Journal of Phonetics 42 (2014) 1–11

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Spatiotemporal coupling between speech and manual motor actions

Benjamin Parrell [a,*], Louis Goldstein [a,b], Sungbok Lee [a], Dani Byrd [a]

[a] University of Southern California, Los Angeles, CA 90089, USA
[b] Haskins Laboratories, New Haven, CT 06511, USA

ARTICLE INFO

Article history:
Received 20 November 2012
Received in revised form
5 November 2013
Accepted 7 November 2013

ABSTRACT

Much evidence has been found for pervasive links between the manual and speech motor systems, including evidence from infant development, deictic pointing, and repetitive tapping and speaking tasks. We expand on the last of these paradigms to look at intra- and cross-modal effects of emphatic stress, as well as the effects of coordination in the absence of explicit rhythm. In this study, subjects repeatedly tapped their finger and synchronously repeated a single spoken syllable. On each trial, subjects placed an emphatic stress on one finger tap or one spoken syllable. Results show that both movement duration and magnitude are affected by emphatic stress regardless of whether that stress is in the same domain (e.g., effects on the oral articulators when a spoken repetition is stressed) or across domains (e.g., effects on the oral articulators when a tap is stressed). Though the size of the effects differs between intra-and cross-domain emphases, 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. The results of this study support the idea that implementation of prosody is not domain-specific but relies on general aspects of the motor system.

1. Introduction

A large body of research has begun to investigate the links between spoken language and other parts of the human motor system, particularly manual gesture. This link is pervasive and has been demonstrated in many domains. For example, Iverson and Thelen (1999) proposed that speech and manual gesture are coordinated in a dynamical system, where the individual subparts combine to form functional groups that can be parameterized at the level of control as a single entity (e.g., Turvey, 1990). They propose that the link between the two motor subsystems is present at birth in a basic form and develops over time into the complicated, multifaceted coordination seen in adults. Infants spontaneously coproduce hand and mouth movements, opening their mouth as the hand moves towards the face (Butterworth & Hopkins, 1988). This coupling leads, later, to coordination between babbling and rhythmic hand shaking. Infants produce high rates of rhythmic behavior in the upper limbs immediately prior to the onset of canonical babbling and, after babbling begins, the two tend to co-occur (Ejiri, 1998; Iverson & Fagan, 2004; Iverson, Hall, Nickel, & Wozniak, 2007). Moreover, when babbling does co-occur with upper limb oscillations, it is more adult-like, with shorter durations for both syllables as a whole and the formant-frequency transitions between the syllable onset and vowel than when babbling is produced in isolation (Ejiri & Masataka, 2001). Vocabulary spurts in infants also indicate a temporal coordination of communicative behaviors in multiple modalities (Parladé & Iverson, 2011; see also Yale, Messinger, Cobo-Lewis, & Delgado, 2003; Yale, Messinger, Cobo-Lewis, Oller, & Eilers, 1999).

This extensive coordination between manual and speech motor systems is pervasive throughout the lifespan. In one study, subjects moved either an apple or a cherry towards their mouth while simultaneously producing the syllable /ba/ (Gentilucci, Santunione, Roy, & Stefanini, 2004). Subjects produced a larger lip opening and a higher F2 for /a/ when grasping an apple rather than a cherry; the same effect was found even when subjects merely observed these actions being performed. Adults also coordinate speech with gestures more generally. McNeill (1992) writes that gestures are co-produced with their semantically co-expressive word. For example, 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). This coordination is quite complex, however. In one study, subjects had to point at a smiley face while naming it with either a ‘CVCV or CV’CV word, the two words varying in the position of stress. Results show that the gesture was not aligned with the word as a whole, but with the stressed syllables (Rochet-Capellan, Laboissière, Galván, & Schwartz, 2008). There is some evidence that this coordination between prosodic emphasis and pointing gestures may be mediated by shared neural circuitry between the two domains, as production of both (along with looking at the focused object) activates the left superior parietal lobe, while using syntax to emphasize the same object does not activate this region (Loevenbruck, Dohen, & Vilain, 2009).

* Corresponding author: Tel.: +1 213 740 2986; fax: +1 213 740 9306.
E-mail address: parrell@usc.edu (B. Parrell).

0095-4470/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2013.11.002

One productive way in which the link between speech and manual gesture has been explored in the laboratory with adults is via the examination of repetitive, synchronous speech and finger tapping. This line of inquiry was initiated in work by Kelso, Tuller, and Harris (1983). In this study, subjects were instructed to speak a monosyllabic word repetitively ("stock") and tap their finger in time with their speech. Subjects either spoke or tapped their right hand in an alternating stressed-unstressed pattern and were instructed to keep constant their other production (i.e., tapping unchanged during spoken stress alternation, speaking unchanged during tapped stress alternation). Despite these instructions, subjects consistently produced larger taps when 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-occurred with a stressed tap than an unstressed tap. Kelso and colleagues interpreted these acoustic results as evidence that the speech and finger movements were entrained as a single coordinative structure.

The Kelso et al. (1983) results were replicated by Chang and Hammond (1987), which also showed that the entrainment existed when the tapping was performed with both the right and left hands, indicating the cross-modal effects were not due to anatomical overlap between the motor areas for speech and right-handed movements. This study also examined the temporal coordination between the two systems, finding the initiation of finger taps lagged the acoustic onset of the co-produced word by 30–50 ms. Smith, McFarland, and Weber (1986) extended this paradigm by examining concurrent tapping and speaking not only synchronously but also at differing rates of production. They found more complicated results than in the previous two studies. While all subjects showed the expected positive correlations between the amplitude of speech intensity and finger motion under alternating spoken stress, only half showed this pattern when the finger tap amplitude was alternated. Two of six showed a negative correlation, and one showed no significant correlation at all. When speech and fingertip movements were produced at the same rate, the two systems were tightly aligned temporally (onset of a movement in one domain fell within 1/4 cycle of onset of a movement in the other). However, these movements were produced more asynchronously when they were produced at different rates (i.e. subjects did not produce simple harmonic frequency ratios), indicating a lack of absolute coordination between the two systems but rather weak coupling.

Treffner and Peter (2002) examined both synchronous and alternating production of speech (the syllable /ba/) and finger taps in a task where the speed of production increased during each trial. This increase triggered a switch from anti-phase (alternating) to in-phase (synchronous) productions of finger taps and spoken syllables, consistent with similar experiments on bimanual tapping and limb oscillation (Haken, Kelso, & Bunz, 1985; Turvey, 1990). They also found that, generally speaking, initiation of the finger tap motion preceded jaw opening for in-phase trials, while the jaw motion preceded the finger tap for the anti-phase trials. While this at first seems at odds with the results of Chang and Hammond, it is important to remember that Treffner and Peter used jaw motion as an index of speech motor activity, while Chang and Hammond used the onset of the acoustic amplitude rise for the syllable-initial consonant (/s/), which most likely would precede jaw opening (which is generally associated with the vowel). In fact, Treffner and Peter posit that the finger tap in these types of studies is coordinated not with any particular vocal or physiological event but rather with the perceptual center (p-center) of the syllable. The p-center is the point in a syllable which subjects use to align that syllable to a metronome rhythm (Morton, Marcus, & Frankish, 1976). The p-center is generally located near the onset of the vowel constituting the syllable nucleus (Fowler, 1983), though it does not seem to be tied to a particular kinematic event (de Jong, 1994). Inui (2007) also examined anti-phase coordination between speech and finger tapping and found shorter lags when speech preceded tapping (i.e. the subjects began with a spoken syllable and followed that with a finger tap) than the reverse, which the author took to indicate a tighter coupling in the former case. This is similar to the result found in Smith et al. (1986), where a more consistent entrainment was found between the two domains when stress was consciously placed on spoken rather than tapped repetitions.

