Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 57 (2014) 283–300

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Effects of perturbation and prosody on the coordination of speech and gesture

Heather Leavy Rusiewicz[a,*,1], Susan Shaiman[b], Jana M. Iverson[c], Neil Szuminsky[b]

[a] *Duquesne University, Department of Speech-Language Pathology, 412 Fisher Hall, 600 Forbes Avenue, Pittsburgh, PA 15282, USA*
[b] *University of Pittsburgh, Department of Communication Science and Disorders, 4033 Forbes Tower, 3600 Forbes at Atwood Street, Pittsburgh, PA 15260, USA*
[c] *University of Pittsburgh, Department of Psychology, 3415 Sennott Square, 210 South Bouquet Street, Pittsburgh, PA 15260, USA*

Abstract

The temporal alignment of speech and gesture is widely acknowledged as primary evidence of the integration of spoken language and gesture systems. Yet there is a disconnect between the lack of experimental research on the variables that affect the temporal relationship of speech and gesture and the overwhelming acceptance that speech and gesture are temporally coordinated. Furthermore, the mechanism of the temporal coordination of speech and gesture is poorly represented. Recent experimental research suggests that gestures overlap prosodically prominent points in the speech stream, though the effects of other variables such as perturbation of speech are not yet studied in a controlled paradigm. The purpose of the present investigation was to further investigate the mechanism of this interaction according to a dynamic systems framework. Fifteen typical young adults completed a task that elicited the production of contrastive prosodic stress on different syllable positions with and without delayed auditory feedback while pointing to corresponding pictures. The coordination of deictic gestures and spoken language was examined as a function of perturbation, prosody, and position of the target syllable. Results indicated that the temporal parameters of gesture were affected by all three variables. The findings suggest that speech and gesture may be coordinated due to internal pulse-based temporal entrainment of the two motor systems.

Keywords: Gesture; Perturbation; Prosody; Entrainment; Speech-hand coordination; Capacitance

1. Introduction

The pervasive, but still primarily subjective, finding that speech and gesture temporally overlap is frequently cited as a theoretical pillar of proof regarding the integrative nature of the two systems (e.g., Goldin-Meadow, 1999; Iverson and Thelen, 1999; Krauss et al., 2000; McNeill, 1992). The common observation is that gesture and speech roughly occur at similar times during communication. Even though many individuals observe and state that speech and gesture are produced in tight temporal "synchrony", the mechanism responsible for this potential temporal alignment as well as factors that may affect the coordination of these two actions have not been elucidated. Although recent investigations have examined the effect of variables such as prosodic prominence upon the execution of gestures, there is still relatively sparse and mostly descriptive data on the impact of perturbation of the speech system on the gestural movement, and vice versa. This study seeks to explore the effect of perturbation of the speech stream in addition to the effect of prosodic prominence upon the temporal parameters of corresponding gestures in a controlled paradigm with the goal to inform theoretical postulations regarding the speech and gesture production and add to the growing literature base on the interaction of these two systems.

---

* Corresponding author. Tel.: +1 412 396 4205; fax: +1 412 396 4196.
E-mail address: rusiewih@duq.edu (H.L. Rusiewicz).

[1] This data were collected in partial fulfillment of doctoral studies at the University of Pittsburgh, Pittsburgh, Pennsylvania.

0167-6393/$ - see front matter
http://dx.doi.org/10.1016/j.specom.2013.06.004

1.1. Theoretical accounts of speech–gesture temporal alignment

Despite the relatively recent surge of interest regarding the study of gesture, theoretical postulation regarding the interface of speech and gesture is limited and primarily rooted in a linguistic perspective (e.g., Krauss et al., 2000; McNeill, 1992; de Ruiter, 1998; de Ruiter, 2000). However, recent experimental findings that gestures align with prosodically stressed syllables (de Ruiter, 1998; Esteve-Gibert and Prieto, 2011; Leonard and Cummins, 2011; Rochet-Capellan et al., 2008; Rusiewicz et al., 2013) cannot be easily incorporated into a linear, linguistic-based theory of gesture and spoken language production. Nor can such theories incorporate the findings of Krahmer and Swerts (2007) who found that the production of gestures increased the acoustic prominence of syllables, even when the participants were instructed not to stress the target syllable. Likewise, a cognitive-linguistic account cannot explain the limited literature on speech-gesture perturbations which indicates that speech and gesture remain synchronized or alter the timing of movements when the timing of one of the movements is halted in some way, whether it be as a consequence of a speech error (de Ruiter, 1998), moments of stuttering (Mayberry and Jaques, 2000; Mayberry, 1998), delayed auditory feedback (McNeill, 1992), or perturbing the early part of pointing gestures (Levelt et al., 1985).

Recovery from perturbation can be examined to study the stability of a motor behavior, including speech production. Perturbation studies also provide abundant information regarding the role of sensory feedback for motor control. There is a wealth of data on the sensory information that is utilized by the motor system for upper limb movements. For instance the visual location of a target can be altered during reaching tasks to measure the compensation of movement (e.g., Paulignan et al., 1991; Prablanc and Martin, 1992). In addition, perturbation studies have demonstrated the importance of both visual feedback and proprioceptive sensory information for accurate pointing trajectories (e.g., Bard et al., 1999; Komilis et al., 1993. Typically, this literature is interpreted as evidence for the remarkable adaptability of the motor system as well as the dynamic coordination of structures. Even though examining perturbations can offer insight into the motoric coordination of speech and gesture, this relatively straight-forward paradigm has rarely been utilized as a tool to examine the temporal relationship of speech and gesture. In fact, a systematic investigation of the temporal coordination of speech and gesture following speech perturbation has yet to be conducted.

Though gestures certainly are a unique class of manual movements with the potential for substantial overlay of linguistic information, it is necessary to also consider the underlying motoric form of gestures and the similarity of the interactions of the motor processes for speech and manual movements, gesture or otherwise. For instance, movements such as finger tapping, align prosodically strong syllables and altering temporal parameters as a result of the manipulations in the speech stream (Franz et al., 1992; Kelso et al., 1981; Smith et al., 1986). Another interesting and relevant line of work was embarked upon by Gentilucci and colleagues (e.g., Gentilucci et al., 2001, 2004) which revealed the size of mouth opening and acoustic amplitude of syllables is influenced by the size of grasping movements. Taken together, these data imply that (1) the alteration of the temporal parameters of speech and manual movements due to perturbations as well as (2) the temporal coordination of prosodically prominent syllables and manual movements is at least in part explicable according to a motoric perspective.

1.1.1. Dynamic systems theory and entrainment

When contemplating the motoric coordination of speech and gesture, it is important to consider gesture first as a hand movement that can also be overlaid with language functions. As such, dynamic systems theory (DST) potentially offers the best launching pad for positing and testing the motoric coordination of speech and gesture. In recent decades, nonlinear science focusing on the coordinative structures and processes has emerged as a viable alternative to traditional top-down, executive controlled theories. The application of nonlinear dynamics stemmed originally from work in the area of physics, and later, the coordination of motor control within the human body (see [204] for a review). However, principles of nonlinear dynamics are now being applied to just about any behavior or event, from cognitive-linguistic processing (e.g., Thelen and Smith, 2002) to inanimate properties like biological and chemical processes (e.g., Lorenzo et al., 2007), but not often considered in relation to gestures. Indeed, the applications and specifications of DST even within the particular realms of speech (see van Lieshout, 2004 for a review), limb movement (see Schmidt and Lee, 1999 for review) and even the coordination of speech and manual movements like tapping (e.g., Kelso et al., 1981) are certainly vast and varied and beyond the scope of this manuscript. Hence, the present investigation is focused on studying the phenomenon of *entrainment* within a DST framework.

The idea of entrainment is certainly not new or specific to gestures, yet it is underrepresented in both the literature on spoken language production (though refer Franz et al., 1992; Kelso et al., 1981; Smith et al., 1986; van Lieshout, 2004) and gesture production. The mechanism of entrainment was first presented 70 years ago by von Holst (1973) in response to his observations of the mutual influence of fin movements of swimming fish. The general idea of entrainment is that the preferred temporal pattern of one oscillator will interface with the preferred temporal pattern of the second oscillator, resulting in either an identical rhythmic pattern or a compromise rhythmic pattern somewhere in between the two patterns relative to when they are produced in isolation.

Similarly, Thelen and Smith (2002) iterated “to the degree that two component systems have a history of time-locked activity, they will come to entrain each other and to mutually influence each other". Another definition comes from Clayton et al. (2004) who defined entrainment as "a phenomenon in which two or more independent rhythmic processes synchronize with each other" and in "such a way that they adjust towards and eventually 'lock in' to a common phase and/or periodicity" (p. 1). Merker et al. (2009) used the more descriptive term of "pulse-based rhythmic entrainment" (p. 4) to describe this phenomena.

It is important to note that two oscillators need not be produced in perfect rhythmic harmony in order to be entrained. Instead, it is more likely that two oscillators will share a pulse-based entrainment (Clayton et al., 2004; Merker et al., 2009; Bluedorn, 2002). That is, that the pattern of one behavior (e.g., deictic gestures) co-occurs at certain points in time with the cycle of another behavior (e.g., prosodically prominent syllables in the speech stream). Bluedorn (2002) provided a summary of pulse-based entrainment as follows:

> Entrainment is the process in which the rhythms displayed by two or more phenomena become synchronized, with one of the rhythms often being more powerful or dominant and capturing the rhythm of the other. This does not mean, however, that the rhythmic patterns will coincide or overlap exactly; instead, it means the patterns will maintain a consistent relationship with each other (p. 149).

Entrainment can be external, such that a rhythmic behavior of one species is entrained to that of another. Examples in human behavior are limitless and not restricted to movements. For example, studies demonstrated that conversational partners temporally entrain dialogue externally in a predictable manner (Włodarczak et al., 2012) and that finger tapping entrains to external metronome pulses (e.g., Corriveau and Goswami, 2009), as does repetitive phrase production (e.g., Cummins and Port, 1998). Entrainment can also be internal (i.e., self-entrained), such that one rhythmic pattern of an individual is entrained to another rhythmic pattern within the same individual (e.g., breath groups coordinate with ambulation patterns while jogging). According to Port et al. (1998), (p. 2), self-entrainment is the mutual influence of two oscillators that are actually "parts of a single physical system…when a gesture by one part of the body tends to entrain gestures by other parts of the same body". Port (2003) later proposed that the rhythm of speech is temporally spaced according to regular periodic patterns generated by neurocognitive oscillators. These oscillators produce pulses that "attract perceptual attention (and) influence the motor system…by biasing motor timing so that perceptually salient events line up in time close to the neurocognitive pulses" (p. 599).

Investigation of the impact of pulses in the speech stream, such as prosodically stressed or pitch accented syllables (referred to as *prominent* syllables from this point), on the subsequent timing of gesture may assist in exploring the potential entrainment of speech and gesture. This study aims to explore the effect of prominent syllables in the first and second position of target words produced in multiword controlled utterances upon the timing of deictic gestures similar to an earlier study (Rusiewicz et al., 2013). As aforementioned, a perturbation paradigm can also explore the dynamic entrainment of the speech and gesture systems. Thus, the current study also aims to examine the temporal parameters of gesture as a function of perturbations to the speech system. The overall objective of this work is to more fully develop and test theoretical thought and contemplate practical implications regarding the potential pulse-based internal entrainment of speech and manual movements.

