Articulatory evidence for functional coupling of speech and non-speech motor tasks

*Benjamin Parrell, Louis Goldstein, Sungbok Lee, and Dani Byrd**

Department of Linguistics, University of Southern California
*corresponding author dbyrd@usc.edu

Abstract

Control of speech production is part of the larger motor control system, and as such can be organized into coordinative structures (or functional synergies) with other motor behaviors, that can then be parameterized as single functional units. The current study explores this phenomenon, expanding previous findings with direct kinematic evidence of speech production. Findings indicate that amplitude of repetitive synchronized speech and manual movements covary systematically across repetitions. In addition, magnitude of the movement of both of these effectors is larger when the participant is instructed to place emphasis on a repetition with one effector, but not the other. Thus, control of speech prosody may modulate a functional synergy that is defined over a wide set of articulatory subcomponents, not just the speech motor system.

Index Terms: coordination, coupling, motor control, speech

1. Introduction

Speech production is part of the larger human motor control system. Kelso, Tuller, & Harris [1] demonstrated consequences of this fact in a seminal study in which subjects were instructed to repetitively speak a word and tap their finger in time. When subjects stressed the spoken word, they produced larger finger movements despite instructions to maintain the finger tapping constant; subjects also showed a greater acoustic intensity on spoken syllables when stressing a finger tap, despite instructions to maintain the speech at a constant level. Kelso et al. took this to be evidence that speech articulation, like motor control in general, can enter into *functional coordinative structures* with other motor components. These findings were replicated and presented in more quantitative detail in later work [2,3]. Taken together, these studies imply that a) speech articulators are part of the general motor control system and b) control of speech functions in fundamentally the same way as control of other motor behaviors. These, and other similar findings that we control speech production in the same way as we control other actions has driven the development of Articulatory Phonology [4] and Task Dynamics [5], which apply the principles of self-organized, task-driven coordination [6,7] to the act of speech. This has proven a fruitful perspective and has been harnessed to explain such diverse phenomena as speech errors [8], overlap and coarticulation, allophonic variation [9], and speech prosody [e.g. 10-12]. However, because Kelso et al.'s 1983 study and subsequent replications were based only on the acoustic intensity of spoken syllables, it is not known if the effects can be seen in the behavior of the oral articulators forming constriction gestures. It is possible that the effects are limited to the laryngeal and respiratory [13] subsystems.

2. Methods

The current study examines the movement of both speech and non-speech articulators, rather than relying on speech acoustics as was done in the previous studies on this topic. This allows for a direct comparison between speech and non-speech actions, as well as the opportunity to examine the spatial, temporal, and coordination properties of the actions.

2.1. 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.

2.2. Procedure & Subjects

Four subjects (TA, TB, TC, TD) participated in the current study. Subjects (all right handed) were instructed to tap their right finger on their left shoulder while repeating a monosyllabic word in time with their finger taps when cued by the experimenter. Subjects were presented with a modified clock face with stars at the cardinal points (12:00, 3:00, 6:00 and 9:00) 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 production of finger tapping and speaking at the sweep of the second hand past a star when signaled by the experimenter and continue until the next star, i.e. for a 15 second interval. The subjects were told that when the second hand was at or near the halfway hash mark, they should either (in condition 1) make a single finger tap movement emphatic or (in condition 2) to place an emphatic stress on one repetition of the spoken syllable. In both cases, subjects were instructed to maintain the unemphasized action (tap or syllable) completely unchanged, continuing to repeat it at a constant, even rate. Ten repetitions of the task were collected per block. There were two blocks for each condition (emphatic tap or emphatic syllable), one using the syllable "ma" and one with the syllable "mop," for a total of four blocks and 40 repetitions. The order of conditions was varied by subject.

2.3. Data Analysis

In order to analyze the movement of the lips, a derived variable Euclidean Lip Aperture (LA) was calculated between the transducers placed on the upper and lower lips. For both

LA and finger 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 finger movements, the point of maximum constriction was measured when the finger was touching the shoulder, so the lowering of the finger towards that target constituted the first part of the movement and the raising of the finger, the second. The magnitude of the lip closing gesture for [m] was taken as the difference in position between the onset of the movement and the point of maximum constriction. Likewise, the magnitude of the opening gesture was the difference in position between the point of maximum constriction and the offset of movement. The finger tip movement measured was its lowering, i.e. the magnitude of movement from the gesture onset to the point of maximum constriction. Measurements are shown in Figure 1.