It is not clear, however, how the results of these repetitive production tasks generalize to the interaction between real-world speech and manual actions. First, past experiments (Chang & Hammond, 1987; Kelso et al., 1983; Smith et al., 1986) have generally imposed a rhythmical alternation between stressed and unstressed taps or spoken syllables. While it has been suggested that some languages, including English, show a similar regular pattern of isochronous inter-stress intervals (e.g. Abercrombie, 1967), such proposals have received little empirical support (for a review, see Arvaniti, 2012). On the other hand, it has been well established that both prosodic structure and emphatic stress have large effects on the spatial and temporal production of speech gestures and the coordination patterns between those gestures. Prosodic boundaries both lengthen and increase the spatial magnitude of speech gestures local to them (for a review see Byrd, Krivokapić, & Lee, 2006) and similar effects are seen under emphatic stress/accent (e.g., Beckman & Edwards, 1994). Additionally, the temporal lag between speech gestures often increases near a prosodic boundary (e.g., Byrd, Kaun, Narayanan, & Saltzman, 2000; ; Hardcastle, 1985; Keating, Cho, Fougeron, and Hsu, 2003). Furthermore, the imposition of an alternating stress pattern also confounds results showing covariation of amplitude across speech and manual motor domains; it may be the case that these amplitude correlations are caused by this imposed rhythm, rather than through an intrinsic aspect of their coordination.

The current study presents a first step to understanding how the prosodic structure of speech affects the multimodal coordination of speech and manual gesture. Following previous studies, we continue to use synchronous, repetitive productions of monosyllabic words and finger taps. However, here we move beyond examining the effects of imposed rhythm. While a repetitive sequence of unstressed repetitions is obviously regular, it is not *rhythmic* in the linguistic sense, that is "an ordered recurrence of strong and weak elements" (Fowler, 1983, p. 386). Rather than examining how an ongoing (alternating) rhythmic pattern affects speech and manual gesture, we examine the spatial, temporal, and coordinative effects of *emphatic* stress. By instructing subjects to stress a single spoken or tapped repetition, we can elicit an quasi-linguistic emphatic stress similar to sentence-level stress, or accent, such as would be used to distinguish the phrases "I said I saw the CAT" and "I said I SAW the cat." This type of stress has been shown to cause speech gestures around the stress to increase in magnitude and, more variably, to lengthen (Beckman & Edwards, 1994; de Jong, 1995). The use of a single emphatic stress with repetitive syllables and taps allows for a comparison of the effects of language-like prosodic structure with previous results from repetitive rhythmic tasks. Additionally, this paradigm provides a method for examining the spontaneous spatiotemporal effects of coupling between the speech and manual motor systems. The use of a quasi-linguistic emphatic stress provides a first step towards connecting these types of repetitive tasks with the much more complex relationship between speech and other motor systems that occurs in natural settings.

2. Experimental methods

2.1. Procedure and subjects

Four male, right-handed subjects (TA, TB, TC, TD) participated in the current study, and were paid for their participation. Subjects' ages ranged from 19 to 29. Subjects 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 (the normal locations of 12:00,

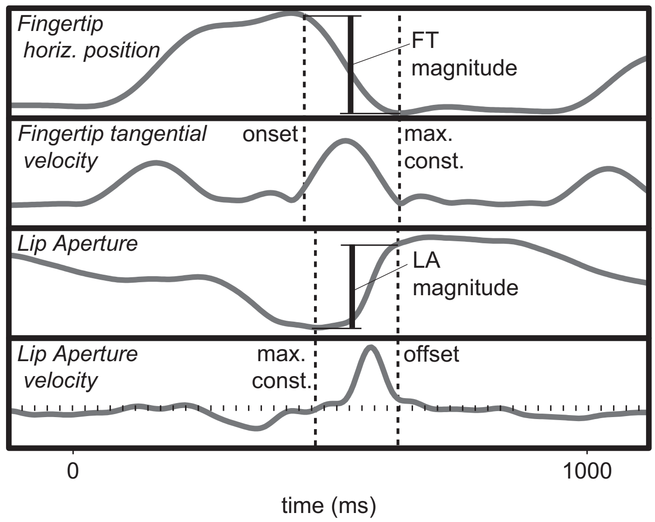

Fig. 1. Schematic of measurements taken from articulatory data for lip aperture and fingertip magnitude. All measurements are based on the zero-crossings in the velocity signal. Fingertip movements are measured from the onset of movement toward the shoulder to the point of maximum constriction. Lip movements measured as the movement from the /m/ to the following vowel.

3:00, 6:00 and 9:00) and with hash marks halfway between each star; a second hand swept around the clock face 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 s 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 emphatically 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. No explicit instruction was given as to whether the task should be completed on a single breath cycle or not. Ten repetitions of the task were collected per block. There were two blocks for each condition (emphatic tap or emphatic syllable), one using the monosyllable word "ma" and one with the monosyllable word "mop," for a total of four blocks and 40 repetitions. The order of conditions was counterbalanced across subjects.

2.2. Kinematic data collection

Kinematic articulator data were collected using an electromagnetic articulometer (Carstens AG500). This device allows three-dimensional tracking of transducers attached to the articulators. For this study, transducers were glued to the upper and lower lips, and the tip of the right index finger. Reference sensors were attached to the nose ridge and behind each ear. Articulatory data were collected at 200 Hz, and acoustic data at 16 kHz. After collection, the articulatory data were 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 processing

Lip activity for forming the constriction for /m/ was measured using lip aperture (LA), defined as the Euclidian distance between sensors placed on the upper and lower lips. For both LA and fingertip (FT) movement, movement speeds derived from the velocity time functions (tangential velocity in the case of the finger tapping data) were used to identify the events associated with individual /m/s and individual taps: the point of speed minimum (point of maximum constriction), the peak speeds both before and after the speed 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 or following minima, respectively). FT movements were defined as the lowering of the finger towards the shoulder, measured from the start of that movement until the point when the finger was touching the shoulder (identified as a minimum in the tangential velocity signal in the same way as the lip closing movement).

The magnitude of the articulatory lip opening gesture for [m] was taken as the difference in LA between the point of maximum constriction and the offset of lip opening movement. The fingertip lowering magnitude was measured, i.e. the magnitude of movement from the gesture onset to the point of maximum constriction (Fig. 1).

To examine the temporal effects of emphatic stress the time between successive repetitions, i.e. the inter-response interval (IRI), was used. This measure has been used in many studies to measure the timing variability in rhythmic tasks produced at rates similar to those in the current study, both for tapping (e.g. Ivry & Keele, 1989; Max & Yudman, 2003; Nagasaki, 1987; Wing & Kristofferson, 1973) and for speech (e.g. Hulstijn, Summers, van Lieshout, and Peters, 1992; Max & Yudman, 2003). Here, IRI was measured as the time between the onset of one repetition and the onset of the following repetition, for both LA and FT movements (Fig. 2). For LA, the onset of the lip opening movement was used. Results presented here measured the onset of the both lip aperture and finger tap movements using the same velocity-threshold algorithm[1] as above.

The stressed repetition was the repetition that showed a larger movement magnitude in the domain where stress was instructed to be placed (e.g., a larger lip movement for trials with instructed spoken stress). On a few trials, no stress could be identified during analysis. These trials were excluded from further analysis. Out of 160 trials total, 9 trials were excluded for subject TA, 1 for TB, 0 for TC, and 10 for TD. All of the excluded trials for subject

[1] As an assurance that these measurements accurately indexed the movement onsets, IRIs were also calculated based on the time of peak velocity of the each movement; these results were not substantially different from those for the velocity-threshold method and as such are not reported upon further.

