Synchronization of Prosodic Stress and Gesture: A Dynamic Systems Perspective

Heather Leavy Rusiewicz

Department of Speech-Language Pathology, Duquesne University, Pittsburgh, Pennsylvania, USA
rusiewih@duq.edu

Abstract

The temporal synchronization of spoken language production and gesture is often proposed as evidence of the integrative nature of these two systems. Intriguing but limited experimental investigation indicates that the synchronization of speech and gesture, and indeed other manual movements, is a function of prosodically prominent syllables in the speech stream. This paper presents existing data from a dynamic systems perspective and the postulation that speech and manual movements, including gestures, are aligned due to internal temporal entrainment of the two motor systems. Implications for the management of individuals with communication disorders are also discussed.

Index Terms: entrainment, prosody, gesture, dynamic systems, synchronization

1. Introduction

Gestures and spoken language co-occur in time, though the mechanism of this temporal coordination is yet unknown. In order for speech and gesture to synchronize in some uniform manner, there must be interaction between the speech and gesture production systems. Yet, few have hypothesized the precise parameters of coordination and even fewer have experimentally investigated the temporal coordination of the two systems. The hypothesis that is explored in the current paper is that the underlying mechanism of synchronization of speech and gesture is a function of internal temporal entrainment of the pulses planned and produced by the two motor systems. It is also proposed that examining the entrainment of speech and gesture is not only important for understanding typical behaviors, but also can impact the assessment and treatment of individuals with communication disorders. Thus, the objective of this manuscript is to meld and refine existing theoretical thought with relevant data to then propel a more predictive theory of entrained manual and speech systems to elucidate both typical and atypical communication processes.

1.1. Gesture theories and synchronization

Regardless of the theoretical approach that one adopts and despite the limited experimental data, there is general agreement that the perceived synchronization of speech and gesture is premier evidence that indeed the two systems are integrated (though see [1], [2], [3], [4], [5]). The traditional hypothesis is that gesture and speech share a seamless interaction at all points within their respective production mechanisms, resulting in a perceived tight temporal synchrony [6], [7]. However this hypothesis leads to nebulous points of interaction, thus making it difficult to make predictions regarding the effects of specific speech and gesture variables upon the timing of the other movement.

Others have posited more specific points of interaction of the two systems from a linguistic perspective. For instance, de Ruiter [8], [9] proposed a model of gesture production with an explicit point of interaction with the speech production system. In short, de Rutter’s Sketch Model is an extension of Levelt’s [10] model of spoken language production and predicts that speech and gesture originate and interact only at the stage of conceptualization when one accesses spatio-temporal and propositional information from working memory. According to de Ruiter, any phonological or motor processes that occur later in the speech system, such as assigning prosodic stress to a syllable, do not influence the timing of gesture. Although other linguistic-based models exist (e.g., [11]), they do not yet account for the role of prosody in the temporal synchronization of gesture and speech.

1.2. Motoric-based gesture theories

Conversely, it is also possible that the interaction between the speech and gesture production systems is not solely at a linguistic level, but also dependent upon motoric processes. Expressly, speech and gesture may synchronize due to a form of entrainment of the speech and manual motor systems rather than an interaction of higher-level phonological processes and the gesture production system. The Sketch Model, like the vast majority of modular, linguistic-focused production models, cannot account for such an interaction. However, even McNeill [7] postulated in his seminal work that gestural strokes coincide with prosodically stressed lexical items according to his rule of phonological synchrony. Though not widely considered in the gesture literature, there are theoretical accounts of entrainment contributing to the synchronization of speech and gesture as described by Tuite [12] and Iverson and Thelen [13].

Tuite’s [12] Rhythmic Pulse Hypothesis asserts that gesture and speech are linked prosodically and that gesture and speech originate from a kinetic base. Tuite argued that this kinetic base is “represented as a rhythmical pulse” (p. 99). The pulse peak corresponds to the stroke portion of the gesture and the intonational peak of spoken language. The pulse may be most simply expressed as a beat movement or may be overlaid with spatial or semantic properties and expressed as a deictic or iconic gesture. Even though Tuite did not directly situate his hypothesis within a dynamic systems perspective, the notions of rhythmical pulses and pulse peaks are very much in line with work in the dynamic systems literature [14], [15], [16], [17], [18], [19], [20].