1.1.2. Entrainment of speech and gesture

In short, dynamic systems theories attempt to explain the behavior and activity of complex systems. Thus, the application to human movement and later speech production was a straight-forward progression of the dynamic systems movement. There have been a number of studies of the dynamic properties of speech following similar paradigms utilized to study manual and limb movements (e.g., Saltzman and Byrd, 2000; Saltzman and Munhall, 1989; Tuller et al., 1982a; Tuller et al., 1982b; Tuller et al., 1983; see Kelso and Tuller, 1984 for review). A limitation of linear models is that they often cannot account for the fluid coarticulatory processes of speech production. A dynamic systems approach can more adequately describe context-dependent coarticulation by viewing speech production as a coordinative process.

It is proposed that speech articulators can also be thought of as coordinative structures. Saltzman and Byrd (2000) nicely summarize Saltzman and Munhall's (1989) task-dynamic model of speech production as follows, "in the task-dynamic model of speech production, articulatory movement patterns are conceived of as coordinated, goal-directed gestures that are dynamically defined… in particular, they have been modeled as critically damped oscillators that act as point attractors" (p. 501). The coordinative structures work together to produce articulatory gestures which are "changes in the cavities of the vocal tract – opening or closing, lengthening or narrowings, lengthenings or shortenings" (Liberman and Whalen, 2000, p. 188). The vocal tract configurations for a given gesture are created by goal-directed movements of the various structures involved in speech such as the velum, parts of the tongue, mandible, and lips. Although, many theorists have applied extensive and sophisticated mathematical work to the relative phase of oscillators involved in speech and various types of action, that is not the role of dynamic systems theory in this research endeavor. The reader is referred to the following references for greater detail on specific dynamic pattern models of speech production and more general motor behaviors (Haken et al., 1985; Saltzman and Munhall, 1989).

Recovery from perturbation can be examined to study the stability of a motor behavior, including speech production. Perturbation studies also provide abundant information regarding the role of sensory feedback for motor control. There is a wealth of data on the sensory information that is utilized by the motor system for upper limb movements.

For instance the visual location of a target can be altered during reaching tasks to measure the compensation of movement (e.g., Paulignan et al., 1991; Prablanc and Martin, 1992). In addition, perturbation studies have demonstrated the importance of both visual feedback and proprioceptive sensory information for accurate pointing trajectories (e.g., Bard et al., 1999; Komilis et al., 1993). Later, a perturbation study that randomly applied a load to the wrist during pointing in an effort to examine the resultant effects on the timing of speech is reviewd (Levelt et al., 1985).

There is also a plethora of experimental findings that demonstrate the importance of various types of sensory feedback on speech motor control. There are two broad categories of speech perturbation studies, those that directly affect the biomechanics of speech and those that affect the auditory feedback of speech. Many investigators have manipulated the biomechanics of articulator movement by introducing a bite block (Folkins and Zimmerman, 1981; Kelso et al., 1981), applying a mechanical load to an articulator, most often the mandible or lower lip, either in a consistent or transient manner (Abbs and Gracco, 1982, Abbs and Gracco, 1984; Folkins and Abbs, 1975; Gracco and Abbs, 1985, 1986; Kelso and Tuller, 1984; Kelso et al., 1984; Shaiman, 1989; Shaiman and Gracco, 2002), or by removing afferent information from the motor system by using local anesthetics like Xylocaine and nerve blocks (Kelso and Tuller, 1983; see Abbs et al., 1976 for review). Repeatedly, it has been demonstrated that accurate acoustic goals can be attained even in the face of these biomechanical perturbations. Therefore, there are many configurations of the vocal tract that are permissible for attaining an acoustic goal. In sum, "this ability to use different movements to reach the same goal under different conditions, called motor equivalence, is a ubiquitous property of biological motor systems" (Guenther et al., 1998, p. 6).

The sensory feedback that results from biomechanical perturbations is received and processed much more quickly than perturbations of auditory information. Auditory perturbations include manipulations of the feedback of fundamental frequency (e.g., Brunett et al., 1998), masking noise (e.g., Ringel and Steer, 1963; Kelso and Tuller, 1983), and the feedback of the auditory signal with the presence of a delay (e.g., Howell and Archer, 1984; Stuart et al., 2002). Even though there is a longer latency of response for auditory perturbations, the sensory information provided by the acoustic signal is clearly important for monitoring the accuracy of speech production. Most importantly for the present investigation, the "effects of delayed auditory feedback on speech are simply another example of deterioration in the performance of any serially organized motor behavior when a competing, rhythmic, out-of-synchrony event is co-occurring" (Smith, 1992, p. 253). As stated in the Introduction, DAF causes a temporal disruption of speech characterized by decreased speech rate, increased speech errors (e.g., phoneme exchanges), increased vocal intensity, and increased dysfluencies (e.g., Burke, 1975; Howell and Archer, 1984; Stuart et al., 2002). Thus, if speech and gesture are coupled oscillators that entrain and mutually influence the timing of one another, the temporal disruption of speech that results from DAF should also affect the temporal pattern of gesture production.

Though not yet widely tested in the gesture literature, there are motor-based accounts that specifically focus on the entrainment of speech and gesture that can potentially unify the literature on motoric coordination of manual movements, including gestures and spoken language production (Iverson and Thelen, 1999; Rusiewicz, 2011; Tuite, 1993). Tuite (1993) was the first to assert that gesture and speech are linked prosodically and that gesture and speech originate from a kinesic base according to his underspecified, nonetheless intriguing, Rhythmical Pulse Model. Regardless of the type of gesture, Tuite theorized that the gestural stroke "tends to coincide with the nuclear syllable of the accompanying tone group" (p. 100). Even though Tuite did not directly situate his hypothesis within a dynamic systems perspective, the notion of coordination of rhythmical pulses is very much in line with work in the dynamic systems literature (Port, 2003; Barbosa, 2002; Cummins and Port, 1996; Jones and Boltz, 1989; Large and Jones, 1999; O'Dell and Nieminen, 1999).

Another theory of gesture production rooted in dynamic systems theory of entrained systems was presented by Iverson and Thelen (1999). These theorists stated it is "through rhythmical activity, and later through gesture, the arms gradually entrain the activity of the vocal apparatus…this mutual activation increases as vocal communication through words and phrases becomes more practiced, leading to a tight synchrony of speech and gesture" (p. 36). In addition, this is the only theory that makes predictions about the simultaneous development of the speech and gesture production systems in early childhood, though the authors made no explicit statements about what affects the synchronization of speech and gesture produced by older children and adults. Likewise, neither Iverson and Thelen, nor Tuite (1993) considered the role of perturbation on the temporal entrainment of gestures.

Recently, the Theory of Entrained Manual and Speech Systems (TEMSS) Rusiewicz, 2011 posited speech and gesture as two internal, couple oscillators based upon amalgamation of prior theoretical notions (Iverson and Thelen, 1999; Port, 2003; Tuite, 1993). The TEMSS proposed that the production of speech, marked by prominent acoustic events such as pitch accented and/or stressed syllables, acts as a rhythmic pulse and influences the temporal pattern of coinciding manual movements, including, but not limited to gestures. Specifically the most salient portion of the speech unit (e.g., a prominent syllable) entrains the most salient portion of the gesture (e.g., a gestural stroke) resulting in the perceived temporal coordination of speech and gesture. Additionally, Rusiewicz (2011) hypothesized that perturbations in the speech stream will result in temporal disruptions in the gesture system due to the nature of coordinative rhythmic structures that entrain and mutually influence the timing of the other. Thus, by measuring the temporal parameters of gesture following manipulation of not only prosodic prominence, but also speech perturbation, one can begin an exploration of the mechanism of speech/gesture coordination and add to the relatively sparse literature on the interface of speech and gesture as a function of perturbation (McNeill, 1992; de Ruiter, 1998; Mayberry and Jaques, 2000; Mayberry, 1998; Levelt et al., 1985).

Two tenets of the TEMSS (Rusiewicz, 2011) were tested in the present investigation and correspond to the following research questions:

1. Are the temporal parameters of deictic gestures affected by perturbations in the affiliated spoken production in a controlled paradigm?
2. Are the temporal parameters of deictic gestures affected by prosodic prominence (e.g., increased temporal alignment of deictic gestures and syllables with prosodic stress compared to the same syllables produced with no stress)?

The time to produce the speech required in each trial, as well as the syllables and words therein, was expected to increase when an auditory delay was present relative to when there was no delay. Likewise, the temporal variables associated with the production of the corresponding deictic gesture also were predicted to be longer for DAF conditions compared to normal auditory feedback (NAF) conditions. In addition, no difference was expected to be found for the time between gestures for DAF versus NAF conditions due to the tight temporal entrainment of the two systems even with perturbation as described in previous work (McNeill, 1992; Mayberry and Jaques, 2000; Mayberry, 1998). The temporal relationship between speech and gestures was predicted to also change as a function of prosodic prominence. Specifically, it was expected that the interval between target syllables and affiliated gestures would be reduced for responses with prosodic prominence and that gestures would take longer for second position syllables with prominence due to the pulse-based coordination of speech and gesture movements.

2. Methods

2.1. Participants, design, and variables

Fifteen monolingual American English speakers (four male and eleven female) ranged in age between 22 and 31 years ($M = 25.1$ years, $SD = 3.2$ years) and completed between 12 and 17 years of education ($M = 16$ years, $SD = 1.5$ years). Each participant was paid fifteen dollars and five also received extra course credit. All participants passed a hearing and vision screening.

A three-way ($2 \times 2 \times 2$), within group repeated measures design was employed. The three independent variables were speech perturbation (i.e., presence or absence of 200 ms auditory delay), prosodic prominence (i.e., present or absent), and syllable position (i.e., first or second). Delayed auditory feedback (DAF) was chosen as the perturbation method to build upon McNeill (1992) earlier descriptive work with DAF as well as to capitalize on the effects on the timing of speech production such as decreased speech rate, prolonged voicing, and increased disfluencies (e.g., Stuart et al., 2002). Contrastive pitch accent was chosen to systematically manipulate prosodic prominence and also allow the same phonetic constructions to be produced both with and without prominence. Syllable position was manipulated to control for the strong trochaic bias in American English and to eliminate the likelihood that individuals aligned gestures with first position syllables simply because over 90% of English content words begin with a stressed syllable (Cutler and Carter, 1987). The dependent variables measured including the durations of (1) total gesture time (i.e., gesture onset to gesture offset) and (2) gesture launch time (i.e., gesture onset to gesture apex) and the interval of (3) gesture launch (i.e., gesture onset to gesture apex) midpoint to vowel midpoint.

2.2. Procedures

2.2.1. Experimental setup and stimuli

Experimental setup and stimuli were similar to those fully described in Rusiewicz et al. (2013). Participants were seated in a chair facing a 18-inch by 24-inch piece of Plexiglas onto which the stimuli were back-projected (see Fig. 1). Position from the screen was individually calibrated by measuring 2 in from the maximum extension of a pointing gesture to the midpoint of the screen. The individuals' response were digitally recorded at a sampling rate of 48 kHz with a head-worn hypercardioid condenser microphone (Crown CM-312AHS). A second microphone routed the individual's acoustic signal to the Facilitator (KayPENTAX™, Model 3500) which both amplified the individual's loudness level to 70 dB SPL via Sony MDR-V6 Monitor Series Headphones and also delayed the acoustic signal by 200 ms in the perturbation trials. The amplification level was chosen based upon prior research that demonstrated increased effects of DAF at louder

Fig. 1. Equipment setup with participant facing the stimuli presentations and the theremin capacitance antennas housed behind the screen.

levels, optimally at 70 dB SPL (Elman, 1983; Howell and Sackin, 2002) Likewise, an auditory delay of 200 ms was chosen because it approximates the average length of a syllable (Turvey, 1990) causes the most consistent and largest speech disruptions in typical speakers (e.g., Marslen-Wilson and Tyler, 1981), and was also the auditory delay described by McNeill (1992) earlier work.