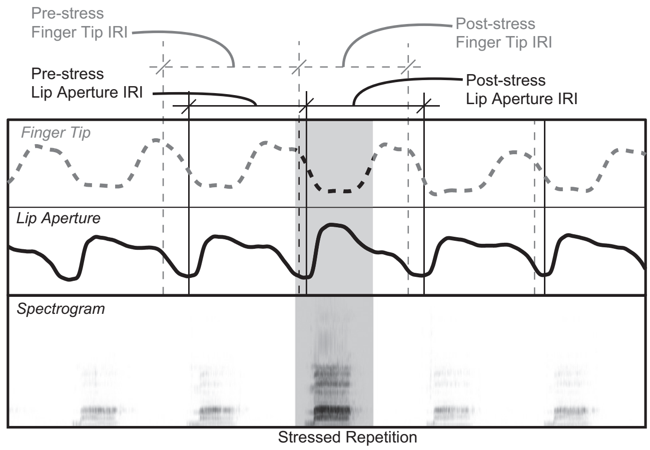

Fig. 2. Example of measuring inter-response intervals (IRIs). IRIs are defined separately for lip aperture and fingertip movements. Fingertip IRIs are defined as the period from the onset of one fingertip lowering movement to the next. Lip aperture IRIs are defined as the period from the onset of one lip opening movement to the next. The pre-stress IRI is the IRI immediately preceding an emphatic stress in either domain; the post-stress IRI is that which includes the repetition with emphatic stress.

TD came from one block of trials with spoken stress, where the subject did not perform the task correctly. In general, for all subjects stress was produced as instructed. On visual observation of the data, there seemed to be larger amplitude motions for both articulators toward the beginning and end of each trial. To control for possible edge effects, the first and last two repetitions were excluded from all analyses.

In order to test whether the presence of a coda in the stimuli ("mop" versus "ma") had an effect on the duration or magnitude of articulator movements, two separate repeated measures ANOVAs were conducted with movement magnitude and duration of all the repetitions in the dataset as the dependent variables and the stimulus word and articulator (FT, LA) as factors. For both words' magnitude, the only significant factor is the articulator ($F(1,3) = 103.37$, $p < 0.001$), with the fingertip movements ($M = 62$ mm) being much larger than lip opening ($M = 14$ mm). Neither the stimulus word nor the interaction between the factors is significant. For duration, neither factor nor their interaction is significant. Based on these results, we conclude that the stimulus word used affected neither the duration nor the magnitude of the movements in this study and, as such, do not include it as a factor in further analyses.

2.4. Data analysis

As prosodic structure and emphatic stress in speech have spatial, temporal, and coordinative effects on speech gestures, all three aspects are examined in the current study. Because subjects produced speech and finger taps at a self-selected rate, we also examine the rate of production for each subject to make sure that these rates are similar between the subjects of the current study and to the rates reported in previous work. This is important, as rate has been shown to influence both the spatial and temporal aspects of speech movements as well as the effects of prosodic structure and emphatic stress. For all the statistical tests detailed below, an α value of 0.05 was used. Post-hoc tests were conducted with a Bonferroni adjustment to retain an overall α of 0.05 for that particular analysis.

We first examine the effects of emphatic stress on the spatial domain. In order to do this, the magnitude of the movements of each articulator was examined separately for each subject. For each articulator (LA or FT), a dataset was constructed to compare the magnitude in repetitions with instructed-emphasis to repetitions in which that articulator was unstressed, from both the domain in which the emphatic stress was placed (e.g., fingertip data when the finger tap received emphatic stress) and across domains. Two data points were examined for each trial: the magnitude of the emphasized repetition and the mean magnitude of the unemphasized repetitions from that trial. A repeated-measures ANOVA was conducted with subject included as a random factor, the magnitude of the articulator movements as the dependent variable and stress level (under emphatic stress vs. unstressed), articulator (FT or LA), and stress domain (within-domain stress vs. across-domain stress) as fixed factors.

The above method provides a global comparison of repetitions with emphatic stress to the unstressed repetitions. In order to get a more detailed picture of the local effects around the stressed repetition, the magnitude of the stressed repetition was compared against the magnitude of the immediately preceding and following unstressed repetitions. A repeated measures ANOVA was conducted with subject as a random factor and stress position (pre-stress, stressed, or post-stress), articulator (FT or LA), and stress domain (within- or across-domain stress) as factors, as was done for the global tests.

In order to test the global co-organization of the lip and finger 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.

To examine the temporal effects of emphatic stress, the IRIs before and after the emphatic stress were measured. Note that since the IRI is measured from movement onset, the IRI following the stress includes the stressed repetition itself (see Fig. 2). These were compared against the mean of the unstressed repetitions for that trial. These means excluded the first and last two IRIs in order to prevent interference from possible effects of movement initiation and termination. A repeated-measures ANOVA was conducted with the factors stress position (pre-stress, post-stress, or unstressed), articulator (FT or LA IRI), and stress domain (within- or across-domain stress). Additionally, correlations between concurrent IRIs for FT and LA were calculated across all repetitions for each subject to examine the temporal coordination in the absence of emphatic stress.

To examine the effects of emphatic stress on the temporal coordination between lip and fingertip movements, we measured the lag between these two gestures. Intergestural lag has commonly been used to measure the coordination between two articulators in speech research and other domains (e.g. Baader, Kazennikov, & Wiesendanger, 2005; Carey, 2000; Goldstein, Chitoran, & Selkirk, 2007; Gracco & Löfqvist, 1994; Munhall, Löfqvist, & Kelso, 1994), particularly in the relevant studies on speech and manual coordination (e.g. Chang & Hammond, 1987, Hulstijn et al., 1992, Inui, 2007). To facilitate comparison with this previous work and speech and manual motor interaction, we chose to calculate intergestural lag as the time between the onset of the opening movement for LA and the onset of the FT movement. We also calculated the relative phase between the articulators as a

Table 1
Mean of trial durations, number of repetitions per trial, and repetition duration for all subjects. Subjects produced around 15–21 repetitions per trial, with average repetition durations of roughly 500–800 ms

	Trial duration (s)			# of repetitions			Repetition duration (ms)		
	M	Min	Max	*M*	Min	Max	*M*	Min	Max
TA	9.8 (0.8)	8.5	12.6	15.2 (1.2)	13	19	644 (19)	615	670
TB	13.4 (0.6)	11.9	14.4	16.8 (1.3)	14	19	780 (50)	709	887
TC	10.5 (1.1)	8.5	12.9	21.0 (1.8)	18	26	501 (27)	460	588
TD	9.8 (0.9)	7.2	11.5	19.2 (1.9)	15	23	514 (40)	452	675

percentage of the overall lip aperture cycle by dividing our lag measure by the time between onset of one lip opening and the onset of the next lip opening movement (cf. Saltzman, Löfqvist, Kay, Kinsella-Shaw, & Rubin, 1998); these results were not different from those resulting from the intergestural lag measure and so are not discussed in more detail. The measure of intergestural lag gives a positive value when the fingertip movement is initiated first, and a negative value if the lip movement leads. These particular points were chosen in order to facilitate comparison of the current results to past work that has used the onset of jaw lowering as the index for speech (e.g., Treffner & Peter, 2002). In order to test whether explicit stress had an effect on coordination, the lag for the stressed repetition of each trial was tested against the mean of the unstressed repetition for the same trial. The two initial and final repetitions from each trial were excluded from this mean (recognizing the possibility of boundary effects). These data were then submitted to a two-way repeated measures ANOVA with stress domain (spoken or tapped) and position (stressed or unstressed) as fixed factors.