The only other theory of gesture production rooted in motor behavior was presented by Iverson and Thelen [13]. These authors proposed an entrained system of speech and manual movements during early development that is rooted in dynamic systems theory. Iverson and Thelen 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). An important tenet of this hypothesis is that as the novelty and effort of a behavior decreases, the mutual
entrainment and degree of synchrony between the two effectors will increase.

Also, Iverson and Thelen [13] are among the only theorists to postulate a motoric level of interaction of the speech and gesture systems from a dynamic systems perspective, rather than a conceptual, lexical, or phonological level of interaction. Similarly, this is the only theory that makes predictions about the simultaneous development of the speech and gesture production systems, though the authors made no explicit statements about what affects the synchronization of speech and gesture produced by older children, adults, or those with impaired communication systems.

2. Evidence of entrained manual and speech movements

2.1. Gesture production

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 [21], [22], [23], [24] and that the body moves closely in time with speech (e.g., [25]). Thus, the idea that gestures, especially beat gestures, correspond to the prosody of spoken language is not new. It has long been accepted that beat gestures are tightly linked to stressed syllables and “move to the rhythmical pulsation of speech” ([7], p. 15). However, relatively few experimental investigations of the relationship of gestural movements and prosodic features of speech were conducted to date and little is known about the mechanism of this temporal relationship.

Ten investigations examined the effect of syllable prominence upon the timing of gesture [25], [26], [27], [28], [29], [30], [31], [32], [33], [34]. The majority of these investigators did not experimentally manipulate or control the spoken and gestural responses or measure the precise temporal parameters of the two movements [8], [25], [26], [27], [28], [29], [32]. Despite a number of methodological limitations and variations, all of these more naturalistic paradigms, with the exception of McClave’s earlier study [27], suggested a potential interaction between prosodic prominence and the timing of the associated gesture.

de Ruiter [8], de Ruiter, Capell, et al. [30], Rusiewicz [31], and Leonard and Cummins [33], [34] experimentally manipulated the assignment of prosodic stress and required a controlled response from their participants. Indeed, all found evidence for a coordination of gestures with prosodically stressed syllables, though the methodologies of these investigations varied greatly. For instance, the participants spoke different languages (e.g., Dutch [8], Brazilian Portuguese [30], and American English [31]). Additionally, each study used relatively novel measurement procedures in this area of research such as employing ultrasound [8], infrared diode tracking [30], [33], [34], and electrical capacitance [31]. Each of these studies examined only constrained deictic [8], [30], [31] or beat gestures [33], [34], but required different spoken responses consisting of bisyllabic canonical forms with varying stress (e.g., ‘/Papa/’ [30], two and three word responses with varying lexical or contrastive stress (e.g., the VIOLET crocodile; [8]), seven word responses with varying contrastive pitch accent (e.g., no, the light HOUSE is above the square; [31]), and story reading [33], [34]. Though de Ruiter did not find support for deictic gestures aligning with syllables with lexical stress, all other experiments did demonstrate a coordination of gestures and prosodically prominent syllables. Certainly, considerable work remains to fully examine and understand the synchronization of gestures and prosodic stress, nonetheless the growing literature base suggests that the temporal relationship of the two systems is at least in part dependent upon prominent points within the speech stream.

2.2. Manual movements

Gestures are not unique in their tendency to align with prominent pulses. Likewise, there are data from studies embedded in dynamic systems theory that also support a shared processing of speech and manual motor patterns, particularly for rhythmic movements. For instance, there is growing interest in the effect of external rhythms imposed by a metronome in speech cycling and synchronous speech tasks [35], [36], [37], [38] and for the influence of rhythm on the coordination of speech and manual movements within the field of music [39], [40], [41].

Twenty years ago, Kelso, Fullter, and Harris [42] claimed that disparate motor acts, like speech and finger movements, can be organized as a single coupled unit. Based on this earlier work by Kelso et al., Smith and colleagues [43] found support for both changes in amplitude and frequency of movement due to interactions between repetitive speech and finger tapping performed simultaneously. Eight individuals were instructed to repeat /stak/ and tap their fingers at alternating loud and soft volumes or large and small movements. Even though there was not absolute coordination when the participants were instructed to alter the magnitude of only one of the movements, a systematic change of frequency was observed, such that a frequency somewhere in between the two original frequencies was adopted. Thus, there was an interaction and entrainment between speech and manual movements. Subsequently, Franz, Zelaznik, and Smith [44] also demonstrated that the temporal patterns of repetitive movements of the mandible and single syllable repetition were significantly correlated to the repetitive productions of finger and forearm movements. There is limited experimental and theoretical work on this topic, though what exists is intriguing and holds promise for future investigations, including those employing gestures, not only finger tapping.

