Temporal Coupling Between Speech and Manual Motor Actions

Benjamin Parrell [1]*, Louis Goldstein [1,2], Sungbok Lee [1,3], Dani Byrd [1]

[1] Department of Linguistics, University of Southern California
3601 Watt Way, GFS 301
Los Angeles, CA 90089-1693

[2] Haskins Laboratories
300 George Street
New Haven, CT 06511

[3] Ming Hseih Department of Electrical Engineering, University of Southern California
3740 McClintock Ave, EEB 428
Los Angeles, CA 90089-2562

parrell@usc.edu

Abstract. *Speech production is part of the larger motor control system, and as such can be organized into coordinative structures with other motor behaviors that can then be parameterized as single functional units. Previous work has shown the magnitude of movements in coordinated speech and finger tapping covaries across repetitions. The current study expands on this previous work to examine the temporal coordination of these coordinated movements. Subjects repeated a spoken syllable while tapping their finger in time with their speech. On each trial, they placed an emphatic stress on a single repetition of either a spoken syllable or a finger tap, and were instructed to maintain their production in the other domain unchanged. Results show that explicit stress in either speech or finger tapping generally lengthens the stressed repetition in that domain as well as the synchronous repetition in the other domain. Additionally, stress of either a spoken syllable or finger tap causes some lengthening of the interval between repetitions in both domains. This cross-domain covariation of the interval between repetitions extends to unstressed repetitions as well. These results indicate that control of prosody may be defined over a wide set of articulatory components, not just the speech motor system.*

1. Introduction

Speech has long been studied as part of the motor control system. One particular avenue of research has examined coordination between the speech and non-speech aspects of this system, beginning with work by Kelso, Tuller, & Harris (1983). In this study, subjects were instructed to repetitively speak a monosyllabic word and tap their finger in time with their speech. Subjects either spoke or tapped in an alternating stressed-unstressed pattern, and were instructed to keep constant their other production (tapping during spoken stress alternation, speaking during tapped stress alternation). Despite these instructions, subjects consistently produced larger taps when those were synchronous with a *stressed* spoken word than for those synchronous with an *unstressed* word; similarly, words were produced with greater acoustic intensity when they co-

* Work supported by NIH DC03172.

occurred with a stressed tap than an unstressed tap. These findings were confirmed by our team in the articulatory domain for speech as well (Parrell, Goldstein, Lee & Byrd, 2010). Our previous study additionally showed a correlation in the magnitude of lip and finger movements *in the unstressed repetitions*, from which we concluded that the coupling of the two motor domains exists independently of the effects of explicit emphasis.

Additional support for the shared connection between speech and other motor movement comes for the literature on manual pointing. When pointing at and naming an object with a demonstrative (e.g. "this lamp"), vocalization begins at the apex of the pointing movement (Levelt, Richardson & Heij, 1985). Additionally, and perhaps more directly relevant, is the study of coproduced deictic pointing gestures and speech by Rochet-Capellan and colleagues (Rochet-Capellan, Laboissière, Galván & Schwartz 2008). In this study, subjects had to point at a smiley face while naming it with either a ˈCVCV or CVˈCV word. Results show that the stressed syllables are aligned with the pointing gestures, indicating a coordination between the two actions. There is some evidence that this coordination between prosodic emphasis and pointing gestures may be mediated by shared neural circuitry between the two domains, at least for contrastive stress (Lœvenbruck, Dohen & Vilain 2009).

This body of work has focused mainly on the spatial effects of motor coupling, largely passing over temporal effects. One study that has looked at this aspect of coordination is Smith, MacFarland, & Weber (1986). This study examined the relative timing of tapping and synchronous speech produced repetitively without stress and found that subjects produced consistently aligned productions (within ¼ cycle), though this study used acoustic, rather than articulatory, measures of speech production. This is evidence that there is temporal as well as magnitude entrainment between speech and nonspeech motor actions. This study, however, did not examine timing coordination under stress. Stress, as part of the larger prosodic system, is known to have large effects in nonrepetitive speech including lengthening near a prosodic boundary (for a review see Byrd, Krivokapić & Lee, 2006) and under stress/accent (e.g. Beckman & Edwards, 1994). Rhythmic tapping alone exhibits its own temporal patterns, with stressed repetitions either followed generally by a long-short tapping pattern, though some subjects show a short-long pattern (Nagasaki, 1987). The current study, building on our previous articulatory work, explores the temporal effects of stress in a multi-modal coordinative system of speech and manual motor actions.