3. Results

3.1. Rate of production

While each trial had an available window of duration of 15 s, no subject produced speech and/or tapping during the entire target interval.[2] Total production duration ranged from 7.2 to 14.4 s, with an average of 10.8 s. Subjects produced between 13 and 26 repetitions in that period, averaging 18.2 s. The average repetition duration, defined as the total trial duration divided by the number of repetitions in that trial, was calculated separately for each trial, then a grand mean was calculated; this gave an overall average repetition duration of 590 ms. Results, separated by subject, are presented in Table 1. Subject TC and TD showed relatively short average repetition durations of around 500 ms, while the others showed somewhat longer durations. Subject TB, in particular, showed an average duration of over 750 ms. These durations are consistent with evidence that people have a preferred rate of production in rhythmic tasks of around 2 Hz, with a range of individual variation (e.g. Allen, 1975; Miles, 1937; Woodrow, 1951), and are comparable to the spontaneous, comfortable rates produced by subjects in previous studies on hand/speech rhythmic coordination (Chang & Hammond, 1987; Kelso et al., 1983; Smith et al., 1986).

3.2. Spatial effects

We first examine the global comparison of repetitions under emphatic stress to the mean of all unstressed repetitions. There were main effects of all three fixed factors: stress level ($F(1,3)=13.46$, $p<0.05$), articulator ($F(1,3)=249.40$, $p<0.001$), and stress domain ($F(1,3)=11.57$, $p<0.05$), as well as a significant interaction between stress level and stress domain ($F(1,3)=27.56$, $p<0.05$). Tukey post-hoc tests (adjusted $\alpha=0.0035$) showed that, in the intra-modal stress condition, repetitions with emphatic stress were larger than unstressed repetitions for both finger tap and lip opening movements for all subjects. This result is unsurprising, as the stressed repetition was defined as that with a larger magnitude than the surrounding repetitions (see Section 2.3). Fingertip movements were also larger than lip movements. As can be seen in Fig. 3, all subjects exhibited this same pattern. All subjects also showed a larger difference in magnitude between stressed and unstressed repetitions for the fingertip movement compared to lip movements.

Cross-modal spatial effects of stress are shown in Fig. 4. For three out of the four subjects (TA, TB, TD), fingertip movements were larger when speech is stressed. One of the subjects also showed larger magnitudes of lip movements when finger taps are stressed (TB). Overall, the spatial effects of emphatic stress are much smaller for cross-modal than for intra-modal stress, which is the source of the stress level by stress domain interaction.

The results are broadly similar when we look only at the local effects of emphatic stress. When stress is in the same domain as the articulator movement, there are main effects of stress position ($F(2,6)=13.79$, $p<0.01$) and articulator ($F(1,3)=238.48$, $p<0.001$); two-way interaction effects between stress position and articulator ($F(2,6)=7.06$, $p<0.05$) and between stress position and stress domain ($F(2,6)=14.40$, $p<0.01$); and a three-way interaction between all factors ($F(2,6)=6.51$, $p<0.05$). Results are shown by subject in Fig. 5. As can be seen in that figure, the stressed repetition is generally larger than both pre-and post-stress repetitions. The lip movements are much smaller than the fingertip movements, and the cross-modal effect of emphatic stress on movement magnitude is smaller than the intra-modal effect. All subjects show a significant effect of emphatic stress in the same domain as the articulator movement with the stressed repetition being larger than pre- and post-stress positions. In the cross-domain condition, the effects are smaller (the source of the interactions with stress domain). On an individual level, post-hoc tests comparing the stressed to the pre-and post-stress positions (adjusted $\alpha=0.0035$) demonstrate that three of the subjects (TA, TB, TD) show significantly larger magnitudes of the stressed FT repetition, and two subjects (TB, TD) show significantly larger magnitudes for the same position for LA. The one subject who shows no cross-articulator effects, TC, is also the one whose intra-articulator effects are smallest in magnitude; that subject appears to only weakly emphasize the selected repetition.

[2] Though there were no explicit instructions to produce the 15 s of repetitive speech in one continuous expiration, all of the subjects did so. This is most likely the reason why no subject produced a full 15 s of speech and tapping.

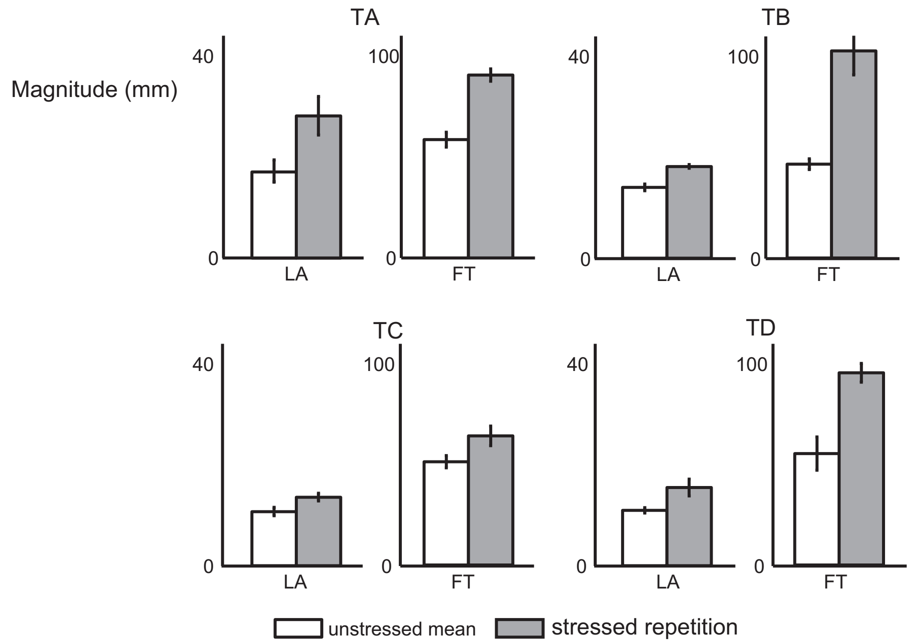

Fig. 3. Comparison of the magnitude of repetitions with emphatic stress against the mean magnitude of unstressed repetitions when stress occurs in the same domain as the movement. Means and standard deviations shown. All subjects show larger movements of both lip aperture and fingertip under stress.

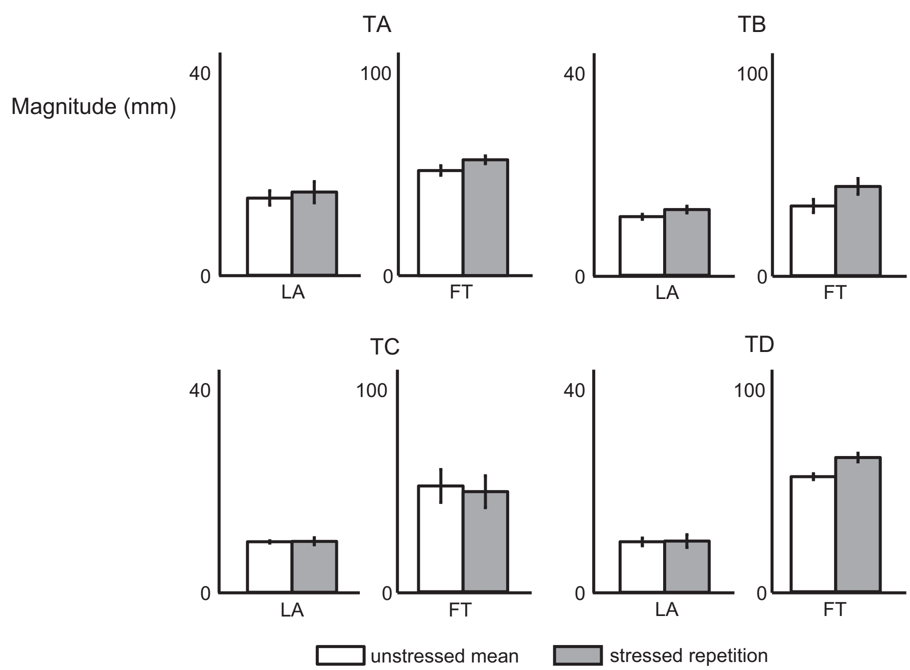

Fig. 4. Comparison of the magnitude of repetitions with emphatic stress against the mean magnitude of unstressed repetitions when stress occurs in the opposite domain as the movement. Means and standard deviations shown. There is an overall effect of larger movements under cross-domain stress, but subjects differ slightly in this effect.