A novel methodology was employed for measuring the temporal parameters of the deictic gestures. The gesture movement, specifically the apex (i.e., point of maximum hand and finger extension) was captured using capacitance sensors placed behind the Plexiglas within a modified theremin device (Theremax PAiA 9505KFPA) and recorded using a Dataq acquisition system (WinDaq/Lite, DI-400; Dataq Instruments WinDaq Acquisition, Version 3.07). The reader is referred to the PAiA website for technical details and schematics of the design and mechanics of the theremin device (http://www.paia.com/theremax.asp). The primary modification to the theremin was the use of only one of the two capacitance antennae and the placement of the antenna behind the Plexiglas screen rather than attached to the theremin itself. Capacitance is the ability to store and modulate electrical charge while a theremin is traditionally a musical instrument that uses capacitance to create a warbling-like sound that can shift in frequency and loudness as the hands move in approximation, but never in contact, to the sensors. The theremin not only produces a musical sound as output, but also a linear voltage trace. Thus, as the hand and finger of a pointing gesture get closer to the capacitance sensor, the voltage increases. The time of maximum voltage is equal to the time of gesture apex. The theremin was able to reliably detect movement less than 1 cm in relation to the capacitance antenna. The oscillators of the theremin were calibrated initially and then periodically checked for drift, though none was noted during the duration of the study. Additionally, an optical reflective sensor (Fairchild QRD1114), situated on the convexed starting position on the armrest, measured the time of gesture onset and offset by an infrared emitter and detector recording the time an object moves further away than (i.e., gesture onset) or returns to (i.e., gesture offset) 0.05 in from device surface.

As illustrated in Fig. 2, each stimulus was presented via Stimulate software (Necessity Consulting) and consisted of first a screen with a question prompt followed by a second screen with a picture array that required a spoken and gestural response. A third black screen with a random interstimulus interval (ISI) between 1 and 4 s was then presented. The experiment was self-paced by programming the ISI screen to appear only after the participant's gesture offset was recorded. Thirty trials in each condition were produced with DAF and the other thirty without DAF resulting in one hundred and twenty total trials produced under the influence of DAF and one hundred and twenty without the influence of DAF. Thus, each individual was presented with 60 total stimuli that elicited pitch accent on the first syllable position. Thirty of these responses were produced under the non-altered auditory feedback condition and the other thirty under the DAF condition. The 30 stimuli that manipulated pitch accent placement on the first syllable of compound words in the NAF condition were the same 30 stimuli in the DAF condition. This was also the case for the stimuli with manipulation of pitch accent on the second syllable, neutral pitch accent on the first syllable, and neutral pitch accent on the second syllable.

Prosodic prominence (i.e., contrastive pitch accent) and syllable position were manipulated within compound word pairs (see Appendix) embedded within seven word utterances. Compound word pairs such as *lifeboat/lifeguard* and *toothbrush/paintbrush* were selected due to ability to emphasize the contrastive element of the word pair (e.g., *Is that a life'boat?; No, that is a lifeGUARD'*) and to make comparisons of the same phonetic structure both with and without pitch accent. An additional advantage of manipulating pitch accent on the second syllable of the compound word is the ability to place accent on a normally unstressed syllable and reducing the potential confounding effect of preferred trochaic stress patterns in American English. Thus, fifteen of the word pairs elicited first position prominence and fifteen elicited second position prominence.

Prosodic prominence was manipulated via the initial questions prompt (e.g., *Is the football above the square*?). When the picture array was then shown (see Fig. 2), the participant was instructed to point to the target picture and respond accordingly. Therefore, if a football was indeed over the square, a neutral response was elicited (e.g., *Yes, the football is above the square*). If instead a baseball was above the square, the response held contrastive pitch accent and prominence (e.g., *No, the 'BASEball is above the square*). Of the 120 unique stimuli, 30 required a pitch accent on the first syllable e.g., *No, the 'HOTtub is above the star*) and 30 required pitch accent on the second syllable (e.g., *No, the life'BOAT is before the cross*). The remaining trials were identical but were produced without contrastive accent or illustrations of the compound word stimuli were presented two at a time, one stimulus acted as the target and the other as a distractor. The illustrations were presented in one of four possible locations in relation to one of eight possible shapes. A variety of shapes, locations, and target distractor illustrations were chosen and randomly organized to increase attention to the task and reduce the risk of rote phrasing patterns.

2.2.2. Familiarization, practice, and experimental phases

The participants were first familiarized with all target words and accompanying illustrations. If any picture was labeled incorrectly, the correct word was provided verbally by the examiner. They were also familiarized with the DAF by reading the first four sentences of the Rainbow Passage with a 200 ms auditory delay at 70 dB SPL.

Standardized instructions were digitally presented along with a single audiovisual model shown within the instructions. The standard pointing gesture with the index finger

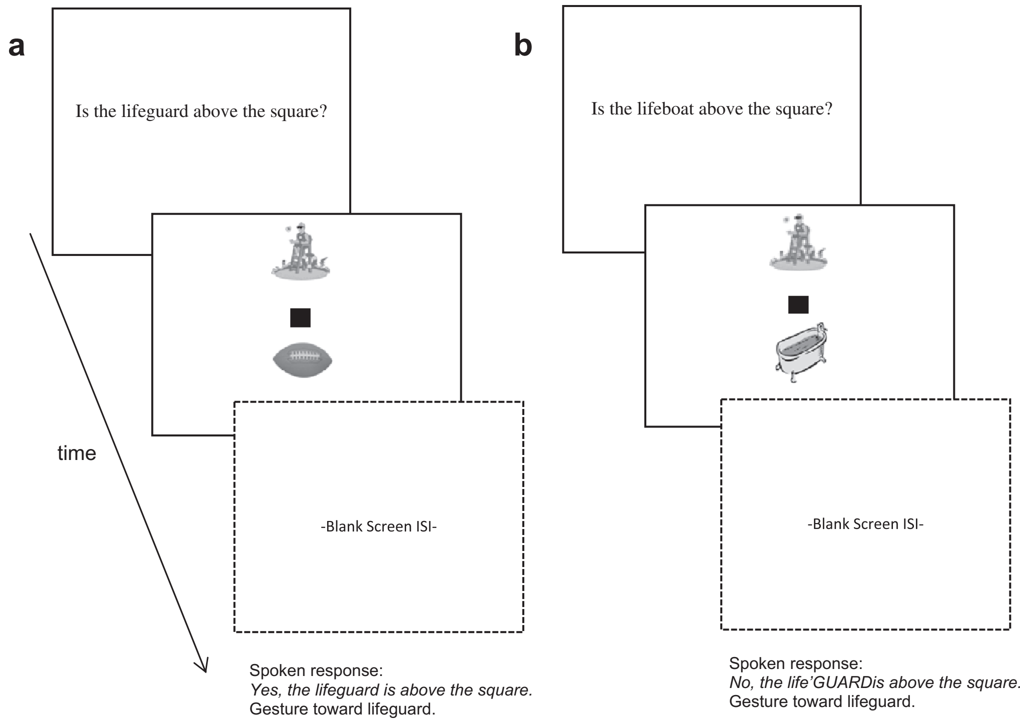


Fig. 2. Examples of two experimental trials consisting of a question prompt, illustrations, and a blank screen. The target is *guard* for each trial, produced with (a) no prominence (i.e., neutral) and (b) with contrastive prominence.

extended was shown in this model, though specific instructions to complete the deictic gesture in this form were not provided. The participants were instructed to point and produce the spoken response, though no details on the timing of these responses was given. Prior to the experimental trials, the individuals completed eight practice trials with target compound words not of the experimental set. Verbal feedback was provided only during these practice trials to correct inaccurate responses (e.g., incorrect placement of contrastive pitch accent). The participants were also given a deceptive motivation within the instructions both to point to and label the stimuli. They were told their responses would be shown to two examiners after the experiment was completed. In this explanation, one of these examiners would only see their responses and the second would only hear their responses. Further, the participants were told that these examiners must identify the chosen stimuli pictures. Therefore, both the gestured and spoken responses were imperative.

The 240 experimental trials required approximately 60 min and participants were provided breaks as requested. Perturbation was overlaid in counterbalanced blocks of 20 trials as standard in prior investigations with manipulations of delayed auditory feedback (e.g., Howell and Sackin, 2002). In other words, 20 trials were presented without DAF, followed by 20 with DAF, etc. The presentation order of the perturbation condition was randomized across participants so that seven began the experiment with a non-perturbed block of trials and eight began with a perturbed set. Syllable position and prominence were also randomly distributed. The beginning of each trial was marked by a 6 s presentation of the question prompt (e.g., *Is the bathtub above the square?*) followed by the 7 s presentation of the pictures and shape. The beginning of each response was marked by the lift of the hand from the neutral position. The termination of both the trial and response was marked by the return of the hand to the neutral position. Thus, the trials were self-paced. After the optical sensor recorded the termination of the gesture, a blank screen was presented for a randomly determined interstimulus interval (ISI) to allow adequate time to return to the neutral position. The ISI varied between 1 and 3 s. The rationale for randomizing the ISI was to further reduce anticipation of response initiation (e.g., Franks et al., 1998).

2.3. Data reduction and analyses

The acoustic and gesture data channels were synchronized with the Dataq system. Two time points for each target syllable (i.e., vowel onset and vowel offset), two time points for each sentence (i.e., sentence onset and sentence offset), and three time points for each gesture (i.e., gesture onset, gesture apex, gesture offset) were recorded. Adobe Audition 3.0 software was used to obtain the time of word onset and vowel midpoint of the target syllable from a wide-band spectrographic display with a 300 Hz bandwidth filter. Each vowel of the target syllable was isolated according to criteria described by Shriberg et al. (2003). The vowel onset, offset and total vowel duration were measured for each of the target syllables. Vowel midpoint was calculated by dividing the duration of each vowel by two.

Though prosodic prominence is potentially marked by several acoustic variables, vowel midpoint was chosen as the dependent variable to allow for a single time point measure that reflects the overall vowel duration as used in previous work (Rusiewicz et al., 2013). Vowel duration is frequently cited as the most consistent acoustic correlate of prosodic stress (e.g., van Kuijk and Boves, 1999). Another reason vowel midpoint was chosen was that prosodic prominence affects the vowel/nucleus of a syllable to a much greater degree of than the rime, syllable, or word duration (e.g., van Kuijk and Boves, 1999). Lastly, Krahmer and Swerts (2007) recently provided strong support for choosing vowel duration as the acoustic dependent measure. Their findings indicated that vowel durations were longer for syllables produced with a gesture even when the participants were instructed to actually produce the affiliated syllable without prominence. Vowel midpoint was calculated from the vowel duration measures to provide a single time point to relate to the gesture launch midpoint, hence providing an interval measure of synchrony.

The data were also analyzed to assure that DAF affected the overall spoken production and indeed perturbed speech. Thus, the total sentence duration and mean time error (MTE) per trial was calculated for each trial to validate the effect of DAF upon the temporal execution of speech production. MTE is a measure often used to reflect the expected lengthened duration to complete a spoken language production under the influence of DAF compared to NAF conditions. Elman (1983) defines MTE as the “mean difference between the time it takes a subject to complete a pattern with DAF and the time it takes with normal auditory feedback” (p. 109).