2. Methods

2.1. Procedure and subjects

Subjects were instructed to rhythmically tap their right finger on their shoulder while synchronously repeating a monosyllabic word at a self-selected rate. Subjects were presented with a clock face with "stars" at 3:00, 6:00, 9:00, and 12:00 and with hash marks halfway between each star, on which a second hand swept continuously in the clock-wise direction. Subjects were instructed to begin tapping and speaking when the sweeping second hand reached one star, and continued until it reached the next (i.e., for 15 seconds). At a subject-selected point near the marked midway point between stars, the subject either make a single finger tap emphatic (in condition 1) or place an emphatic stress on one repetition of the spoken syllable (in condition 2). Subjects were instructed to maintain their repetitions of the unstressed modality unchanged at a steady rate. 10 trials were collected per block. There were two blocks per condition (finger or spoken stress), one using the syllable "ma" and one using the syllable "mop," giving a total of four blocks and 40 trials. Four right-handed subjects (TA, TB, TC, TD) participated.

2.2. Data collection

Articulatory data was collected using an electromagnetic articulometer (Carstens AG500). This device allows three-dimensional tracking of transducers adhered to the articulators. For this study, transducers were adhered to the upper and lower lips, and the tip of the right index finger. Reference sensors were adhered to the nose ridge and behind each ear. Articulatory data was collected at 200 Hz, and acoustic data at 16 kHz. After collection, the articulatory data was smoothed with a 9th-order Butterworth low pass filter with a cut-off frequency of 5 Hz, rotated to match the subject's occlusal plane, and corrected for head movement using the reference sensors.

2.3. Data Analysis

Movement of the lips was calculated using a derived measure of Lip Aperture (LA), defined as the Euclidian distance between sensors placed on the upper and lower lips. For both LA and fingertip (FT) movement, the following points were identified from the velocity trajectories (tangential velocity was used for the multidimensional finger tapping data): the point of velocity minimum (point of maximum constriction), the peak velocities both before and after the velocity minimum, and the onset and offset of movement (defined as the point at 20% of the difference in speed between speed maxima and the preceding of following minima, respectively). For FT movements, the point of maximum constriction was measured when the finger was touching the shoulder (identified by a plateau in the velocity signal), so the lowering of the finger towards that target constituted the first part of the movement and the raising of the finger, the second.

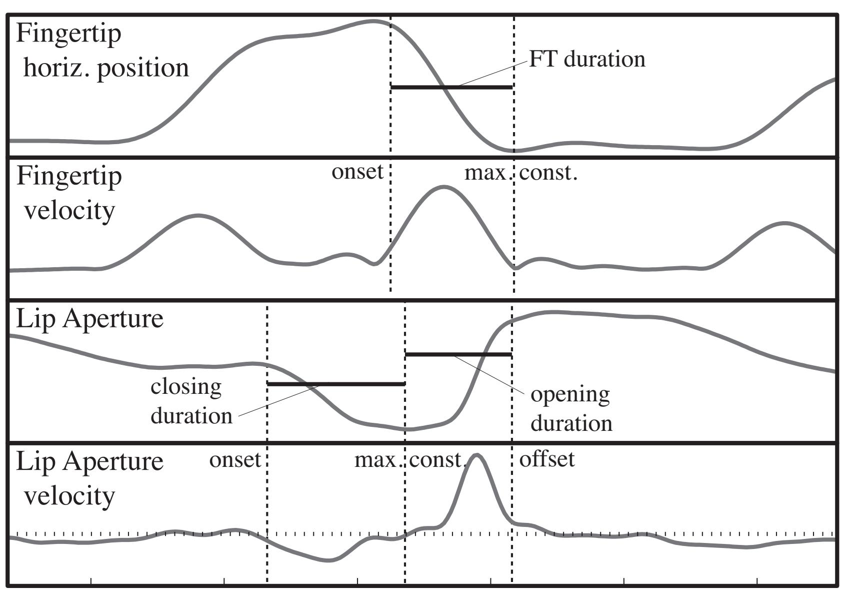

Figure 1. Labeling and measurements made from articulatory data. Location of relevant articulatory events is based on the velocity signal for both lip and finger movements; horizontal motion of the finger tip shown for illustration only. Both opening and closing durations of syllable-initial /m/ were measured.

Duration of LA movement was split into two parts: the closing motion into the constriction for /m/ and the opening motion from the constriction for /m/ into the following vowel /a/. The closing motion was defined as the time between onset of movement to the point of maximum constriction and the opening motion as the time between the point of maximum constriction and movement offset. FT movement duration was calculated as the time from the velocity minimum at the moment of maximal distance from the shoulder to the velocity minimum where the finger contacted the shoulder (Figure 1).