In sum, despite instructions to make no change in the effector system having no instructed explicit stress, repetitions of that articulator do in fact have greater magnitudes when concurrent with an active emphasis in the other effector system. These effects, while present in three of four subjects, are smaller in magnitude (roughly 25%) than the effects seen for intra-articulator stress. (The relatively small magnitude of the cross-articulator effects combined with the small magnitude of subject TC's intra-articulator effects (see Figure 5) could lead to any cross-articulator effect for that subject being lost in noise.)

With regard to the spatial covariation between *unstressed* repetitions, three out of four subjects show a significant correlation between the magnitude of LA and FT movements (TA: $r^2=0.11$, $p<0.0001$; TB: $r^2=0.11$, $p<0.0001$; TC: $r^2=0.08$, $p<0.0001$; TD: n.s.).

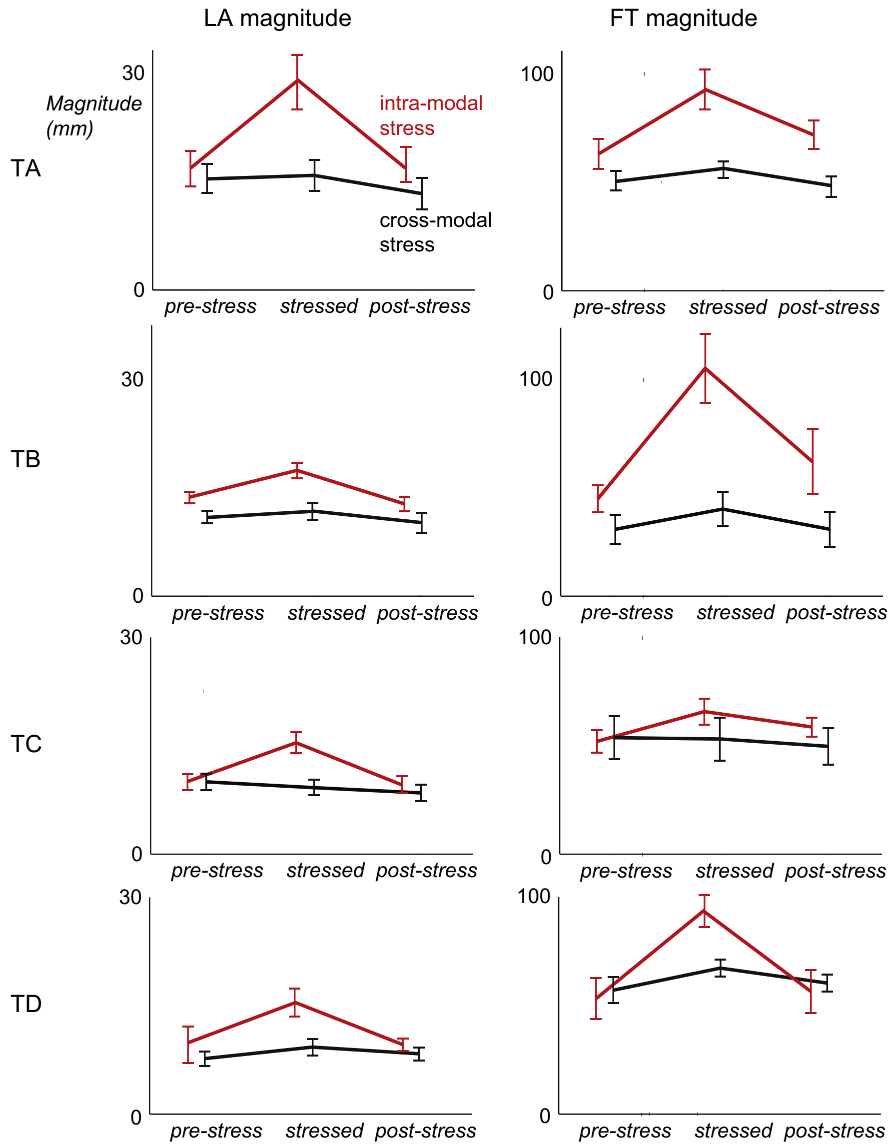

Fig. 5. Comparison of magnitude of pre-stress, stressed, and post-stress repetitions of lip and fingertip movements. Means and standard deviations shown. All subjects show larger movements of stressed repetition when stress is in the same domain as the movement. All subjects except TC also show larger movements of the stressed repetition when stress occurs in the other domain.

3.3. *Temporal effects*

As for the effects of emphatic stress on IRI, there is a significant effect of stress position ($F(2,6)=7.75$, $p<0.05$) and no main effect of articulator or stress domain. There is a significant interaction between stress position and stress domain ($F(2,6)=14.51$, $p<0.01$). While only a few of the individual post-hoc tests reached significance (adjusted $\alpha=0.0035$), the main effect of stress position across subjects can be seen in Fig. 6, where pre-stress IRIs generally lengthened under intramodal emphatic stress, and returned to the mean duration in the post-stress IRI.

A different pattern emerges under cross-modal emphatic stress, however. Unlike the consistent spatial effects found across subjects for intra-modal stress, subjects show somewhat different patterns in the effects of cross-modal emphatic stress on IRI. For fingertip movements under spoken stress, two subjects (TA, TB) show lengthening of the post-stress IRI only, and one subject (TC) shows lengthening of the pre-stress IRI. Similarly, when lip movements are affected by fingertip stress, two subjects (TA, TC) show lengthening of the post-stress IRI and one (TD) shows lengthening of the pre-stress IRI. The effect of cross-modal stress on IRI is somewhat different from that of magnitude: where magnitude showed concurrent, though reduced, effects cross-modally, the duration results indicate similar magnitude effects for intra- versus cross-articulator comparisons, though the effect appears to often be delayed in the domain lacking explicit stress: the intra-articulator effect is larger in the pre-stress repetition while the cross-articulator effect is larger in the post-stress repetition (hence the interaction between position and stress domain).

As for magnitude, it is important to know whether the temporal variation caused under emphatic stress is caused solely by the emphatic stress itself or whether it is a general property of the coordination between the two domains. To test this, correlations between concurrent IRIs for FT and LA were calculated across all repetitions for each subject. All subjects showed significant positive correlations between the two measures (TA: $r^2=0.07$, $p<0.0001$; TB: $r^2=0.29$, $p<0.0001$; TC: $r^2=0.32$, $p<0.0001$; TD: $r^2=0.23$, $p<0.0001$). These results are comparable with the correlations in magnitude between the two domains (Section 3.2). These results indicate that the temporal coordination between speech and finger taps in this task were in fact prevalent throughout the task, and not an epiphenomenal effect of explicit stress.