All gesture time points were automatically extracted by the Dataq system. As noted previously, the time of gesture onset and offset were obtained from the optical hand rest sensor and the time of gesture apex corresponded to the maximum output voltage signal of the theremin. Total gesture time (ms), gesture launch time (ms), and gesture launch midpoint (i.e., the temporal midpoint of the gesture launch) were calculated from these time points. Lastly, the interval measure of gesture launch midpoint to vowel midpoint (GLM-VM) was calculated by taking the absolute value of the time between these two moments. Although gesture apex was chosen as the dependent measure in earlier work (de Ruiter, 1998; Esteve-Gibert and Prieto, 2011; Rochet-Capellan et al., 2008; Rusiewicz et al., 2013), review of the resultant data and consideration of temporal entrainment of the systems prompted reconsideration of a more appropriate time point within the gesture. For instance, inspection of the entire gesture trajectory that accompanied these multiple productions of lengthier, more natural stimuli (Rusiewicz et al., 2013) compared to previous work (e.g., 5, 7, 9, 10) highlighted the concern that the maximum point of extension may not be the ideal dependent measure. Rather, a measure of the entire gesture stroke, or the portion of the gesture that corresponds to the most meaningful part of the movement and includes the gesture apex, may be a better reflection of the synchronization of speech and gesture. The midpoint of the gesture launch (i.e., time from gesture onset to gesture apex) is a possible time interval to use. The measure may be less sensitive to idiosyncratic movement patterns as well, for instance if a participant moved their arm/hand toward the screen but pulsed a bit closer prior to their return to the rest position. Additionally, gesture launch midpoint is a more analogous gesture measure to the acoustic measure since vowel midpoint was chosen to capture stress changes across an entire vowel instead of choosing a single time point in the signal like peak fundamental frequency. In comparison to gesture apex (i.e., the endpoint of a gestural stroke), gesture launch midpoint is potentially more analogous to a pulse point within the gesture movement, rather than the endpoint of the movement with the apex as demonstrated in our previous work (Rusiewicz et al., in press). Also, considering the midpoint of the span gesture onset to apex is a better counterpart to the chosen acoustic measure given that vowel midpoint was chosen to capture stress changes across an entire vowel instead of choosing a single time point in the signal like peak fundamental frequency. Thus, the motivation for analyzing GLM-VM is that this measure may be a more accurate single time point to analyze for each deictic gesture if one is interested in examining the pulse of movement rather than the arguable termination of the meaningful and programmed portion of the movement. Though, future work can and should investigate the motor linkages of these two systems by employing kinematic and discrete phase measures such as peak velocity, gesture launch midpoint was chosen for the current study given the methodological setup and limitations.

Interrater reliability was completed for the acoustic measures associated with three (20%) randomly chosen participants using descriptive statistics and Pearson product-moment correlation coefficient. Adequate reliability was attained for both vowel onsets and offsets ($r = .984$) and sentence onsets and offsets ($r = .998$). Additionally, similar means and standard deviations of the on and offset points were noted for the two raters. The mean vowel time points for the first author were 1766.31 ms ($SD = 424.57$ ms) for onsets compared to 1773.72 ms ($SD = 428.56$ ms) for the second rater. Likewise, the vowel offset measures were similar for the first ($M = 1947.97$ ms; $SD = 449.58$ ms) and second judge ($M = 1952.26$ ms; $SD = 445.91$ ms).

Trials were excluded if they were produced in error as determined by two independent raters. A third rater was used to provide a consensus rating in the case of disagreement between the first two raters. Disagreement occurred on 8.3% of the 3600 total trials. Trials produced in error were excluded along with their paired responses in other conditions, yielding 169 trials (4.7% of the 3600 total trials) incorrectly produced due to incorrect stress placement (77 trials), produced an error on the target such as a hesitation or speech error (78 trials), failure of the DAF equipment (7 trials), and incomplete or inaccurate gestures such as head scratching before pointing or not fully lifting and extending the hand and finger during the trial (7 trials). The number of excluded trials is less than the 6.5% in a similar study using more simplistic stimuli (Rochet-Capellan et al., 2008). The final dataset with erroneous trials and their counterparts in associated conditions removed yielded 2924 responses for analyses. These data points were aggregated for each participant and then compared using three-way repeated measures analysis of variance (ANOVAs) with the significance level set at $\alpha = .05$.

Data are presented for sentence durations to validate the perturbation method and for vowel durations to validate the manipulation of prosodic prominence and speech perturbation. Descriptive statistics are provided for each of the dependent measures as well as the results of the ANOVAs. The results were assessed for deviation from normality with no extreme outliers found, though some concern regarding normality for some conditions of some of the dependent measure were determined. Thus, the data were transformed by computing the base 10 logarithm of $(x + 1)$ to increase normality. Though the results remained unchanged for total gesture time and gesture launch time, this transformation did change the results of the ANOVA for GLM-VM interval. After log transformation, Shapiro–Wilk tests of normality were nonsignificant for all eight conditions for GLM-VM interval, indicating an adequate normal distribution for subsequent analysis. As a consequence, the results presented for GLM-VM interval are from the transformed dataset whereas the results for total gesture time and gesture launch time are from the non-transformed dataset.

3. Results

3.1. The effects of perturbation on spoken responses

The data first were analyzed to evaluate the effects of DAF to assure that each subject's speech indeed was perturbed, as evidenced by the production of longer utterances in the trials produced with DAF compared to these same trials produced with no auditory delay as measured by MTE. As expected, MTE was positive for all participants, indicating that the time from utterance onset to offset was longer for responses spoken with DAF than those same responses without DAF for all participants ($M = 867.53$ ms; $SD = 681.52$ ms). A three-way ANOVA was performed to assess the impact of DAF, as well as prominence and syllable position on sentence duration of the spoken responses. As expected, only two main effects were significant. Indeed, sentences were significantly longer in duration when produced with DAF [$F(1,14) = 24.049$, $p < .0001$]. Sentences were significantly longer in the DAF condition ($M = 3123.35$ ms; $SD = 214.05$ ms) than the NAF condition ($M = 2256.80$ ms; $SD = 68.91$ ms). Additionally, vowels were longer when produced with an auditory delay ($M = 265.84$ ms, $SD = 122.73$ ms) compared to without auditory feedback ($M = 194.69$ ms, $SD = 85.42$ ms; see Fig. 3).

3.2. The effects of prominence and position on spoken responses

Also as expected, sentence durations were longer for the contrastive pitch accent trials ($M = 2874.92$ ms; $SD = 148.33$ ms) than for neutral stress trials ($M = 2505.23$ ms; $SD = 121.19$ ms); [$F(1,14) = 39.420$, $p < .0001$]. Likewise, vowel durations were significantly increased for pitch accented syllables ($M = 248.82$ ms, $SD = 118.27$ ms) relative to neutral syllables ($M = 177.16$ ms, $SD = 78.55$ ms) [$t(1461) = -19.314$, $p < .001$]. This difference was also significant [$t(1461) = -19.618$, $p < .001$]. Unexpectedly, second position vowels averaged 235.07 ms ($SD = 109.55$ ms) compared to 19.38 ms ($SD = 99.80$ ms) for first position syllables [$t(1461) = -11.17$, $p < .011$] (Fig. 5). Moreover, second position syllables significantly differed in duration as a function of contrast [$t(730) = 17.781$, $p < .001$]. Vowel durations were almost 100 ms longer on average for second position syllables with contrastive pitch accent ($M = 275.59$ ms, $SD = 113.36$ ms) relative to when the second position syllables were neutral ($M = 195.22$ ms, $SD = 87.19$ ms).

3.3. Gesture duration measures

The mean durations (ms) for total gesture time, gesture launch time, and GLM-VM interval are displayed in Fig. 4 for each of the three factors; perturbation, prominence, and position.

3.3.1. Total gesture time

Total gesture time differed as a function of perturbation, prominence, and position. As predicted, the time to complete a gesture was longer for sentences produced with DAF ($M = 1609.40$ ms; $SD = 150.55$ ms) relative to NAF ($M = 1500.03$ ms ($SD = 103.39$ ms), though failed to reach significance [$F(1, 14) = 4.434$, $p < .054$]. However, a significant main effect of total gesture time as a function of *contrast* [$F(1,14) = 10.087$, $p < .007$] and *syllable position* [$F(1,14) = 6.344$, $p < .025$] was found. On average, gestures 155 ms longer for utterances produced with contrastive pitch accent ($M = 1632.26$ ms, $SD = 145.36$ ms) than the same utterances produced without contrastive pitch accent ($M = 1477.16$ ms, $SD = 109.85$ ms). As shown in Fig. 5, a significant two-way interaction of *contrast x position* was analyzed utilizing a post hoc analysis of simple main effects using the Bonferroni correction with significance set at .025 with second syllable position reaching significance [$t(30) = 17.50$, $p < .0001$] but not first syllable position [$t(30) = 1.57$, $p < .063$]. In other words, total gesture time was longest for trials that held contrastive pitch accent on the second syllable ($M = 1671.87$ ms, $SD = 140.28$ ms) with an average 79 ms difference between the total gesture time for first and second position condition when produced with contrastive stress though only a 2 ms difference for first and second position condition when produced with neutral stress.

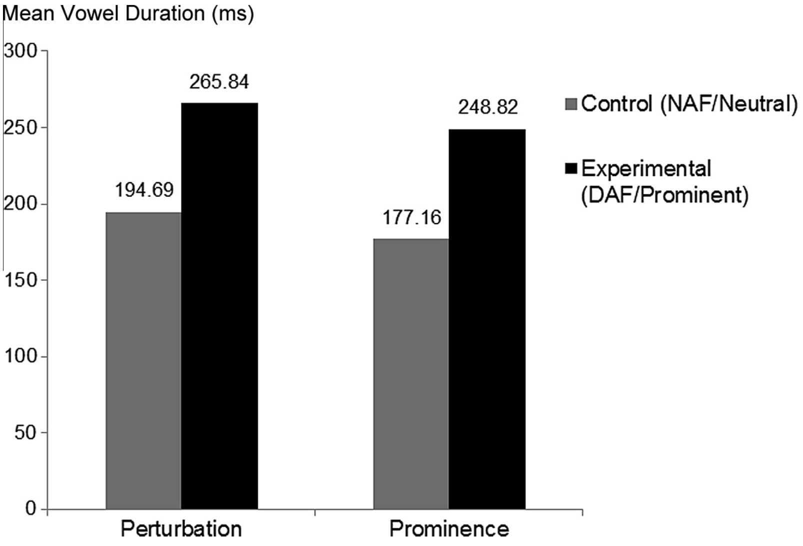


Fig. 3. Mean vowel durations (ms) for the perturbation and prominence conditions.