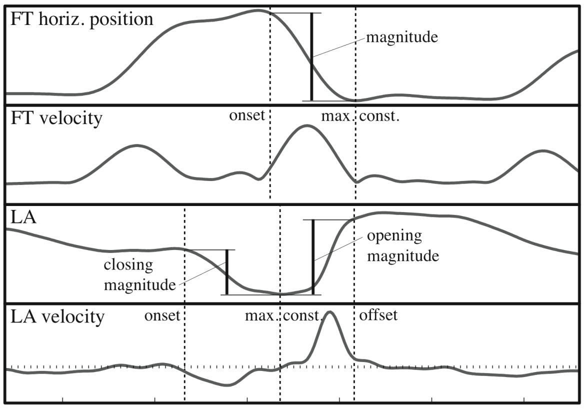

Figure 1: *Labeling and measurements made from data change to closing and opening.*

3. Results

While the purpose of the current study is to examine the possible coupling between speech and manual motor systems, it is necessary first to examine the data to ensure that our participants are performing the requested tasks and that our measurement techniques capture known intra-system effects. In order to do this, the magnitude of the movement of each articulator was examined separately and for each subject. For each articulator, a dataset was constructed to compare the magnitude of the repetition *when the subject explicitly emphasized that articulator* to the unstressed repetitions. Two data points were taken from each trial: the magnitude of the emphasized repetition and the mean magnitude of the unemphasized repetitions. Paired t-tests were conducted for each subject for a) LA closing magnitude, b) LA opening magnitude, and c) fingertip movement magnitude. These tests confirmed the emphasized repetition of both [m] and <tap> to be larger in magnitude than the unemphasized repetitions for almost all cases ($p < 0.02$). The exceptions were LA closing magnitude for subjects TA ($p < 0.07$), TB ($p < 0.80$), and TD ($p < 0.08$).

Using the same technique, tests were conducted for each articulator comparing the movement of that articulator when *the other was explicitly emphasized* to repetitions with no emphasis. For the fingertip, three subjects showed significantly larger movement when speech was stressed ($p < 0.01$, TC n.s.d.). Additionally, all subjects showed a larger magnitude of lip aperture movement when the finger tap was emphasized (though there were differences in which portion of the lip movement was effected [TA & TB, LA opening $p < 0.02$; TC & TD, LA closing $p < 0.01$]. No subject showed significant differences in both LA closing and opening). Recall that these effects occurred despite explicit instructions to the subjects to keep the non-emphasized action consistent and unchanged.

On visual observation, there seemed to be larger amplitude motions for both articulators toward the beginning and end of each trial. In order to control for this, the magnitude of the stressed repetition was compared against the magnitude of the immediately preceding and following unstressed repetitions. Three-way ANOVAs were conducted with this stress position, subjects, and intended stress (either on the finger movement or speech) as factors. As with the t-tests, separate tests were conducted for each articulator measurement (LA closing, LA opening, FT). For LA closing, there is a main effect of both subjects and intended stress ($F_{subjects}(3,392) = 43.31$, $p < 0.0001$; $F_{stress}(1,392) = 11.91$, $p < 0.001$) , but no effect of position ($F(2,392) = 0.3$, n.s.). For LA opening, all there is a main effect for all factors ($F_{subjects}(3,392) = 364.32$, $p < 0.0001$; $F_{stress}(1,392) = 411.49$, $p < 0.0001$; $F_{position}(3,392) = 184.06$, $p < 0.0001$); the same is true for FT ($F_{subjects}(3,392) = 41.58$, $p < 0.0001$; $F_{stress}(1,392) = 379.56$, $p < 0.0001$; $F_{position}(3,392) = 200.44$, $p < 0.0001$). Results for Tukey post-hoc tests conducted with an alpha value of 0.05 are presented in the tables below (* reflects a significant result). The intra-articulator stress results are presented in Table 1, and those for inter-articulator stress in Table 2.