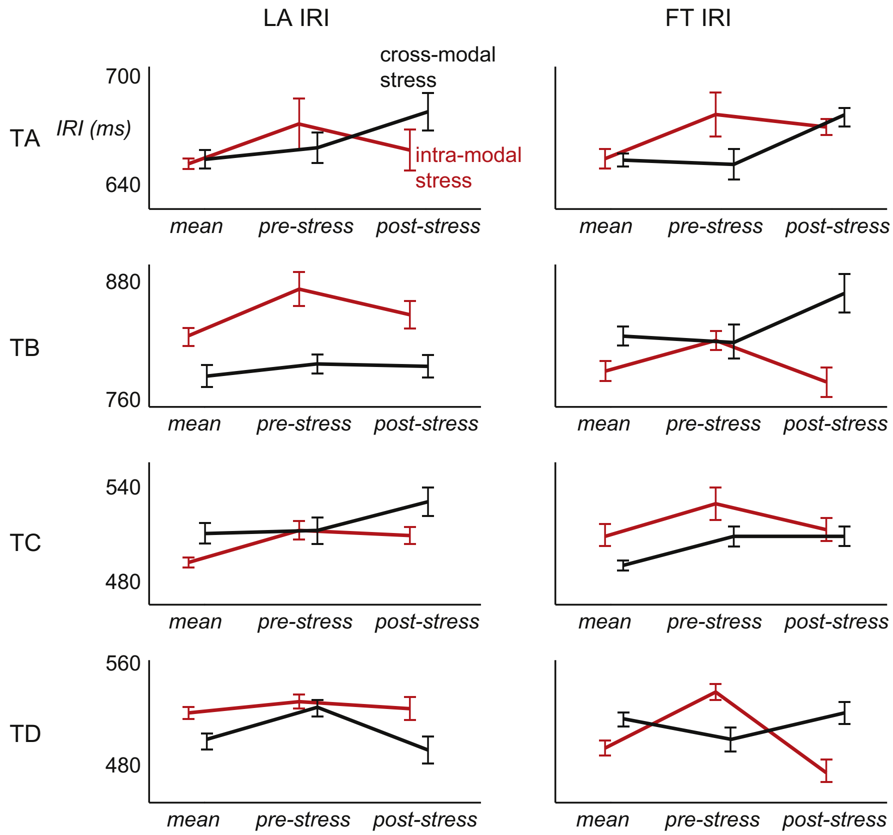

Fig. 6. Durational patterns of pre- and post-stress IRIs compared to the mean IRI in positions not adjacent to the emphatic stress. Means and standard deviations shown. All subjects show lengthening of IRIs in stress-adjacent positions in both domains regardless of in which domain the emphatic stress is placed. Subjects differ on the particular implementation of this lengthening.

3.4. Coordination effects

Overall, fingertip movement preceded lip opening by an average of 57 ms. These results are consistent with previous findings that measured lag in a similar manner (Treffner & Peter, 2002; Hulstijn et al., 1992). There is a significant effect of stress position ($F(1,3)=11.68$, $p<0.05$), but no effect of stress domain nor an interaction between the two. For three out of four subjects (TB, TC, TD), the lag was shorter in stressed compared to unstressed repetitions, indicating a more synchronous production. Results by subject are presented in Fig. 7.

4. Discussion

The results indicate, as predicted, the existence of coupling between the speech and manual motor control systems. These findings mirror those found in previous work on repetitive speech and tapping with alternating stress (Chang & Hammond, 1987; Kelso et al., 1983; Smith et al., 1986) but are novel in two important ways: they demonstrate the effects of emphatic stress, similar to that found in spoken language, rather than regular rhythmic alternation, and they show effects in the temporal domain, as well as spatial.

All subjects show increased *magnitude* of the speech articulator movement concurrent with an emphasized finger movement despite instructions to maintain a constant and unchanging syllable production. Additionally, three of four subjects showed the reverse effect in which spoken syllable emphasis caused larger movements in the simultaneous finger tap. A similar pattern of results was found regardless of whether the emphasized repetition was compared against all unstressed repetitions in the same trial or only those in its immediate vicinity. Additionally, the data further indicate that a correlation exists between manual and speech action magnitudes even independently of the effects of an instructed imposition of emphasis or imposed rhythm. The fact that the magnitudes of the two actions are correlated in *unstressed* repetitions provides compelling further evidence, not previously demonstrated, that these two systems are indeed entering into a functional coordination throughout this task.

Compatible effects on movements were found in the *temporal* domain as well. All subjects show significant or near-significant lengthening of IRIs near a stress boundary in both domains, regardless of which domain the instructed stress is placed in, though the lengthening in the unstressed domain may be delayed by one repetition relative to the lengthening in the stressed domain. FT and LA IRIs are also correlated over the length of each trial, just as magnitude is correlated between the two domains.

These spatial and temporal effects are similar to the effects of emphatic stress in spoken language, where the movements of speech articulators under stress are larger and, somewhat more variably, longer than when no emphatic stress is present (de Jong, 1995; Edwards, Beckman, & Fletcher, 1991). These results are consistent with the π-gesture model of speech prosody whereby the pacing of an internal clock that controls the execution of motor gestures locally waxes and wanes as a function of linguistic prosodic structure (Byrd & Saltzman, 2003; Saltzman, Nam, Krivokapić, & Goldstein, 2008). While proposed to account for prosodic effects in speech, the theory predicts that all motor activity controlled by the same clock should show similar effects. The fact that the IRI-lengthening effects are not weaker cross-modally than they are within modes may suggest that the

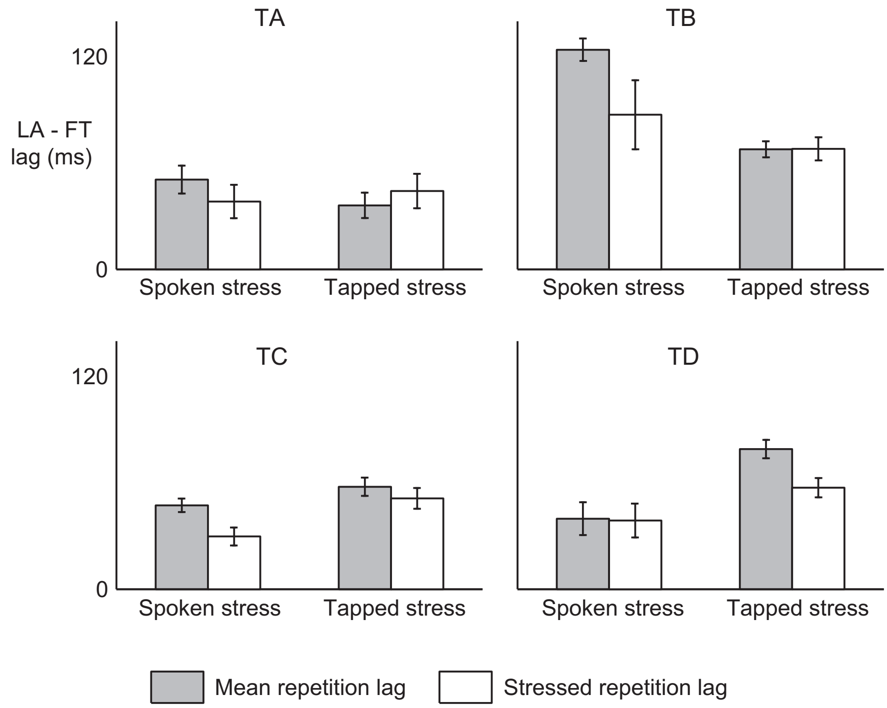

Fig. 7. Inter-gestural lag between lip aperture opening and fingertip movements. Means and standard deviations shown. All subjects show a positive lag, with lip aperture opening occurring after the onset of fingertip lowering. Emphatic stress causes the lag to decrease, which indicates a more synchronous production of the two movements.

finger and lip gestures are influenced by a single clock, although the fact that the cross-modal effects are, in some subjects, delayed by one repetition compared to the intra-articulator effects would need to be reconciled with this single-clock hypothesis.