3.3.2. Gesture launch time

Analyzing gesture launch time enabled an exploration of the arguably more purposeful portion of the gesture movement associated with the entraining pulse of movement. Similar to total gesture time, gesture launch times were longer for the trials spoken with DAF ($M = 834.52$ ms, $SD = 98.64$ ms) compared to gesture launch times without an auditory delay ($M = 789.61$ ms, $SD = 63.82$ ms), though this main effect did not reach significance [$F(1,14) = 1.202$, $p = .291$]. Also like total gesture time, the ANOVA revealed a significant main effect of *contrast* [$F(1,14) = 17.880$, $p < .001$]. On average, gesture launch times were 120 ms longer for trials produced with contrastive pitch accent compared to when the same sentences were produced without pitch accent. No significant main effect was found for *position* for gesture launch time [$F(1,14) = .728$, $p = .41$]. In fact, gesture launch times for first and second position syllables differed only by 17 ms on average. Yet, there was a significant two-way interaction between *contrast* and *syllable position* [$F(1,14) = 5.910$, $p < .029$]. The average gesture launch time was shortest for trials produced with no pitch accent on the second syllable ($M = 741.20$ ms, $SD = 60.43$ ms) but longest for trials produced with pitch accent on the second syllable ($M = 899.50$ ms, $SD = 91.11$ ms). As shown in Fig. 5, mean gesture launch time decreased from first to second position conditions for neutral syllables but increased from first to second position conditions for accented syllables. The significant two-way interaction of *contrast x position* was further analyzed utilizing a post hoc analysis of simple main effects using the Bonferroni correction with significance set at .025. The effect of contrast was significant for both first [$t(30) = 2.49$, $p < .009$] and second syllable position [$t(30) = 4.88$, $p < .0001$]. Explicitly, the time for an individual to move from gesture onset to gesture apex was longer for trials produced with contrastive pitch accent than those produced with no accent, regardless of the position of the accented syllable.

3.4. Gesture launch to vowel midpoint (GLM-VM) interval

The intervals from gesture launch midpoint to vowel midpoint ($M = 2.62$, $SE = .08$) were significantly longer for the DAF trials compared to the no delay trials ($M = 2.44$, $SE = .07$) [$F(1,14) = 32.932$, $p < .001$]. Thus, counter to predictions, there was a difference in the temporal alignment of speech and gesture as a function of perturbation. However, results were consistent with the predictions for the factors of prominence and position. GLM-VM was significantly shorter for syllables with contrastive pitch accent ($M = 2.50$, $SE = .08$) relative to the same syllables produced without pitch accent ($M = 2.57$, $SE = .07$) [$F(1,14) = 27.848$, $p < .001$]. The interval from gesture launch midpoint and vowel midpoint was also significantly shorter on average for first position syllables ($M = 2.40$, $SE = .08$) compared to second position syllables ($M = 2.67$, $SE = .08$) [$F(1,14) = 5.301$, $p < .037$]. There were no significant interactions between the three variables for GLM-VM interval. The significant main effects and descriptive data for syllable position and contrast for GLM-VM interval offers further support that individuals altered the temporal parameters of their deictic gesture to coordinate with the temporal parameters of the spoken response. The significant main effect of speech perturbation for GLM-VM interval also indicates that the time between the gesture and target syllable becomes longer when speech is produced under the influence of an auditory delay.

4. Discussion

4.1. Summary

The temporal parameters of speech and gestures were altered by an auditory delay form of perturbation as well as the presence of contrastive pitch accent. Both sentence durations and vowel durations were significantly longer

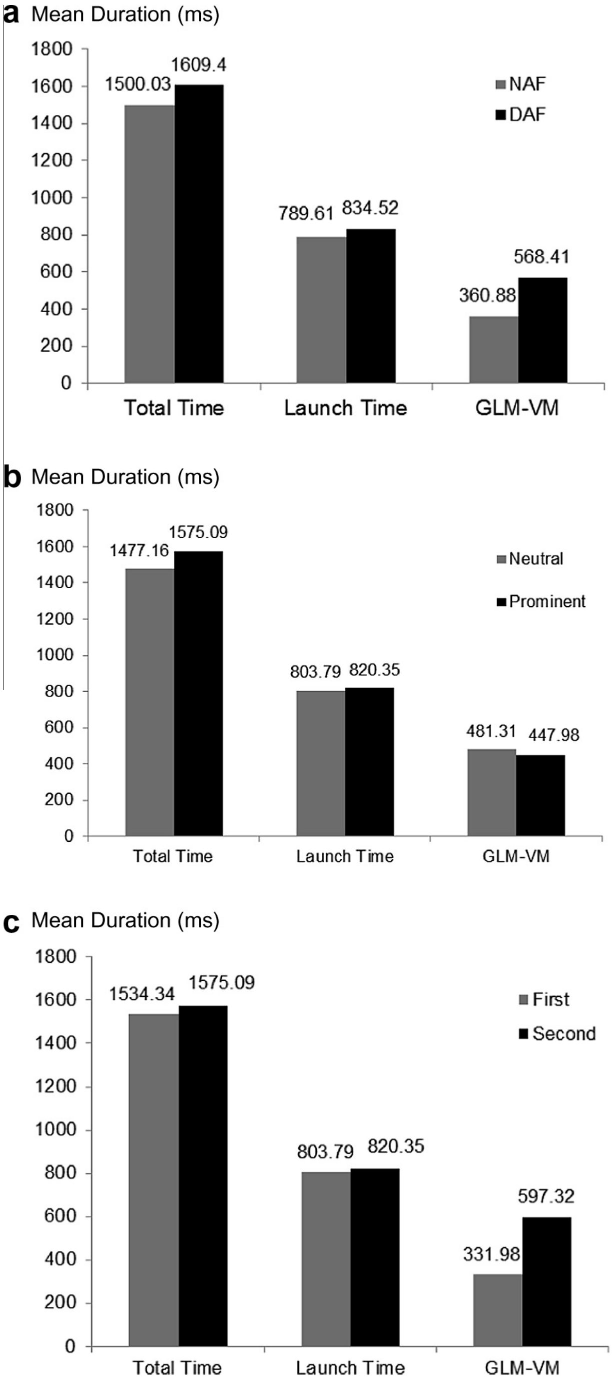


Fig. 4. Mean total gesture times (ms), mean gesture launch times (ms), and mean non-transformed GLM-VM intervals (ms) for (a) perturbation, (b) prominence, and (c) position conditions.

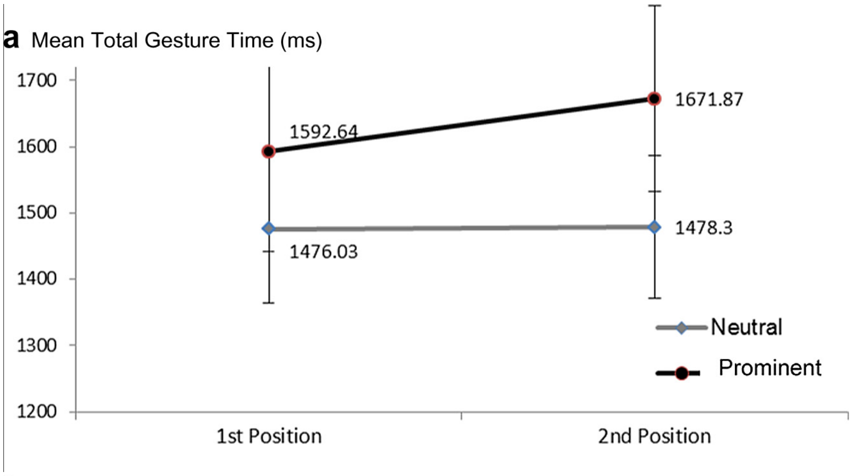
for DAF trials compared to NAF trials, as well as for target syllables with contrastive prominence compared to neutral syllables. Speech perturbations not only elongated the spoken responses, but lengthened total gesture time and gesture launch time, albeit not significantly. Contrary to expectations, speech perturbations did significantly lengthen GLM-VM intervals compared to the NAF conditions. In other words, the time between the midpoint of the gesture launch and of the target vowel did not remain tightly coordinated in the face of perturbation. Although, the gesture took longer to execute as a result of the temporal change of the speech system, the anchor points of gesture launch midpoint and vowel midpoint did not demonstrate continued temporal alignment during perturbation, contrary to previous descriptive accounts (McNeill, 1992; Mayberry and Jaques, 2000; Mayberry, 1998).

Mean total gesture time and gesture launch time were also significantly longer for utterances produced with contrastive pitch accent and as predicted GLM-VM intervals were shorter for the contrastive condition than the neutral condition. Therefore, the data indicated greater coordination of deictic gestures with prominent syllables similar to prior research (de Ruiter, 1998; Esteve-Gibert and Prieto, 2011; Leonard and Cummins, 2011; Rochet-Capellan et al., 2008; Rusiewicz et al., 2013) and theoretical postulations (Port, 2003; Rusiewicz, 2011). Though syllable position was initially included as an independent variable to control for preferred stress patterns in American English and to explore effects of prominence, rather than simply explaining the findings as a result of the lexical onset, a number of intriguing findings and interactions related to second position syllables emerged. Total gesture time was significantly longer for second position syllables than first position syllables, with the greatest mean gesture time observed for prominent syllables in the second position.

4.2. Perturbation findings, implications, and limitations

This was the first study to systematically investigate the effect of speech perturbation upon affiliated gestures, though the fundamental prediction that speech and gesture would be coordinated when speech was perturbed was not upheld. In fact, the opposite finding was noted with speech and gestures reduced in coordination when influenced by perturbation as measured by the interval between gesture launch midpoint and vowel midpoint. However, gestures were indeed temporally altered (i.e., gestures lengthened as the spoken response lengthened) as a function of speech perturbation. Qualitative changes in total gesture time and gesture launch time suggest that there is potentially some interaction and feedback between the two motor systems.

As borrowed from established literature on sensory feedback on speech movements (see Smith, 1992 for review) the speech system was perturbed in order to test the theoretical prediction that a disruption of the spoken production will also disrupt manual gesture production secondary to the coupling of these two oscillatory systems. However, the data do not yield straight-forward conclusions regarding the theoretical level of interaction between speech and gesture. If one only considers the finding that the intervals between gesture launch midpoints with vowel midpoints increased with perturbation rather than remaining coordinated in time, then a linguistic-based account such as the Sketch Model (de Ruiter, 1998, 2000) seems to be supported. To elaborate, the data may suggest that the formulation, planning, and programming of a gesture


a Mean Total Gesture Time (ms)

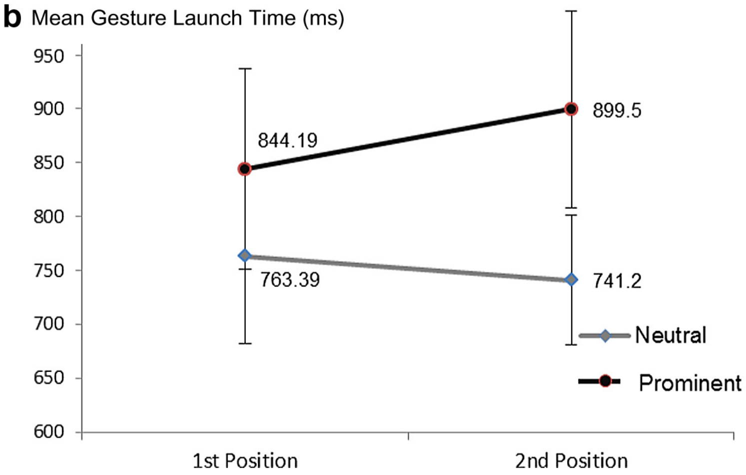

b Mean Gesture Launch Time (ms)

Fig. 5. Interaction effects for (a) total gesture time and (b) gesture launch time (ms). A significant two-way interaction of *position x contrast* is observed for both factors. Error bars correspond to one standard deviation.

and its lexical affiliate are initiated simultaneously in the early stage of conceptualization, then interaction of the two systems ceases during the later formulation/planning phases.

If gesture and speech cease interaction before formulation and planning of speech and gesture, then speech perturbation should not affect the execution of the gesture because there is no feedback mechanism from the formulator to the gesture planner. Therefore, even though speech is perturbed and consequently lengthened in the presence of DAF, the gesture continues uninterrupted because the gesture system has not received feedback that the lexical affiliate will be reached later. As a result, the interval between a gesture launch midpoint and the target vowel would be longer for DAF trials than for NAF trials, as observed in this study. However, there are several arguments to be made that may still suggest dynamic motor-level interactions in the form of temporal entrainment rather than a linguistic, linear model.