Similarly, the simultaneous occurrence of a manual pulse movement may actually increase oral motor effort according to a recent series of experiments conducted by Krahmer and Swerts [45]. Even though Krahmer and Swerts investigated beat gestures it is arguable that the constrained nature of the task was aligned with the finger tapping tasks described above. Dutch speakers were asked to repeat Amanda went to Malta in a variety of conditions. The participants were instructed to either stress A’MANnda or ‘MALta while producing 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. Interestingly, the production of a beat also changed the acoustic parameters of the vowel, even when the individuals were instructed to produce a beat on an unstressed syllable. In other words, individuals could not restrain from placing prosodic stress on a syllable when a manual movement was produced simultaneously.

Additional support for the entrainment of the systems comes from recent investigations that demonstrated tight spatiotemporal coordination of speech and manual movements, particularly for prominent pulses within the production systems. For instance, Gentilucci and colleagues
demonstrated that the size of grasped objects affects the size of mouth opening and the amplitude of syllable production [46], [47]. When a small object like a cherry was picked up by the hand, simultaneous syllable production was significantly less according to acoustic and kinematic measures, as compared to when an apple was picked up.

3. Dynamic systems theory and entrained speech and gesture systems

If gestures and other manual movements temporally co-occur with prosodically prominent syllables, as the existing literature implies, it is not clear why this alignment exists. The understanding of the mechanism of gestures aligning with prosodically stressed syllables remains where it was in 1969 when Dittmann and Llewellyn [48] stated “the really interesting question which this research has raised are those of why body movements should be located as they are in the rhythmic stream of speech” (p. 104). A dynamic systems perspective, particularly theories of entrainment of systems, holds potential explanatory power for the underlying mechanism of speech and gesture synchronization.

3.1. Understanding temporal entrainment

In recent decades, nonlinear science 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, motor control. However, principles of nonlinear dynamics are now being applied to just about any behavior or event, from cognitive-linguistic processing (e.g., [49]) to inanimate properties like biological and chemical processes (e.g., [50]). Most often the coupling of two effectors (i.e., fingers, limbs, lower lip, mandible, etc.) is studied and the temporal rhythms of the two also play a role in the stability of the behaviors. There are many classic examples that exemplify these somewhat abstract ideas including bimanual coordination of finger-thumb closure [51], finger-oscillation [51], finger tapping [52], and pendulum swinging [53]. It is thought that certain patterns of coordination, maximize the efficiency of the coordinative structures, termed oscillators, performing the behavior.

The oscillators are said to be entrained when they influence one another mutually to “produce a single coordinated behavior, synchronous in time and space” ([13], p. 28). Similarly, Telen and Smith (p. 304) 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 clear definition comes from Clayton, Sager, and Will [54] who defined entrainment as “a phenomenon in which two or more independent rhythmic processes synchronize with each other” (p. 1) and in “such a way that they adjust towards and eventually ‘lock in’ to a common phase and/or periodicity. Merker, Madison, and Eckerdal [55] used the more descriptive term of “pulse-based rhythmic entrainment” (p. 4) to describe this phenomena. Likewise, Phillips-Silver, Aktipis, and Bryant [41] provided a thorough, nonmathematical framework for understanding the concept of entrainment across a wealth of phenomena. 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 when they are produced in isolation.

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 [39], [54], [55]. That is, that the pattern of one behavior (e.g., beat gestures) co-occurs at certain points in time with the cycle of another behavior (e.g., pitch accented syllables in the speech stream). Bluedorn [39] 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 person or species is entrained to that of another oscillating event. Examples are limitless but include fireflies synchronizing illumination, synchronization applause patterns, and even parrots “dancing” to music [56], tapping (e.g., [57]) or repetitive phrase production along with a metronome is another example of external entrainment [37]. 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. For example breath groups tend to synchronize with ambulation patterns while jogging.

According to Port et al. ([36], 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 [20] later proposed that the rhythm of speech is temporally spaced according to regular periodic patterns generated by neurocognitive oscillators. These oscillators produce pulses, similar to Tuite’s view, 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). According to Port and as supported by the available empirical work, a salient perceptual event corresponds to such moments as the onset of a prominent vowel. Port states that “for English, this is especially true for syllables with pitch accent or stress” (p. 609).