An alternative possibility is that the implementation of a single emphatic repetition in one domain introduces a perturbation to the coordinative dynamics of the system as a whole. When external mechanical perturbations are imposed on the production of a rhythmic task, they alter its timing, changing the duration of repetitions immediately following the perturbation. This has been found for both speech (e.g. Gracco & Abbs, 1989; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Saltzman, Löfqvist, Kay, Kinsella-Shaw, & Rubin, 1998) and non-speech motor tasks (e.g. Kay, Saltzman, & Kelso, 1991). When a perturbation is introduced on only one articulator during coordinated multi-articulator movements (such as bimanual tapping), a steady phase-relation between the articulators is generally re-established after a period of instability (e.g., Court, Bennett, Williams, & Davids, 2002; Post, Peper, & Beek, 2000; Scholz, Kelso and Schöner, 1987), though the phase of the new pattern may differ slightly from the original (e.g. Kay et al., 1991; Saltzman et al., 1998). Relevant to the results of this study, both the perturbed articulator and the unperturbed articulator are affected. For example, when a perturbation is applied to the lips during production of a voiceless stop, the timing and duration of the associated laryngeal movements are changed (Munhall et al., 1994). Additionally, when a perturbation is given to only one arm in a task where subjects are waving both arms in a coordinated pattern (either in- or anti-phase), the perturbed arm speeds up while the unperturbed arm slows down to restore the target coordination phase (de Poel, Peper, & Beek, 2007). This is somewhat similar to the temporal pattern we see for post-stress IRIs, where they are generally lengthened in the *unstressed* domain but not the *stressed* domain. It may be the case that when the emphatic stress is planned in one domain, the lengthening for that domain occurs mainly before the stressed repetition itself; the lengthening in the *other* domain might be a perturbation response to restore a target relative phase between the two. This suggestion could be tested by introducing mechanical perturbations to both the lips and finger in a similar task to that used in this study and comparing the results of these imposed changes to the effects of explicit emphasis as shown here.

The two alternatives outlined above raise two related, but distinct, interpretations of our findings. One interpretation is that the synchronization task itself simply causes amplitudes of finger and lip movements to be organized into a single oscillatory task. The other is that the functional task of prosody—grouping information and highlighting salient information—harnesses a broad set of body components, including those not normally considered part of the speech system. These hypotheses are not mutually exclusive—the existence of a general cross-domain coupling of the two motor systems does not imply the absence of any prosodic effects; neither does evidence for the recruitment of the prosodic system imply that well-established, domain-general coupling mechanisms cease to be present.

It may seem strange for us to suggest that the prosodic system in speech is being used in this somewhat unnatural task of repeating ‘ma’ and ‘mop.’ After all, there was only one emphatic stress produced during a 10–15 s string of unstressed, identical monosyllabic words, which is quite different from the meaningful multi-layered prosodic structure present in language. However, it is clear that emphatic stress in the current study has similar spatial and temporal consequences as those demonstrated for sentence-level focus in speech, including increases in spatial magnitude and duration (de Jong, 1995). Moreover, like the prosodic aspects of natural speech, the emphatic stress in the current study is not a simple alternating rhythmic pattern but rather occurs at an identifiable location in time (either associated with a specific word token or with a co-word tap).

The prosodic interpretation also fits with previous findings in the literature. When single syllables and finger taps are produced synchronously, rhythmic spoken stress (as implemented by alternating loud and soft syllables) has a more consistent effect on finger tapping than explicit stress on finger taps has on speech (Smith et al., 1986). This finding is consistent with the data in this study, where a larger and more consistent effect of cross-modal stress was found for fingertip movements compared to lip movements (i.e. for spoken stress compared to tapped stress). Additionally, when subjects are instructed to produce a series of spoken syllables with finger taps placed between syllable repetitions, they produce the movements more synchronously than when they are given the reverse instructions, which may indicate a tighter coupling when attention is focused on producing speech as compared to finger movements (Inui, 2007). Prosody is known to recruit articulators across subsystems of the speech production mechanism (oral articulators: e.g., Byrd & Saltzman, 1998; Beckman & Edwards, 1992; Edwards et al., 1991; Fougeron & Keating, 1997; larynx: e.g., Cho & Jun, 2000; velum: e.g., Fougeron, 2001; Byrd, Tobin, Bresch, & Narayanan, 2009; lungs: Ladefoged, 1967; and also the oro-facial musculature: e.g., McClean & Tasko, 2002, 2003), and the effects of prosody have also been shown to extend to articulators not directly tied to speech production. For example, when subjects produce pointing gestures while speaking, the timing of those hand movements is tightly coupled with spoken stress (Rochet-Capellan et al., 2008). Prosodic boundaries and lexical stress in speech also influence the duration and timing of manual gestures in parallel with intonation contours (Esteve-Gibert and Prieto, 2013)). That study showed that deictic pointing gestures are timed to occur with the intonational peak of a stressed syllable, and that both the manual gesture and intonation peak are retracted when the accented syllable occurs in phrase-final position. Phrase boundaries have also been shown to increase the duration of manual gestures (De Ruiter, 1998), similar to the lengthening of speech articulator movements in phrase-final position. Particularly relevant to the current study, emphatic stress in speech has been shown to lengthen co-produced deictic pointing gestures, though the effects on the timing of the manual movement remain unresolved (Rusiewicz, Shaiman, Iverson, & Szuminsky, 2013, 2014). The stronger effects of spoken stress on manual movement than of manual stress on speech found in the current study are difficult to explain under the hypothesis that *solely general* coupling principles are driving the coordinative effects between the two motor domains. The prosody hypothesis, on the other hand, provides a simple explanation of this asymmetry: stress in the speech domain is part of the prosodic structure of speech, which has large influences across motor domains, while stress in the manual motor domain does not harness this prosodic architecture.

This support for a prosodic basis of some of the demonstrated cross-modal effects of stress seen in this study does not preclude the probability that general coupling dynamics also play a role in the cross-domain coordination. Even if the use of prosody can explain some of the asymmetries between stress in the two domains, general coupling dynamics are obviously playing a role in the domain-general effects. The coupling hypothesis can provide an explanation for the smaller increase in amplitude in the unstressed domain than that when stress was in the same domain as the movement. This difference can be explained if the coupling between the two domains is asymmetrical, with the domain receiving cognitive attention (i.e. the domain receiving stress) exerting larger effects on the unattended domain than vice versa. Such asymmetrical effects have been found for bimanual tapping based on the handedness of subjects (Spijkers & Heuer, 1995). For right-handed subjects, the left limb assimilated more to movements of the right limb than vice versa. This effect, while somewhat structural, is mainly due to attention (Rogers, Bradshaw, Cunnington, & Phillips, 1998).

The correlations in magnitude and duration seen during the non-emphasized portions of our trials would be the expected outcome with general coupling. Alternatively (or additionally) under a prosodically-oriented account, these small variations could be seen as micro-prosodic fluctuations. One possible way of distinguishing between the two 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 hypothesis—that cross-modal amplitude effects result solely from (attentionally-asymmetric) coupling—does predict an effect on finger movement.

5. Conclusion

In sum, the current study demonstrates that the coordination between speech and finger movements in a repetitive task exists in both space and time. The implementation of emphatic 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 slightly between subjects, similar variability exists for the temporal implementation of stress within a single domain as well. The fact that, despite these differences, we see cross modal effects for all subjects points to an inherent coordination between the two domains. Asymmetrical effects between stress in spoken and manual domains may arise from recruitment of the prosodic system.

