatory phonology. *Language and Linguistics Compass* 4(9). 818–830. https://doi.org/10.1111/j.1749-818X.2010.00236.x.

Iskarous, Khalil & Marianne Pouplier. 2022. As time goes by: A critical appraisal of space and time in articulatory phonology in the 21st century. Manuscript. UCLA.

Katsika, Argyro. 2012. *Coordination of prosodic gestures at boundaries in Greek*. New Haven, CT: Yale University dissertation.

Katsika, Argyro. 2016. The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics* 55. 149–181. https://doi.org/10.1016/j.wocn.2015.12.003.

Katsika, Argyro. 2018. The kinematic profile of prominence in Greek. Presentation. Speech Prosody 2018, Poznań, PL.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmertest package: tests in linear mixed effects models. *Journal of Statistical Software* 82. 1–26.

Lee, Wai-Sum. 2003. A phonetic study of the neutral tone in beijing mandarin. *ICPhS* 15. 1121–1124.

Lin, Hua. 2000. Stress and the distribution of the neutral tone in mandarin. In De Bao Xu (ed.), *Chinese phonology in generative grammar*, 139–161. Leiden: Brill.

Nam, Hosung & Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. *ICPhS* 15(1). 2253–2256.

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.r-project.org.

Saltzman, Elliot, Hosung Nam, Jelena Krivokapic & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. Presentation. Speech Prosody 4. Campinas, BR.

Tiede, Mark. 2005. Mview: software for visualization and analysis of concurrently recorded movement data. New Haven, CT: Haskins Laboratories .

Westbury, John, Paul Milenkovic, Gary Weismer & Raymond Kent. 1990. X-ray microbeam speech production database. *The Journal of the Acoustical Society of America* 88(S1). S56. https://doi.org/10.1121/1.2029064.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang […] Hiroaki Yutani. 2019.
    Welcome to the tidyverse. *Journal of Open Source Software* 4(43). 1686.
    https://doi.org/10.21105/joss.01686.
Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.
Zhang, Muye, Christopher Geissler & J. A. Shaw. 2019. Gestural representations of tone in man-
    darin: evidence from timing alternations. *ICPhS* 19. 1803–1807.