First, it is possible that there is indeed interaction between the two systems at or below a point of formulation, but that the planning of a deictic gesture is unique and for more simplistic than other gestures in its level of convention and automatic planning, particularly in the recurring task requirement of this experiment. A deictic gesture motor program is arguably simpler and more rote than an arbitrary and idiosyncratic iconic gesture, potentially leading to a deictic gesture being less susceptible to changes in its execution, namely time. An iconic gesture may be more likely to exhibit changes in the gesture launch period as well as holds within the movement to maintain coordination.

Differences in planning between gesture types may also reconcile the findings between this experiment and McNeill (1992) exploratory study of the effect of DAF on gesture production. In contrast to the controlled paradigm and use of deictic gestures in the present research, McNeill first had individuals spontaneously narrate a cartoon. McNeill stated that “gesture is still synchronous with the coexpressive part of the utterance” (p. 275). Interestingly, McNeill did not observe the same synchrony, albeit vaguely described in his experiments, in a second protocol. In the second experiment, participants were trained to recite memorized utterances while executing a series of prescribed fluid, continuous, iconic gestures. Even though the gestures were iconic, they were perhaps accessed and planned differently because they were set and memorized in their form, rather than created spontaneously as in McNeill’s first experiment. Not only was the automaticity and constraint of the responses similar to the current study, the results were as well. In contrast to McNeill's first study in which the spontaneous gestures remained temporally aligned with the spoken targets, there was noted asynchrony in the second experiment. Consequently, the dissociation between spontaneous and controlled gestures may explain the differences between the findings of the present experiment and Mayberry and colleagues' (Mayberry and Jaques, 2000; Mayberry, 1998) descriptions of perceived cessation and retained synchronization of gesture during stuttering for children and adults while spontaneously describing a cartoon. In short, the availability of feedback to and from the speech system may be different for automatic, conventional gestures compared to individually specified, spontaneous gestures. Natural, spontaneous production of all types of gestures in future empirical work could further explore these disparate findings.

A second possible argument that motor entrainment of speech and gesture is a viable theoretical construct is that DAF may not be the appropriate method of perturbation. For instance, there is a wide range of effects of DAF upon typical speakers (Smith, 1992). Even though the MTE was positive for all fifteen participants, there was a wide range of mean durational difference between NAF and DAF trials (i.e., 120–2346 ms). Therefore, it is possible that individuals with greater speech disruptions when faced with DAF may also be differentially susceptible to gesture disruptions. Though post-hoc analysis of the correlation of sentence duration mean time error and gesture launch time mean time error was not significant ($r = .42$; $p = .13$). Future investigations could screen individuals for the susceptibility to DAF to assure speech is adequately perturbed or study the coordinative effects of perturbation across manual and speech movements without the additional manipulation of other variables.

Moreover, DAF is a continuous and relatively slow perturbation method and may not be appropriate for studying the effect of speech perturbation on gesture timing. Specifically, the feedback mechanism shared between the gesture and speech systems may not be continuous as suggested by Levelt et al. (1985) in their study of perturbed wrist movements during a pointing and two-word speech response (i.e., *this light*, *that light*). Speech and gesture were temporally aligned when no load was applied and when the load was applied halfway through the gesture. Conversely, speech was perturbed only when a load was applied to the wrist at the beginning of a deictic gesture. Levelt et al. summarized, "the experimental findings show… that once the pointing movement had been initiated, gesture and speech operate in almost modular fashion… there is nevertheless, evidence to suggest that feedback from gesture to speech can come into play during the first milliseconds of gesture execution" (p. 162).

Evidence for the significant effects of early compared to late perturbation of oral movements was also provided by Gracco and colleagues (e.g., Abbs and Gracco, 1984). Continuous DAF does not enable exploration of more specific points of sensorimotor interactions between the speech and manual systems. Future investigations could perturb speech in a precise, fleeting moment by imposing transient mechanical perturbations of the articulators, such as the mandible (e.g., Shaiman, 1989) or lower lip (e.g., Abbs and Gracco, 1984).

Moreover, investigating the effects of gesture perturbation by extending Levelt and colleagues' work or by instituting new methodology and measures (Levelt et al., 1985) can lend insight on the interactive nature of speech and gesture. One methodology for studying deictic gestures that stands out as both valid and feasible is the double-step perturbation paradigm (e.g., Prablanc and Martin, 1992; Bard et al., 1999). This paradigm is used to examine goal-directed movement, such as reaching. Instead of imposing a direct load to the limb, the movement of the arm/hand is perturbed by changing the location of a target on a visual display. Not only is the trajectory of the arm/hand changed, but there is also an observed hesitation and lengthening of limb movement during reaching tasks. A double-step perturbation paradigm may be modified for a simultaneous speech and gesture task with systematic changes in the visual display. For example, the time of target change could differ between no shift of target, early shift, and later shift.

A third argument for continued consideration of temporal entrainment based upon the current perturbation manipulations are the resultant durational changes of deictic gestures. There were increases in overall duration of the complete gesture and gesture launch time for perturbed trials. Indeed the differences in duration for NAF and DAF conditions were not significant, but they were consistent. Notably, it is quite possible that a larger sample size may be necessary to detect the effect of DAF upon the duration of gesture.

4.3. Prosodic prominence and position findings, implications, and limitations

Certainly additional empirical investigation of the effects of perturbation on corresponding speech and gestures will illuminate potential mechanisms of entrainment. However, the import of prosody in the coordination of the rhythmic oscillators of speech and manual movements, such as gestures, cannot be underestimated or overlooked. One of the earliest proposals regarding the temporal relationship of speech and gesture is that manual and even other body movements co-occur with the suprasegmental properties of speech such as intonation and rhythm (Birdwhistell, 1952; Birdwhistell, 1970; Kendon, 1972; Kendon, 1980) and that movements of the body synchronize with speech (Condon and Ogston, 1966; Dittman and Llewellyn, 1969). The current data add further support to these early and continued hypotheses regarding the relationship of prosody and manual movements. Accordingly, prosodic prominence affected the timing of deictic gestures similar previous empirical work (de Ruiter, 1998; Esteve-Gibert and Prieto, 2011; Rochet-Capellan et al., 2008; Rusiewicz et al., 2013; Bull and Connelly, 1985; Loehr, 2004; Loehr, 2007; McClave, 1994; McClave, 1998; Nobe, 1996; Yasinnik et al., 2004) and in support of theories of temporal entrainment of speech and manual movements in adults (Port, 2003; Rusiewicz, 2011). Though the perturbation findings did not overwhelmingly support tight entrainment, the combination of those suggestive, though flawed, data and the import of pulses within the speech stream predicting the temporal execution of gestures is certainly more in line with motor based accounts of speech and gesture interaction, namely internal pulse-based entrainment of the two systems.

Interestingly, there was an effect of contrastive pitch accent in the predicted direction as indicated by the interval between gesture launch midpoint and vowel midpoint. GLM-VM intervals were shorter on average for contrastive trials than for neutral trials and also shorter for first position syllables than second position syllables, indicating increased coordination of gestures and prominent syllables. In fact, review of the descriptive data for the transformed dataset showed that shortest intervals were noted for prominent first position syllables without auditory delay while the largest intervals were noted for neutral second position syllables produced with DAF. Whereas, gesture launch times were longer for trials with contrastive pitch accent and were longest for accented second position syllables. Total gesture time was also significantly longer for prominent second position syllables but not for prominent first position syllables. These increased gesture durations demonstrate that the time to execute a gesture took longer when a syllable not only was produced in the later of the two positions, but also when produced with the longest vowel duration across all contrastive and syllable position conditions (i.e., second position accented syllables). In other words, second position accented syllables were the latest and longest in duration and so too were the gesture launch and total gesture time, providing evidence that the gesture movement is pulsing with prominent syllables, especially in the second syllable position, though may not always be aligned with specific anchor points within the speech and gesture movements.

The findings that gesture launch and total gesture times were significantly different for second position stressed syllables offer support for a mechanism of lower-level entrainment of the speech and gesture, such as that proposed by Tuite (1993) and Iverson and Thelen (1999). If there were no entrainment and the motor processes proceeded according to a purely linguistic account that views gesture and speech more from a linear perspective and less from a coordinative, rhythmic, system of movement, then one would anticipate that that the GLM-VM intervals would be shortest for first position syllables regardless of stress assignment and that gesture times would be consistent and not change as a function of contrast, position, or perturbation.

Continuing this line of reasoning, it is possible that gestures coincided with the second position pitch accented syllables because they were produced with greater duration and possibly greater motoric effort than their first syllable position counterparts. In other words, the *pulse* in the speech stream would be the vowel with the greatest duration and/or effort as a result of motoric behaviors. One could argue that the exaggerated contrastive pitch accent on the second position syllables in this rote, repetitive, and constrained task were the only syllables that were strong enough attractors to entrain the corresponding pulse (i.e., gesture launch) of the deictic gesture consistently.de Ruiter (1998, p. 61) also posed such an explanation for his finding that pointing gestures co-varied in time with peak syllables in a contrastive stress task. He stated:

> When peak syllables are produced in speech, the outgoing airflow is maximal. The motor system could therefore plan the moment of maximal exhalation to co-occur with the outward motion of the hand. Since the moment of maximal exhalation varies with the location of the peak syllable, the phonological synchrony rule might be explained by a phase locking mechanism at the level of motor planning, instead of a higher level synchronization process.

Syllables with longer durations typical require more exhalation than syllables that are shorter. For example, Tuller et al. (1982a) demonstrated that acoustic durations were longer and muscular activity for a myriad of oral muscles was longer and higher in amplitude for stressed vowels compared to unstressed vowels.

The simultaneous occurrence of a gesture may actually increase this oral motor effort according to a recent series of experiments by Krahmer and Swerts (2007). Speakers were asked to repeat *Amanda went to Malta* in a variety of conditions. The participants were instructed to either stress *Aman'da* or *Mal'ta* while at the same time producing either a manual beat gesture, an eyebrow movement, or a head nod. These visual beats were produced either congruently or incongruently with the target stressed syllable. As one would expect, the perception of prominence was enhanced when a visual beat occurred on the target syllable. However, the production of a beat changed the acoustic parameters of the vowel, even when the individuals were instructed to produce a beat on an unstressed syllable. Vowels were lengthened when a beat gesture was produced at the same time, even when the speaker was consciously attempting to produce the vowel as unstressed.

This distinct difference in the production parameters of stressed and unstressed syllables, in this case duration, is further supported by literature examining the physiologic differences of the production of iambic versus trochaic items. Kinematic research on speech production conducted by Goffman and Malin (1999) demonstrated that iambs were distinct from trochees in a surprising way. Iambs (e.g., *puhPUH*) were more stable and displayed high-amplitude modulation for both preschool aged children and young adults. Children were more likely to produce movements that were indistinct across the two syllables for the trochaic forms. In other words, children produced the strong-weak items more like strong-strong items while they distinctly produced modulated weak-strong patterns for the iambic forms. Goffman and Malin conjectured that "in trochees children can rely on existing resources and routines of the general motor system, thereby minimizing the degree of reorganization or modification required…the production of iambs, however clearly violates this basic organization" (p. 1013).

Although the participants in the present study were young adults, Goffman and Malin's hypothesis could apply to these findings. Rather than a developmental process necessitating the increased modulation of prosodic prominence for iambs, the individuals in the current study increased the modulation of contrastive pitch accent for trochees that would not typically be represented and programmed organized in that fashion with stress on the second syllable.

