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
Available online 27 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
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,