Sub.	Meas.	$M_{stress} > M_{pre}, M_{post}$	M_{pre}	M_{stress}	M_{post}
TA	LA closing	n.s.	10.2	11.8	15.7
TA	LA opening	*	16.9	28.9	17.5
TA	FT	*	62.7	92.4	71.6
TB	LA closing	n.s.	10.7	9.5	7.9
TB	LA opening	*	13.4	17.1	12.6
TB	FT	*	44.4	105.3	61.5
TC	LA closing	n.s.	7.5	8.1	8.8
TC	LA opening	*	10.2	12.8	9.8
TC	FT	*	52.2	66.0	58.9
TD	LA closing	n.s.	5.2	7.1	4.9
TD	LA opening	*	9.5	15.4	9.8
TD	FT	*	53.3	93.8	56.6

Table 1. *ANOVA results for intra-articulator stress condition.*

LA closing is not significant for any subject. LA opening and FT movement significant for all subjects when that articulator is stressed, with the stressed repetition have a significantly greater magnitude than either unstressed repetition. Further, three of the four subjects also show effects of inter-articulator stress. For LA opening, TA shows a significant effect and subject TB and TD, while not significant, show trends in the same direction. For FT subjects TB and TD show significant effects and TA shows a similar trend. Despite instructions to make no change in the effector system without explicit stress, repetitions of that articulator have greater magnitudes when synchronous with a controlled emphasis in the other articulator. These effects are somewhat smaller in magnitude than the effects seen for intra-articulator stress.

Sub.	Meas.	M_{stress} > M_{pre}, M_{post}	M_{pre}	M_{stress}	M_{post}
TA	FT	n.s. (trend)	51.5	55.7	47.2
TA	LA closing	n.s.	10.4	9.8	8.7
TA	LA opening	*	14.5	15.6	13.1
TB	FT	*	31.2	40.4	31.0
TB	LA closing	n.s.	10.0	9.8	7.0
TB	LA opening	n.s. (trend)	10.9	11.8	10.2
TC	FT	n.s.	54.1	53.1	49.9
TC	LA closing	n.s.	6.8	6.4	6.6
TC	LA opening	n.s.	9.1	9.2	8.5
TD	FT	*	57.1	67.4	60.4
TD	LA closing	n.s.	5.6	6.1	6.7
TD	LA opening	n.s. (trend)	7.6	9.2	8.2

Table 2. *ANOVA results for inter-articulator stress condition.*

In order to test the global co-organization of the lip and fingertip movements in the absence of the addition of emphasis, a linear regression was calculated between the magnitudes of the two movements for only those repetitions where the subject did *not* place an explicit emphasis on either articulator. These regressions were calculated for LA closing and LA opening separately, with each measure independently regressed against FT movement. These results are presented in Table 3. The R^2 values are nearly always significant, though relatively low, ranging from 0.05 to 0.22. (Note that those trials with no significant difference are the same comparisons where t-tests found no significant differences due to inter-articulator stress).

Sub.	LA meas.	R^2	F	p <
TA	LA closing	0.0021	0.93	n.s.d
TA	LA opening	0.1079	53.02	0.0001
TB	LA closing	0.0078	4.64	0.05
TB	LA opening	0.1088	74.78	0.0001
TC	LA closing	0.2184	219.45	0.0001
TC	LA opening	0.0779	67.02	0.0001
TD	LA closing	0.0536	40.88	0.0001
TD	LA opening	0.0001	0.09	n.s.d.

Table 3. *Results of linear regression of LA magnitude (2 measures) and FT magnitude*

4. Discussion

As predicted, the results indicate the existence of coupling between the speech and manual motor control systems. All subjects show increased magnitude of the speech articulator movement contemporaneous with an emphasized finger movement despite instructions to maintain a constant and unchanging syllable production. Three of four subjects additionally showed the reverse effect in which spoken syllable emphasis cause larger movements in the simultaneous finger tap. These results were found regardless of whether the emphasized repetition was compared against all unstressed repetitions in the same trial or only those in its immediate vicinity. These results mirror those found in previous work [1-3] but for the first time demonstrate the effect on the *movement* of the speech articulators, whereas only acoustics was examined in the Kelso et al. work. Additionally, the data here further indicate that a correlation exists between manual and speech actions even independently of the effects of an instructed instance of emphasis. The fact that the magnitudes of the two actions are correlated in unstressed repetitions provides further evidence, not previously demonstrated, that these two systems are indeed entering into a functional coordination throughout this speech task.

There are two related, but distinct, interpretations of our findings that we plan to test in future analyses and experiments. One interpretation is that the synchronization task itself simply causes amplitudes of finger and lip movements to be coupled to one other. The other is that the functional task of prosody harnesses a broad set of body components, including those not normally considered as part of the speech system. This prosody hypothesis is a natural one in face of the fact that prosody is already known to recruit virtually all parts of the speech system (oral articulator [e.g., 10,14-16], larynx [e.g. 17], lungs [18]), and also the oro-facial musculature [e.g. 19,20]. If this hypothesis is correct, then the correlations seen during the non-emphasized portions of our trials can be seen as micro-prosodic fluctuations. One way of distinguishing the hypotheses would be to create amplitude variations in speech movement that are *not* associated with prosody. For example, the syllables /ma/ and /mi/ have different magnitudes of lip movement, even when produced in the same prosodic context. The prosody hypothesis predicts that such amplitude variations would not be reflected in finger movements, while the alternative (that amplitude effects results from synchronization per se) does predict an effect on finger movement.