This mechanism also could unite the present data set and the findings of Rochet-Capellan and colleagues (2008). In their study, the deictic gesture apex was synchronized with the maximum jaw displacement first syllable when stressed and the return gesture was synchronized with the maximum jaw displacement when the second syllable was stressed. There are two points of interest in relating Rochet-Capellan et al.'s study to this experiment. First, the bisyllabic nonwords, *papa* and *tata*, employed as stimuli by Rochet-Cappelan et al. are not only more simplistic than the stimuli of Experiment 1, but arguably they were not manipulating prosody from a phonological encoding standpoint, rather they were instructing the participants to increase their motoric effort on a given syllable. Thus, if speech and gesture synchronize due to entrainment of the two motor systems, it is not a surprise that Rochet-Capellan et al. found evidence for synchrony of deictic gestures and syllables with maximum displacement, and possibly greater exhalation. In the present study, individuals lengthened the pulse of the manual movement (i.e., gesture launch) to correspond with the later, lengthened, and markedly prominent pulse within the speech stream (i.e., accented second position syllable).

Second, participants produced the bisyllabic stimuli in isolation in the investigation conducted by Rochet-Capellan and others. One would expect the second syllables to be produced with longer duration due to final syllable lengthening. However, this was not the case. When the syllables were stressed, the first position syllable jaw opening was longer (183 ms) than then second position syllable (144 ms). The stressed syllables were always longer in duration than the unstressed syllables. Thus, Rochet-Capellan and others found evidence that deictic gestures synchronize with syllables of increased motoric effort and perhaps Experiment 2 did so as well.

Continuing this line of reasoning, it is possible that gestures coincided with the second position pitch accented syllables because they were produced with greater duration due to increased prominence and arguable greater motoric effort than their first syllable position counterparts. One could argue that the relatively exaggerated contrastive pitch accent on the typically unstressed second position syllables in this rote, repetitive, and constrained task were the only syllables that were strong enough attractors to entrain the corresponding pulse (i.e., gesture launch) of the deictic gesture consistently. Furthermore, the influence of gesture and speech may be a "two-way street" (Rochet-Capellan et al., 2008). The simultaneous occurrence of a gesture may actually increase the affiliated pulse and corresponding effort of the speech system similar to a recent series of experiments by Krahmer and Swerts (2007). Speakers were asked to repeat *Amanda went to Malta* in a variety of conditions. Target vowels were lengthened when a beat gesture was produced at the same time, even when the speaker was consciously attempting to produce the vowel as unstressed. Thus, one can conjecture that the second position syllables were also more prominent than first position accented syllables *because of* the gesture that was produced in alignment with the second position syllables.

4.4. Limitations and directions for future research

The results of the study are intriguing and suggestive of internal pulse-based temporal entrainment of the speech and gesture motor systems. Still, several limitations of the current methodology and analyses beyond those identified already motivate more research in this area. As discussed in earlier in this discussion, the limitations of the study primarily center on the measurement of the speech and gesture signal and the required response. For instance, vowel midpoint and gesture launch midpoint were chosen as anchor points to investigate the coordination of these two rhythmic oscillators as motivated by previous data, theoretical thought, and the confines of the present methodology. Though gesture launch midpoint was a straightforward temporal point that was found to align with prosodically prominent syllables, future investigations should make use of more sensitive methodology such as a high-speech optical motion capture system to allow for kinematic measures on the resultant movement data. Measures of peak velocity and the variability of movements (i.e., discrete phase measures) will better elucidate the coordination of the speech and manual systems. Using such methods and analyses can also explore the rhythmic coordination of speech and manual movements while retracting from the assumption that two anchor points must be absolutely synchronized to be entrained.

This study, like earlier work (Rusiewicz et al., 2013), was unique in that a controlled protocol was employed while incorporating multiword, more natural responses compared to prior experiments assessing the relationship of prosody and gesture. Still, these data cannot reflect spontaneous speech and gestures, or different gesture types. It is particularly important that future investigations focus on beat gestures given the long-standing assumption that beat gestures are timed to the rhythmic aspect of speech (e.g., McNeill, 1992).

Without a doubt, it is imperative to consider the tremendous and vast impact of cognitive-linguistic processes upon the planning and execution of gestures. On the other hand, it is also important to strip down the gesture movement, to being just that - a manual movement, with the potential for overlaid cognitive and linguistic function. The dynamic coordination of speech production (i.e., typically single repetitive syllables) and manual movements, most often finger tapping observed in controlled paradigms (Smith et al., 1986; Chang and Hammond, 1987; Hiscock and Chipuer, 1986) may be extended to examining gestures. Such studies could also offer insight on the more natural and ubiquitous use of manual movements (i.e., gestures) during communication as a reflection of being two coordinated internal coupled oscillators. Lastly, future investigations and refined models of entrained manual and speech systems will lend insight to the interface of biomechanic and psycholinguistic processes in typical populations. Importantly, much like our understanding of the linguistic interaction of gesture and speech has positively influenced the management of those with language disorders such as aphasia, examination of the internal spatiotemporal entrainment of speech and gesture will also expand our virtually nonexistent empirical literature regarding the impact of manual movements frequently employed in the management of children and adults with speech disorders (Peter and Raskind, 2011).

Acknowledgements

This project was supported by the University of Pittsburgh School of Rehabilitation and Health Sciences Development Fund. We are grateful for the time of the participants and notable contributions to all stages of this work by Christine Dollaghan, Thomas Campbell, J. Scott Yaruss, and Diane Williams. We also thank Jordana Birnbaum, Megan Pelletierre, and Alyssa Milloy for assistance in data analysis and entry and Megan Murray for assistance in the construction of this manuscript. Portions of this work were presented at the 2009 and 2010 Annual Convention of the American-Speech-Language-Hearing Association and the 2011 conference on Gesture and Speech in Interaction (GESPIN).

Appendix A.

Compound word stimuli.

Stimuli set 1: second syllable contrastive prominence		Stimuli set 2: first syllable contrastive prominence	
Bathrobe	Bathtub	Suitcase	Briefcase
Football	Footprint	Jukebox	Mailbox
Grapevine	Grapefruit	Birdhouse	Doghouse
Icecube	Icecream	Toothbrush	Paintbrush
Lifeboat	Lifeguard	Football	Baseball
Blackbird	Blackboard	Bluebird	Blackbird
Toothpaste	Toothbrush	Lighthouse	Greenhouse
Thumbtack	Thumbprint	Cupcake	Pancake
Birdcage	Birdbath	Stoplight	Flashlight
Lighthouse	Light bulb	Footprint	Handprint
Eggnog	Eggplant	Keyboard	Surfboard
Fishbowl	Fishhook	Notebook	Matchbook
Seagull	Seahorse	Wheelchair	Highchair
Snowball	Snowflake	Horseshoe	Snowshoe
Teapot	Teacup	Hottub	Bathtub

References

Abbs, J.H., Folkins, J.W., Sivarajan, M., 1976. Motor impairment following blockade of the infraorbital nerve: Implications for the use of anesthetization techniques in speech research. J. Speech Hear. Res. 19, 19–35.

Abbs, J.H., Gracco, V.L., 1982. Evidence for speech muscle functional compartmentalization: Theoretical and methodological implications. J. Acoust. Soc. Am. 71, S33–S34.

Abbs, J.H., Gracco, V.L., 1984. Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech. J. Neurophysiol. 51, 705–723.

P.A. Barbosa, Explaining cross-linguistic rhythmic variability via coupled-oscillator model of rhythm production, Paper presented at the Speech Prosody, Aix-en-Provence, France, 2002.

Bard, C., Turrell, Y., Fleury, M., Teasdale, N., Lamarre, Y., Martin, O., 1999. Deafferentation and pointing with visual double-step perturbations. Exp. Brain Res. 125, 410–416.

Birdwhistell, R.L., 1952. Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture. University of Louisville Press, Louisville.

Birdwhistell, R.L., 1970. Kinesics and Context: Essays on Body Motion Communication. University of Pennsylvania Press, Philadelphia.

Bluedorn, A.C., 2002. The Human Organization of Time: Temporal Realities and Experience. Stanford University Press, Stanford, CA.

Bull, P., Connelly, G., 1985. Body movement and emphasis in speech. J. Nonverbal Behav. 9, 169–187.

Burke, B.D., 1975. Susceptibility to delayed auditory feedback and dependence on auditory or oral sensory feedback. J. Com. Dis. 8, 75–96.

Chang, P., Hammond, G.R., 1987. Mutual interactions between speech and finger movements. J. Motor Behav. 19, 265–275.

Clayton, M., Sager, R., Will, U., 2004. In time with the music: the concept of entrainment and its significance for ethnomusicology. ESEM Counterpoint 1, 1–82.

Condon, W.S., Ogston, W.D., 1966. Sound film analysis of normal and pathological behavior patterns. J. Nerv. Mental Dis. 143, 338–347.

Corriveau, K.H., Goswami, U., 2009. Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. Cortex 45, 119–130.

Cummins, F., Port, R.F., 1996. Rhythmic commonalities between hand gestures and speech. In: Proceedings of the Eighteenth Meeting of the Cognitive Science Society, Lawrence Erlbaum, London, pp. 415–419.

Cummins, F., Port, R.F., 1998. Rhythmic constraints on stress timing in English. J. Phonetic 26, 145–171.

Cutler, A., Carter, D.M., 1987. The predominance of strong initial syllables in the English vocabulary. Comput. Speech Lang. 2, 133–142.

de Ruiter, J.P., 1998. Gesture and Speech Production, Unpublished Doctoral Dissertation. Katholieke Universiteit, Nijmegen, Germany.

de Ruiter, J.P., 2000. The production of gesture and speech. In: McNeill, D. (Ed.), Language and Gesture. Cambridge University Press, Cambridge, UK, pp. 284–311.

Dittman, A.T., Llewellyn, L.G., 1969. Body movement and speech rhythm in social conversation. J. Pers. Soc. Psychol. 11, 98–106.

Elman, J.L., 1983. Oral vs. manual tapping with delayed auditory feedback as measures of cerebral dominance. J. Speech Lang. Hear. Res. 26, 106–110.

Esteve-Gibert, N., Prieto, P., 2011. Temporal alignment of gesture and speech in Catalan-babbling children. In: Proceedings from the 2nd Gesture and Speech in Interaction (GESPIN), Bielefield, Germany. Retrieved from: <http://coral2.spectrum.uni-bielefeld.de/gespin2011/final/Esteve-Gibert_Prieto.pdf>.

Franks, I.M., Nagelkerke, P., Ketelaars, M., van Donkelaar, P., 1998. Response preparation and control of movement sequences. Can. J. Exp. Psych 52, 93–101.

Folkins, J.W., Abbs, J.H., 1975. Lip and jaw motor control during speech: Responses to resistive loading of the jaw. J. Sp. Hear. Res. 18, 207–220.

Folkins, J.W., Zimmermann, G.N., 1981. Jaw-muscle activity during speech with the mandible fixed. J. Acoust. Soc. Amer. 69, 1441–1445.

Franz, E.A., Zelaznik, H.N., Smith, A., 1992. Evidence of common timing processes in the control of manual, orofacial, and speech movement. J. Motor Behav. 24, 281–287.

Gentilucci, M., Benuzzi, F., Gangitano, M., Grimaldi, S., 2001. Grasp with hand and mouth: Kinematic study on healthy subjects. J. Neurophysiol. 86, 1685–1699.