The theoretical concept of pulse, be it Tuite’s [12] or Port’s [20] conception, holds a prosodically prominent (i.e., salient) syllable as the pulse that can then entrain other oscillators. Even though Port’s hypothesis is based upon work with speech cycling tasks (e.g., [35]) which require the simultaneous production of short phrases (e.g., Dig for a duck) with an external oscillator (i.e., a metronome), his broader intent is to explain self-entrainment of internal oscillators across systems. He summarizes, “these oscillations can be described as neurocognitive because they represent major neural patterns somewhere in the brain and are cognitive because they may be time-locked to events across many different modalities-audition, cyclic attention, speech, limb motion, etc.” (p. 609).

3.2. Developing a theory of speech and gesture entrainment

The current proposal is that speech and gesture are two internal, coupled oscillators according to this evolving Theory of Entrained Manual and Speech Systems (TEMSS) (Figure
1. By integrating the hypotheses put forth by Port [20], Tuite [12], and Iverson and Thelen [13] a more elaborative and predictive hypothesis of speech and gesture entrainment can be developed. It is proposed via the TEMSS that speech and gesture, and indeed other movements like finger tapping, are temporally synchronous due to entrainment of the two motor systems. It is also proposed that this entrainment occurs as a result of the self-entrainment of two internal coupled oscillators. That is the rhythmic 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 gestures. Continued exploration of the neurophysiologic basis of the proposed pulses is beyond the scope of the current discussion, though certainly warranted. Likewise, future discussion and assessment of the effect of entrainment of systems on neurocognitive processing will further elucidate this theoretical mechanism. For instance, the entrainment of prominent pulses potentially supports the “listener” as well as the “speaker” by enhancing attention and memory processes (e.g., [17], [18]).

![Figure 1: Self-entrainment of speech and gesture oscillators.](image)

The current discussion focuses on the theoretical spatiotemporal entrainment of the manual and speech systems. Specifically, the most salient portion of the accented syllable (e.g., vowel midpoint) entrains the most salient portion of the gesture (e.g., peak velocity of gesture stroke), resulting in the perceived temporal synchronization of speech and gesture. It can be further hypothesized that speech acts as the entraining pulse because it is a more continuous and rhythmic behavior than gesture. That is, speech is produced with an acceptable rhythmic structure and the gestures themselves are almost always transient and non-obligatory. Yet, it is certainly possible that the effects of coordination is a two-way street and that the motor pattern of gesture may also affect the spatiotemporal parameters of speech (e.g., [30], [45]).

Viewing the synchronization of speech and gesture from a motor entrainment perspective also explains prior research that revealed that not only are hand gestures synchronized with pitch accented syllables, but eyelinks, head movements, and even torso movements are as well [21], [22], [25], [26]. If there was a unique and absolute cognitive-linguistic relationship responsible for gesture-speech synchrony as posited by theorists like McNeill [7], de Ruiter [8], [9] and Krauss, Chen, & Gottesman [11], then such observations would be difficult to explain.

### 3.3. Predictions of TEMSS

Amalgamation of the theoretical work by Iverson and Thelen [13] Tuite [12], and Port [20] offers testable predictions for future empirical work on the entrainment of speech and gesture. Though rudimentary, some of the basic predictions of an integrated interpretation of these theories would include:

1. The entrainment and subsequent synchronization of speech and gesture is dependent upon prominent pulses that act as periodic attractors within two rhythmic systems.
2. A pulse for a gesture system corresponds to the purposeful stroke and to the vowel with the greatest motoric effort within a unit of speech (e.g., accented syllables which require increased amplitude, fundamental frequency, and duration).
3. Variability of synchrony will be least for the strongest attractors (e.g., stronger the prosodic prominence → stronger the synchrony).
4. As the stability and strength of one motor behavior increases, so will the likelihood that two systems will entrain (e.g., increased task automaticity → increased coordination; increased semantic information and novel planning of the movement → reduced coordination).
5. Pulses can be coordinated between multiple internal oscillators manifested in oral, manual, facial, and other body movements (e.g., movements like finger tapping and eyebrow raises also entrained with prominent pulses in the speech stream).
6. Perturbations in the speech system will have corresponding disruptions in the gesture system (e.g., lengthened speech production → lengthened gesture).
7. The spatial parameters of movement may also be affected by the entrainment of the manual and speech systems (e.g., producing an anterior to posterior movement of the hand → facilitates an analogous movement of the articulators while perturbing posterior to anterior movement).