Acknowledgments

Work supported by Grant no. DC03172 (Byrd) from the National Institutes of Health. The authors thank Tine Mooshammer for an extremely thoughtful feedback that significantly improved the presentation of this material. We also thank the Journal reviewers for their insightful feedback.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: University Press.
Allen, G. (1975). Speech rhythms: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, *3*, 75–86.
Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373.
Baader, A. P., Kazennikov, O., & Wiesendanger, M. (2005). Coordination of bowing and fingering in violin playing. *Brain Research. Cognitive Brain Research*, *23*(2-3), 436–443.
Beckman, M., & Edwards, J. (1992). Intonational categories and the articulatory control of duration. In: Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 359–375). Tokyo: Ohmsha, Ltd.
Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In: P. Keating (Ed.), *Phonological Structure and Phonetic Form—Papers in Laboratory Phonology III* (pp. 7–33). Cambridge: University Press.
Butterworth, G., & Hopkins, B. (1988). Hand–mouth coordination in the new-born baby. *British Journal of Developmental Psychology*, *6*, 303–314.
Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In: M. B. Broe, & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 70–87). Cambridge: Cambridge University Press.
Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America*, *120*, 1589–1599.
Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, *26*, 173–199.
Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*, 149–180.
Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics*, *37*, 97–110.
Carey, D. P. (2000). Eye–hand coordination: Eye to hand or hand to eye?. *Current Biology*, *10*(11), R416–R419.
Chang, P., & Hammond, G. R. (1987). Mutual interactions between speech and finger movements. *Journal of Motor Behavior*, *19*, 265–274.
Cho, T., & Jun, S. -A. (2000). Domain-initial strengthening as enhancement of laryngeal features: aerodynamic evidence from korean. *UCLA working papers in phonetics*, pp. 57–70.
Court, M. L. J., Bennett, S. J., Williams, A. M., & Davids, K. (2002). Local stability in coordinated rhythmic movements: Fluctuations and relaxation times. *Human Movement Science*, *21*(1), 39–60.
de Jong, K. (1994). The correlation of p-center adjustments with articulatory and acoustic events. *Perception & Psychophysics*, *56*(4), 447–460.
de Jong, K. (1995). The supraglottal articulation of prominence in english: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, *97*, 491–504.

de Poel, H., Peper, C., & Beek, P. (2007). Handedness-related asymmetry in coupling strength in bimanual coordination: Furthering theory and evidence. *Acta Psychologica*, *124*(2), 209–237.

De Ruiter, J. P. (1998). *Gesture and speech production* (Unpublished doctoral dissertation). Nijmegen, the Netherlands: Katholieke Universiteit.

Edwards, J., Beckman, M., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, *89*, 369–382.

Ejiri, K. (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica*, *55*, 226–237.

Ejiri, K., & Masataka, N. (2001). Co-occurences of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, *4*, 40–48.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*, 850–864.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*, 109–135.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*, 3728–3740.

Fowler, C. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*(3), 386–412.

Gentilucci, M., Santunione, P., Roy, A., & Stefanini, S. (2004). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *European Journal of Neuroscience*, *19*, 190–202.

Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In W. Trouvain & W. Barry (Eds.), *Proceedings of the XVIth international congress of phonetic sciences* (pp. 241–244).

Gracco, V. L., & Abbs, J. H. (1989). Sensorimotor characteristics of speech motor sequences. Experimental Brain Research, 75(3), 586–598.

Gracco, V., & Löfqvist, A. (1994). Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience*, *14*, 6585–6587.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347–356.

Hardcastle, J. W. (1985). Some phonetic and syntactic constraints on lingual co-articulation during /kl/ sequences. *Speech Communication*, *4*, 247–263.

Hulstijn, W., Summers, J., van Lieshout, P., & Peters, H. (1992). Timing in finger tapping and speech: A comparison between stutterers and fluent speakers. *Human Movement Science*, *11* (1–2), 113–124.

Inui, N. (2007). Interactions of speech and manual movement in a syncopated task. *Perceptual and Motor Skills*, *105*, 447–457.

Iverson, J., & Fagan, M. (2004). Infant vocal-motor coordination: Precursor to the gesture-speech system?. *Child Development*, *75*, 1053–1066.

Iverson, J., Hall, A., Nickel, L., & Wozniak, R. (2007). The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. *Brain and Language*, *101*, 198–207.

Iverson, J., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11), 19–40.

Ivry, R., & Keele, S. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, *1*(2), 136–152.

Kay, B., Saltzman, E., & Kelso, J. A. S. (1991). Steady-state and perturbed rhythmical movements: A dynamical analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 183–197.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. -S. (2003). Domain-initial articulatory strengthening in four languages. In *Papers in Laboratory Phonology 6* (pp. 145–163).

Kelso, J. A. S., Tuller, B., & Harris, K. (1983). A "dynamic pattern" perspective on the control and coordination of movement. In: P. MacNeilage (Ed.), *The production of speech* (pp. 138–173). New York: Springer-Verlag.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(6), 812–832.

Ladefoged, P. (1967). *Three areas of experimental phonetics*. Oxford: Oxford University Press.

Levelt, W., Richardson, G., & Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*(2), 133–164.

Loevenbruck, H., Dohen, M., & Vilain, C. (2009). Pointing is 'Special'. In: S. Fuchs, H. Loevenbruck, D. Pape, & P. Perrier (Eds.), *Some aspects of speech and the brain* (pp. 211–258). New York: Peter Lang Verlag.

Max, L., & Yudman, E. A. (2003). Accuracy and variability of isochronous rhythmic timing across motor systems in stuttering versus nonstuttering individuals. *Journal of Speech, Language, and Hearing Research*, *46*(1), 146–163.

McClean, M., & Tasko, S. (2002). Association of orofacial with laryngeal and respiratory motor output during speech. *Experimental Brain Research*, *146*, 481–489.

McClean, M., & Tasko, S. (2003). Association of orofacial muscle activity and movement during changes in speech rate and intensity. *Journal of Speech, Language, and Hearing Research*, *46*, 1387–1400.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Miles, D. (1937). Preferred rates in rhythmic response. *The Journal of General Psychology*, *16*(2), 427–469.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (p-centers). *Psychological Review*, *83*(5), 405.

Munhall, K., Löfqvist, A., & Kelso, J. A. S. (1994). Lip–larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America*, *95*(6), 3605–3616.

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

Parladé, M., & Iverson, J. (2011). The interplay between language, gesture, and affect during communicative transition: A dynamic systems approach. *Developmental Psychology*, *47*, 820–833.

Post, A., Peper, C., & Beek, P. (2000). Relative phase dynamics in perturbed interlimb coordination: The effects of frequency and amplitude. *Biological Cybernetics*, *83*(6), 529–542.

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

Rogers, M. A., Bradshaw, J. L., Cunnington, R. C., & Phillips, J. G. (1998). Inter-limb coupling in coordinated bimanual movement: Attention and asymmetries. *Laterality*, *3*(1), 53–75.

Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2014). Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Communication*, *57*, 283–300.

Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2013). Effects of prosody and position on the timing of deictic gestures. *Journal of Speech, Language, and Hearing Research*, *56*(2), 458–470.

Saltzman, E., Löfqvist, A., Kay, B., Kinsella-Shaw, J., & Rubin, P. (1998). Dynamics of intergestural timing: A perturbation study of lip–larynx coordination. *Experimental Brain Research Experimentelle Hirnforschung Expérimentation Cérébrale*, *123*(4), 412–424.

Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of the Speech Prosody 2008 conference*, Campinas, Brazil.

Scholz, J., Kelso, J. A.S, & Schöner, G. (1987). Nonequilibrium phase transitions in coordinated biological motion: Critical slowing down and switching time. *Physics Letters A*, *123*(8), 390–394.

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

Spijkers, W., & Heuer, H. (1995). Structural constraints on the performance of symmetrical bimanual movements with different amplitudes. *The Quarterly Journal of Experimental Psychology*, *48A*(3), 716–740.

Treffner, P., & Peter, M. (2002). Intentional and attentional dynamics of speech–hand coordination. *Human Movement Science*, *21*, 641–697.

Turvey, M. T. (1990). Coordination. *American Psychologist*, *45*, 938–953.

Wing, A., & Kristofferson, A. (1973). Response delays and the timing of discrete motor responses. *Attention, Perception, & Psychophysics*, *14*, 5–12.

Woodrow, H. (1951). Time perception. In: S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). Oxford, England: Wiley.

Yale, M., Messinger, D., Cobo-Lewis, A., & Delgado, C. (2003). The temporal coordination of early infant communication. *Developmental Psychology*, *39*, 815–824.

Yale, M., Messinger, D., Cobo-Lewis, A., Oller, D., & Eilers, R. (1999). An event-based analysis of the coordination of early infant vocalizations and facial actions. *Developmental Psychology*, *35*, 505–513.