Gentilucci, M., Santunione, P., Roy, A.C., Stefanini, S., 2004. Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. Eur. J. Neurosci. 19, 190–202.

Goffman, L., Malin, C., 1999. Metrical effects on speech movements in children and adults. J. Sp. Lang. Hear. Res. 42, 1003–1015.

Goldin-Meadow, S., 1999. The role of gesture in communication and thinking. Trends Cogn. Sci. 3, 419–429.

Gracco, V.L., Abbs, J.H., 1985. Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. J. Neuropath. 54, 418–432.

Gracco, V.L., Abbs, J.H., 1986. Variant and invariant characteristics of speech movements. Exp. Brain. Res. 65, 155–166.

Guenther, F.H., Hampson, M., Johnson, D., 1998. A theoretical investigation of reference frames for the planning of speech movements. Psych. Rev. 105, 611–633.

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

Hiscock, M., Chipuer, H., 1986. Concurrent performance of rhythmically compatible or incompatible vocal and manual tasks: evidence for two sources of interference in verbal–manual timesharing. Neuropsychologia 24, 691–698.

Howell, P., Archer, A., 1984. Susceptibility to the effects of delayed auditory feedback. Percepti. Psychophysics. 36, 296–302.

Howell, P., Sackin, S., 2002. Timing interference to speech in altered listening conditions. J. Acoust. Soc. Am. 111, 2842–2852.

Iverson, J.M., Thelen, E., 1999. Hand, mouth and brain: the dynamic emergence of speech and gesture. J. Consciousness Stud. 6, 19–40.

Jones, M.R., Boltz, M., 1989. Dynamic attending and response to time. Psychol. Rev. 96, 459–491.

Kelso, J.A.S., Tuller, B., Harris, K.S., 1981. A "dynamic pattern" perspective on the control and coordination of movement. In: MacNeilage, P.F. (Ed.), The Production of Speech. Springer-Verlag, New York, pp. 137–173.

Kelso, J.A., Tuller, B., 1983. "Compensatory articulation" under conditions of reduced afferent information: A dynamic formulation. J. Sp. Hear. Res. 26, 217–224.

Kelso, J.A., Tuller, B., 1984. Converging evidence in support of common dynamical principles for speech and movement coordination. Amer. J. Physio.: Regulat. Integrat. Compar. Physio. 246, R935–R938.

Kelso, J.A., Tuller, B., Vatikiotis-Bateson, E., Fowler, C.A., 1984. Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. J. Exper. Psych.: Hum. Percept. Perf. 6, 812–832.

Kendon, A., 1972. Some relationships between body motion and speech. In: Seigman, A.W., Pope, B. (Eds.), Studies in Dyadic Communication. Pergamon, Elmsford, NY, pp. 177–210.

Kendon, A., 1980. Gesticulation and speech: Two aspects of the process of utterance. In: Key, M.R. (Ed.), The Relationship of Verbal and Nonverbal Communication Mouton, The Hague, the Netherlands, pp. 207–227.

Komilis, E., Pelisson, D., Prablanc, C., 1993. Error processing in pointing at randomly feedback induced double step stimuli. J. Motor Behav. 25, 299–308.

Krahmer, E., Swerts, M., 2007. The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. J. Mem. Lang. 57, 396–414.

Krauss, R.M., Chen, Y., Gottesman, R.F., 2000. Lexical gestures and lexical access: a process model. In: McNeill, D. (Ed.), Language and Gesture. Cambridge University Press, Cambridge, UK, pp. 261–283.

Large, E.W., Jones, M.R., 1999. The dynamics of attending: how people track time varying events. Psychol. Rev. 106, 119–159.

Leonard, T., Cummins, F., 2011. The temporal relation between beat gestures and speech. Lang. Cogn. Proc. 26, 1457–1471.

Levelt, W.J.M., Richardson, G., La Heij, W., 1985. Pointing and voicing in deictic expressions. J. Mem. Lang. 24, 133–164.

Liberman, A.M., Whalen, D.H., 2000. Of the relation of speech to language. Trends Cog. Sci. 4, 187–196.

D.P. Loehr, Gesture and Intonation: Doctoral dissertation, Georgetown University, Dissertation Abstracts International, 65 (2004) 2180. (UMI No. 3137056).

Loehr, D.P., 2007. Aspects of rhythm in gesture and speech. Gesture 7, 179–214.

Lorenzo, S., Diederich, E., Telgmann, R., Schutte, C., 2007. Discrimination of dynamical system models for biological and chemical processes. J. Comput. Chem. 28, 1384–1399.

Marslen-Wilson, W.D., Tyler, L.K., 1981. Central processes in speech understanding. Philos. Trans. R. Soc., Ser. B 295, 317–332.

Mayberry, R.I., Jaques, J., 2000. Gesture production during stuttered speech: insights into the nature of gesture-speech integration. In: McNeill, D. (Ed.), Language and Gesture. Cambridge University Press, Cambridge, pp. 199–214.

Mayberry, R.I., Jaques, J., DeDe, G., 1998. What stuttering reveals about the development of gesture speech relationship. New Direction Child Dev. 79, 77–87.

McClave, E., 1994. Gestural beats: the rhythm hypothesis. J. Psycolinguist. Res. 23, 45–66.

McClave, E., 1998. Pitch and manual gestures. J. Psycolinguist. Res. 27, 69–89.

McNeill, D., 1992. Hand and Mind. University of Chicago Press, Chicago.

Merker, B., Madison, G., Eckerdal, P., 2009. On the role and origin of isochrony in human rhythmic entrainment. Cortex 45 (2009), 1–17.

S. Nobe, Representational gestures, cognitive rhythms, and acoustic aspects of speech: a network/threshold model of gesture production: Doctoral dissertation, University of Chicago, Dissertation Abstracts International, 57 (1996) 4736. (UMI No. 9636827).

O'Dell, M., Nieminen, T., 1999. Coupled oscillator model of speech rhythm. In: Ohala, J.J., Hasegawa, Y., Ohala, M., Granville, D., Bailey, A.C. (Eds.), . In: Proceedings of the XIVth International Congress of Phonetic Sciences, 2. American Institute of Physics, New York, pp. 1075–1078.

Paulignan, Y., Jeannerod, M., MacKenzie, C., Marteniuk, R., 1991. Selective perturbation of visual input during prehension movements. Exp. Brain Res. 87, 407–420.

Peter, B., Raskind, W.H., 2011. Evidence for a familial speech sound disorder subtype in a multigenerational study of oral and hand motor sequencing ability. Top. Lang. Dis. 31, 145–167.

Port, R.F., 2003. Meter and speech. J. Phonetic 31, 599–611.

Port, R., Tajima, K., Cummins, F., 1998. Speech and rhythmic behavior. In: Savelsburgh, G.J.P., van der Maas, H., van Geert, P.C.L. (Eds.), The Nonlinear Analysis of Developmental Processes. Royal Dutch Academy of Arts and Sciences, Amsterdam, pp. 5–45.

Prablanc, C., Martin, O., 1992. Automatic control during hand reaching at undetected two-dimensional target displacements. J. Neurophysiol. 67, 455–469.

Ringel, R.L., Steer, M.D., 1963. Some effects of tactile and auditory alterations on speech output. J. Sp. Hear. Res. 13, 369–378.

Rochet-Capellan, A., Laboissierre, R., Galvan, A., Schwartz, J., 2008. The speech focus position effect on jaw-finger coordination in a pointing task. J. Speech Lang. Hear. Res. 51, 1507–1521.

Rusiewicz, H.L., 2011. Synchronization of speech and gesture: a dynamic systems perspective. In: Proceedings from the 2nd Gesture and Speech in Interaction (GESPIN), Bielefield, Germany. Abstract retrieved from <http://coral2.spectrum.uni-bielefield.de/gespin2011/final/rusiewicz.pdf>.

Rusiewicz, H., Shaiman, S., Iverson, J., Szuminsky, N., 2013. Effects of prosody and position on the timing of deictic gestures. J. Speech Lang. Hear. Res. 56, 458–470.

Saltzman, E., Byrd, D., 2000. Task-dynamics of gestural timing: Phase window and multifrequency rhythms. Hum. Movement Sci. 19, 499–526.

Saltzman, E.L., Munhall, K.G., 1989. A dynamical approach to gestural patterning in speech production. Ecolog. Psych. 1, 330–380.

Schmidt, R.A., Lee, T.D., 1999. Motor Control and Learning: A Behavioral Emphasis, third ed. Human Kinetics, Champaign, IL.

Shaiman, S., 1989. Kinematic and electromyographic responses to perturbation of the jaw. J. Acoust. Soc. Am. 86, 78–88.

Shaiman, S., Gracco, V.L., 2002. Task-specific sensorimotor interactions in speech production. Exp. Brain Res. 146, 411–418.

Shriberg, L.D., Campbell, T.F., Karlsson, H.B., Brown, R.L., McSweeney, J.L., Nadler, C.J., 2003. A diagnostic marker for childhood apraxia of speech: the lexical stress ratio. Clin. Linguist. Phonetics 17, 549–574.

Smith, A., 1992. The control of orofacial movements in speech. Crit. Rev. Oral Bio. Med. 3, 233–267.

Smith, A., McFarland, D.H., Weber, C.M., 1986. Interactions between speech and finger movements: an exploration of dynamic pattern perspective. J. Speech Lang. Hear. Res. 29, 471–480.

Stuart, A., Kalinowski, J., Rastatter, M.P., Lynch, K., 2002. Effect of delayed auditory feedback on normal speakers at two speech rates. J. Acoust. Soc. Am. 111, 2237–2241.

Thelen, E., Smith, L.B., 2002. A Dynamic System Approach to the Development of Cognition and Action. MIT Press, Cambridge, MA.

Tuite, K., 1993. The production of gesture. Semiotica 93, 8–105.

Tuller, B., Harris, K.S., Kelso, J.A., 1982a. Stress and rate: Differential transformation of articulation. J. Acoust. Soc. Amer. 71, 1534–1543.

Tuller, B., Kelso, J.A., Harris, K.S., 1982b. On the kinematics of articulatory control as a function of stress and rate. J. Acoust. Soci. Am. 72, S103.

Tuller, B., Harris, K.S., Kelso, J.A., 1983. Converging evidence for the role of relative timing in speech. J. Exp. Psych.: Human Percept. Perf. 9, 829–833.

Turvey, M.T., 1990. Coordination. Am. Psychol. 45, 938–953.

van Kuijk, D., Boves, L., 1999. Acoustic characteristics of lexical stress in continuous telephone speech. Speech Commun. 27, 95–111.

van Lieshout, P.H.H.M., 2004. Dynamic systems theory and its application in speech. In: Maassen, B., Kent, R., Peters, H., van Lieshout, P., Hulstijn, W. (Eds.), Speech Motor Control: In Normal and Disordered Speech. Oxford University Press, New York, pp. 51–82.

von Holst, E., 1973. Relative coordination as a phenomenon and as a method of analysis of central nervous system function. In: Martin, R. (Ed.), The Collected Papers of Erich von Holst: The Behavioral Physiology of Animal and Man. University of Miami Press, Coral Gables.

Włodarczak, M., Šimko, J., Wagner, P., 2012. Syllable boundary effect: temporal entrainment in overlapped speech. Paper presented at speech prosody. In: 6th International Conference, Shanghai, China.

Yasinnik, Y., Renwick, M., Shattuck-Hufnagel, S., 2004. The timing of speech-accompanying gestures with respect to prosody. In: Paper presented at the Meeting the Sound to Sense: 50+ years of Discoveries in Speech Communication, Cambridge, MA.