Systematic inquiry of these predictions among others will offer rich opportunities for compelling scientific explorations of the integration of speech, manual, and linguistic process.

### 4. Implications and future directions

Dynamic systems approaches hold vast promise for understanding complex, divergent, and variable human behaviors, like speech and gesture, that yet have organization, consistency and even coordination of structures emerge from the seeming unlimited degrees of freedom. The study of dynamic coordinative nature of oral and manual movements is blossoming, but still only its provenance.

For instance, the optimal measure of the oscillator pulse within the speech stream and manual movement remains an empirical question. If one is interested in studying single time points of movement, then peak amplitude and/or peak fundamental frequency are alternative acoustic measures. However, based upon recent results [30], [31], [33], [34], the speech and gesture system appear to share dynamic motor linkages. Given that the question is one of movement, studies that have examined kinematic variables such as peak velocity [33], [34], [62] and jaw movement [30] offer the optimal guide for examining the coordination of movement. However, it is also notable that the optimal measure of a pulse, whether perceptual, acoustic, or kinematic, will likely remain elusive given the dynamic nature of entrainment and the unlikelihood that two systems would exhibit exact temporal synchronization. Tightly constrained studies of manual and speech movements may offer the most effective data for understanding the coordination of pulses, whereas more
natural studies of gesture and speech production will provide insight on the complexities and utility of the entrainment across internal systems.

The coordination of two motor systems can also be explored by studying the effects of perturbations. If temporal entrainment of the speech and gesture movements is a viable hypothesis, then one would anticipate that a perturbation of one of the movements would result in a corresponding temporal modification of the affiliate movement. This relatively straight-forward paradigm has rarely been utilized as a tool to examine the temporal relationship of speech and gesture. Only one group of researchers altered the timing of gesture [57] and measured the effects on the timing of speech. Likewise, only one experimental investigation of the effects of speech perturbation on the timing of speech was conducted to date [31]. These experiments, along with observations of the synchronization of gesture during speech dysfluencies produced by adults and children who stutter [60], [61] and ad hoc analyses of speech errors produced by typical adults [8] indicate that speech and gesture remain synchronized even when the timing of one of the movements is altered in some way. Though, systematic investigation of the temporal synchronization of speech and gesture following perturbations will further elucidate the entrainment of these two systems.

Lastly, the development and disruption of entrainment across systems may assist in understanding the underlying mechanisms of communication disorders. Recently, Corriveau and Goswami [56] explored the fairly consistent findings in the literature of slow, less accurate, and less coordinated fine motor skills exhibited by children with SLI relative to their peers. Interestingly, Corriveau and Goswami found that 21 children with SLI were only impaired relative to 21 age- and 21 language-matched peers when asked to manually tap to an external stimulus (i.e., pacing and entraining to a metronome), not when tapping to an internally generated rhythm (i.e., keeping the pace after the metronome was turned off). The relative contribution of the perception and production of rhythm was not parsed within this study, but certainly the findings exemplified the need for additional study on the entrainment of children and adults with communication disorders.

The relevant implication for this line of work is that linguistic and manual movements not only display shared processing in typical development and adult communication processes, but also in the breakdown of communication. Conversely, it is important to consider the potential therapeutic benefits of gesture and other manual movements in the treatment of individuals with prosodic deficits and other communication disorders. For example, two basic tenets of entrainment may be capitalized on in the management of individuals with speech production disorders.

1. Production of a manual pulse will increase the likelihood of a corresponding pulse in the speech system.
2. Spatiotemporal properties of a manual movement will affect the analogous spatial and temporal parameters of a speech movement.

Though the significance of gesture in the assessment and treatment of individuals with a variety of communication disorders such as aphasia and autism spectrum disorders is well established, the motoric coordination of these movements is far less considered in the management of children and adults with neuromotor speech disorders to date (for a review [63]). Thus, continued experimental investigation of, and theoretical postulation regarding the dynamic coordination of speech and gesture will offer significant contributions to a broad and diverse set of disciplines.

5. Acknowledgements

I am grateful for the invaluable support and guidance of Jana Iverson and Susan Shaiman for the conceptual and empirical development of this line of my scholarly work. Thank you also to the thoughtful comments provided by the review committee for GESPIN 2011.

6. References


