THE ROLE OF PROSODIC STRESS AND SPEECH PERTURBATION ON THE TEMPORAL SYNCHRONIZATION OF SPEECH AND DEICTIC GESTURES

by

Heather Leavy Rusiewicz

BA, University of Pittsburgh, 1997

MA, University of Pittsburgh, 1999

Submitted to the Graduate Faculty of

School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

PhD in Communication Science and Disorders

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This dissertation was presented

by

Heather Leavy Rusiewicz

Defense Date:

April 12, 2010

Committee:

Susan Shaiman, PhD, Associate Professor

Jana M. Iverson, PhD, Associate Professor

J. Scott Yaruss, PhD, Associate Professor

Diane L. Williams, PhD, Assistant Professor

Dissertation Direction: Susan Shaiman, Associate Professor

THE ROLE OF PROSODIC STRESS AND SPEECH PERTURBATION ON THE TEMPORAL SYNCHRONIZATION OF SPEECH AND DEICTIC GESTURES

Heather Leavy Rusiewicz, B.A., M.A., CCC-SLP

University of Pittsburgh 2010

Gestures and speech converge during spoken language production. Although the temporal relationship of gestures and speech is thought to depend upon factors such as prosodic stress and word onset, the effects of controlled alterations in the speech signal upon the degree of synchrony between manual gestures and speech is uncertain. Thus, the precise nature of the interactive mechanism of speech-gesture production, or lack thereof, is not agreed upon or even frequently postulated. In Experiment 1, syllable position and contrastive stress were manipulated during sentence production to investigate the synchronization of speech and pointing gestures. An additional aim of Experiment 2 was to investigate the temporal relationship of speech and pointing gestures when speech is perturbed with delayed auditory feedback (DAF). Comparisons between the time of gesture apex and vowel midpoint (GA-VM) for each of the conditions were made for both Experiment 1 and Experiment 2. Additional comparisons of the interval between gesture launch midpoint to vowel midpoint (GLM-VM), total gesture time, gesture launch time, and gesture return time were made for Experiment 2. The results for the first experiment indicated that gestures were more synchronized with first position syllables and neutral syllables as measured GA-VM intervals. The first position syllable effect was also found

in the second experiment. However, the results from Experiment 2 supported an effect of contrastive pitch effect. GLM-VM was shorter for first position targets and accented syllables. In addition, gesture launch times and total gesture times were longer for contrastive pitch accented syllables, especially when in the second position of words. Contrary to the predictions, significantly longer GA-VM and GLM-VM intervals were observed when individuals responded under provided delayed auditory feedback (DAF). Vowel and sentence durations increased both with (DAF) and when a contrastive accented syllable was produced. Vowels were longest for accented, second position syllables. These findings provide evidence that the timing of gesture is adjusted based upon manipulations of the speech stream. A potential mechanism of entrainment of the speech and gesture system is offered as an explanation for the observed effects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Research is not conducted in seclusion and nor is life lived in isolation. I am fortunate to have not only grown in academic knowledge during the pursuit of my doctoral degree, but to be surrounded by such amazing individuals. I am indebted to the University of Pittsburgh for provided an enriching intellectual environment and so much more. I am grateful for my mentors, colleagues, friends, and family for all of their support, seen and unseen.

Thank you to my dissertation committee co-chairs, Susan Shaiman, Ph.D. and Jana Iverson, Ph.D. and my committee members, Diane Williams, Ph.D. and J. Scott Yaruss, Ph.D. You each have inspired me in so many ways. Diane, I am privileged to call you a colleague and friend. You will continue to be a role model as a junior faculty member, but even more as an example of how to balance what is truly meaningful. Scott, your doctoral seminar on theoretical models of speech production first set me on the path to these dissertation questions. I admire and hope to emulate your exceptional ability to meld research and clinical impact in my professional career. Jana, you are an amazing woman. From the first time I met I knew you were a tremendous thinker, but had such a warmth and strong spirit. Your work will continue to inspire me and I hope to continue the multi-faceted friendship we have developed. Sue, you have taken me under your wing and have helped me to fly. This document would truly not be what it is, or even perhaps here at all if it was not for your encouragement on so many levels. You were selfless with your time, your resources, and your meticulous thought on this investigation. I am so fortunate to have learned from you as an undergraduate, graduate, and doctoral student and now as your colleague. In addition to my committee, I would like to acknowledge the invaluable

help of Neil Szuminsky for his expertise and many hours of creating the software and instrumentation necessary to complete this project and Elaine Rubenstein for her patience, time, and instruction regarding the statistical analyses of the investigation.