3. Results

In order to test whether any lengthening occurred under stress, the duration of the stressed repetition of each individual trial was compared against the mean of all unstressed repetitions for the same trial. When the stress was on a spoken syllable, three of four subjects (TA, TB, TD) showed a significantly longer duration for one or both of the LA movements. Subjects TA and TD additionally showed a significantly longer finger tap in this condition when it was synchronous with the stressed repetition than when not. When the fingertip repetition was stressed, only subjects TA and TD showed a longer FT movement (the same who showed FT lengthening in the spoken stress condition). Subjects TB, TC, and TD additionally showed longer LA movements in this condition. In general, subjects who showed an effect of lengthening under intra-modal stress also showed that effect when the other modality was stressed, despite instructions to keep those productions constant and unchanged.

An alternative way to look at temporal variation is to look at the time between successive repetitions, or Inter-impulse Interval (IRI, Nagasaki, 1987). This was measured as the time between the onset of one repetition and the onset of the following, for both LA and FT movements. Subjects in general showed either lengthening of the interval *before* the stressed repetition or of the interval *after* the stress, sometimes with concurrent shortening of the opposing repetition. We can see in the table below that the pre-stress lengthening occurs in two cases (TA and TD, spoken stress), all when the stress was placed on a spoken syllable. Post-stress lengthening occurs in eleven cases, the majority of which occur when the tap was stressed. Both patterns occur in one case (TD, spoken stress). Despite intentional stress in only one domain, IRI lengthening occurs in both domains in the majority of cases.

	Subject TA		Subject TB		Subject TC		Subject TD	
Stress:	*spoken*	*tapped*	*spoken*	*tapped*	*spoken*	*tapped*	*spoken*	*tapped*
LA IRI	+/-	0/+	-/+	-/+	-/+	0/+	+/0	0/+
FT IRI	0/+	0/0	-/+	-/+	-/+	0/0	+/+	-/+

Table 1. Durational effects of stress shown by subject. Lengthening is shown by a "+"; shortening, by a "-"; no change, by "0." In each table cell, prestress IRI is on the left; poststress, on the right.

In addition, the IRIs for LA and FT are significantly correlated with each other during *the entire reiterant utterance* for all subjects regardless of stress. The covariation we see in Table 1, then, is not just a local effect of stress but rather a more pervasive characteristic of the coordination between the two tasks.

4. Conclusion

The current study demonstrates that the coordination between speech and finger movments in a repetitive task exists in the temporal domain in addition to the spatial coordination shown in pervious work for magnitude. The implementation of stress affects both motor domains, indicating a tight connection. This close coupling is seen even in the absence of stress, though it is highlighted under stress. Although the precise patterns of inter-modal temporal effects differ between subjects, similar variability exists for the temporal implementation of stress in any single domain as well. The fact that, despite these differences, we see *some* cross modal effects for all subjects points to an inherent coordination between the two domains. These results are consistent with the idea that speech and nonspeech motor tasks can be parameterized into a single coordinative structure, which can be parameterized as a single functional unit. Additionally, this study supports the idea that implementation of prosody is not domain specific but relies on general aspects of the motor system.

References

Beckman, M., and Edwards, J. Articulatory evidence for differentiating stress categories. In P. Keating, editor, *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III,* pages 7-33. Cambridge: University Press, 1994.

Byrd, D., Krivokapić, J., and Lee, S. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America*, 120(3):1589-1599, 2006.

Kelso, J. A. S., Tuller, B., and Harris, K. S. A "dynamic pattern" perspective on the control and coordination of movement. In P. F. MacNeilage, editor, *The Production of Speech,* pages 138-73. New York: Springer-Verlag, 1983.

Levelt, W. J. M., Richardson, G., and Heij, W. L. Pointing and voicing in deictic expressions. *Journal of Memory and Language*, 24: 133-164, 1985.

Lœvenbruck, H., Dohen, M., and Vilain, C. Pointing is 'special'. In S. Fuchs, H. Lœvenbruck, D. Pape, and P. Perrier, editors, *Some Aspects of Speech and the Brain*, pages 211-258. New York: Peter Lang, 2009.

Nagasaki, H. Correlations of stress and timing in periodic tapping. *Human Movement Science*, 6:161-180, 1987.

Parrell, B., Goldstein, L., Lee, S., and Byrd, D. Articulatory evidence for functional coupling of speech and non-speech motor tasks. In *Proceedings of Speech Prosody V*, 2010.

Rochet-Capellan, A., Laboissière, R., Galván, A., and Schwartz, J. L. The speech focus position effect on jaw-finger coordination in a pointing task. *Journal of Speech and Hearing Research*, 51(6):1507-1521, 2008.

Smith, A., McFarland, D. H., and Weber, C. M. Interactions between speech and finger movements: An exploration on the dynamic pattern perspective. *Journal of Speech and Hearing Research*, 29(4):471-480, 1986.