5. Conclusions

The current study indicates that the control of speech prosody—or at least emphasis as one aspect of prosody—relies to a degree on basic control elements of the motor system. While stress in speech of course also relies on additional linguistic factors such as the control of fundamental frequency for pitch accent, the fact that manual emphasis effects the same qualitative alterations in speech articulation as language-specific stress indicates that the control mechanism underlying both phenomena is shared at some level.

6. Acknowledgements

Work supported by NIH DC03172.

7. References

[1] J.A.S. Kelso, B. Tuller and K.S. Harris, "A "Dynamic Pattern" perspective on the control and coordination of movement", *The Production of Speech*, P.F. MacNeilage, ed., New York: Springer-Verlag, 1983, p. 138-173.

[2] A. Smith, D.H. McFarland and C.M. Weber, "Interactions between speech and finger movements: an exploration on the dynamic pattern perspective", *J Speech Hear Res*, vol. 29, Dec. 1986, p. 471-80.

[3] P. Chang and G.R. Hammond, "Mutual interactions between speech and finger movements", *Journal of Motor Behavior*, vol. 19, 1987, p. 265-274.

[4] C.P. Browman and L. Goldstein, "Articulatory phonology: an overview", *Phonetica*, vol. 49, Jan. 1992, p. 155-80.

[5] E. Saltzman and K. Munhall, "A Dynamical Approach to Gestural Patterning in Speech Production", *Ecologoical Psychology*, vol. 1, 1989, p. 333-382.

[6] N. Bernstein, *The coordination and regulation of movements*, London: Pergamon, 1967.

[7] M.T. Turvey, "Coordination", *American Psychologist*, vol. 45, Jul. 1990, p. 938-953.

[8] L. Goldstein, M. Pouplier, L. Chen, E. Saltzman and D. Byrd, "Dynamic action units slip in speech production errors", *Cognition*, vol. 103, Jun. 2007, p. 386-412.

[9] R. Krakow, "Physiological organization of syllables: a review", *Journal of Phonetics*, Jan. 1999.

[10] D. Byrd and E. Saltzman, "Intragestural dynamics of multiple prosodic boundaries", *Journal of Phonetics*, vol. 26, Jan. 1998, p. 173-199.

[11] D. Byrd and E. Saltzman, "The elastic phrase: modeling the dynamics of boundary-adjacent lengthening", *Journal of Phonetics*, vol. 31, Jan. 2003, p. 149-180.

[12] D. Byrd, S. Tobin, E. Bresch and S. Narayanan, "Timing effects of syllable structure and stress on nasals: A real-time MRI examination", *Journal of Phonetics*, vol. 37, Jan. 2009, p. 97-110.

[13] P. Ladefoged, *A course in phonetics*, Boston: Thomson Wadsworth, 2006.

[14] M.E. Beckman and J. Edwards, "Intonational categories and the articulatory control of duration", *Speech perception, production, and linguistic structure*, Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka, ed., Tokyo: Ohmsha, Ltd., 1992, p. 359-375.

[15] J. Edwards, M. Beckman and J. Fletcher, "The articulatory kinematics of final lengthening", *Journal of the Acoustical Society of America*, vol. 89, 1991, p. 369-382.

[16] C. Fougeron and P.A. Keating, "Articulatory strengthening at edges of prosodic domains", *Journal of the Acoustical Society of America*, vol. 101, Jun. 1997, p. 3728-40.

[17] T. Cho and S.A. Jun, "Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean", *UCLA Working Papers in Phonetics*, 2000, p. 57-70.

[18] P. Ladefoged, *Three areas of experimental phonetics*, Oxford: Oxcord University Press, 1967.

[19] M.D. McClean and S.M. Tasko, "Association of orofacial with laryngeal and respiratory motor output during speech", *Exp Brain Res*, vol. 146, Oct. 2002, p. 481-9.

[20] M.D. McClean and S.M. Tasko, "Association of orofacial muscle activity and movement during changes in speech rate and intensity", *J Speech Lang Hear Res*, vol. 46, Dec. 2003, p. 1387-400.