I set out on this path long ago only by the encouragement by Tom Campbell, Ph.D. and Chris Dollaghan, Ph.D. You have both changed my life in so many incredible ways. I will never be able to express the gratitude I have for the innumerous academic, clinical, and personal opportunities you provided me. I am especially appreciative and fond of all the years spent with you and so many wonderful people at Children's Hospital of Pittsburgh. I also would like to specifically acknowledge the friendship and support of Denise Balason, Jennifer DelDuca, Sharon Gretz, Mitzi Kweder, Tammy Nash, and Dayna Pitcairn.

My colleagues now extend down Fifth Avenue to Duquesne University. My students have been a tremendous motivation. I am particularly indebted to Jordana Birnbaum and Megan Pellettiere for their diligent work with these experiments. I am so thankful to be a faculty member in the department of Speech-Language Pathology at DU and surrounded by such wonderful people. Thank you all and thank you Mikael D.Z. Kimelman and Katie Staltari for your constant support and encouragement.

My love and overwhelming gratitude is given to my family and friends. It is simplest to say that in no way would I be who I am or where I am today without each of you. I am blessed to be part of a joyous and loving family. My grandparents, Bob and Eleanor West, taught me to be kind to others and to dream big. My mom and dad, Kathy and Denny Leavy, provided me with the surest foundation of experiences, values, and the importance of pursuing all goals. My sister, Alyssa, Uncle Bob, Aunt Diane, and cousin, Kristen guided my path and kept me going on that path.

Finally, this project and indeed all that I do is dedicated to my “boys". Jameson and Brenden, you are mommy’s pride and joy and inspire me every day. Scott, there are no words to convey the love, appreciation, and admiration I have for you. Thank you for your patience, for listening to all this talk about prosody and gestures for all these years, and for being my constant best friend.

1.0 INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

We gesture simultaneously while we speak, though it is not clear if manual gesture and spoken language are part of a unified communication system. Intuitively, it appears that the modalities of speech and gesture fuse to express communicative information. Yet, the vast majority of research of the relationship of speech and gesture is based upon just that, intuition and informal observations. In fact, it is not clear whether speech and gesture truly are interactive or at what point during the processing of speech and manual gestures the assumed interaction occurs. The purpose of the present investigation was to measure the effect of three variables, prosodic stress, syllable position, and the temporal perturbation of speech upon the degree of synchronization between the speech and deictic (i.e., pointing) gestures. The aim was to test the notion that the perceived tight temporal synchrony of speech and gesture is evidence of an integrated communication system.

I measured the degree of temporal synchrony between gesture apex (i.e., time of maximum extension of the deictic gesture) and vowel midpoint utilizing a novel methodology. This methodology consisted of capacitance sensors for the temporal measures of gesture movement and acoustic analyses of the temporal parameters of the speech signal. The

investigation was comprised of two experiments. The first was a controlled paradigm of typical adults producing compound words within seven word utterances. Syllable position and the prosodic variable of contrastive pitch accent were manipulated in Experiment 1. In the second experiment, typical adults produced the same compound words within utterances, though with the additional manipulation of speech perturbation. Contrastive pitch accent, syllable position, and temporal perturbation of speech via delayed auditory feedback (DAF) were manipulated in Experiment 2. Syllable position was manipulated to control for a possible confounding effect of word onset synchrony. Prosodic stress and speech perturbation were chosen to not only test whether a predictable temporal relationship between speech and gesture exists, but also to examine the mechanism of this potential interaction. Manipulation of prosodic stress, specifically contrastive pitch accent, investigated the role of the phonological encoder in the timing of manual gestures. In contrast, imposing an auditory delay perturbed the timing of speech production at a lower-level of motor processing (i.e., at the motor programming level) at which speech and gesture are hypothesized to be temporally entrained.

1.2 BACKGROUND

The relationship between speech and gesture has long intrigued scholars from an array of disciplines. A myriad of associations between gesture and speech exist in the literature, such as the evolution of spoken language from manual gestures (Blute, 2006; Corballis, 2003; 2010;

Hewes, 1972; Rizzolatti & Arbib, 1998), the concurrent acquisition of gesture and language development milestones (Bates & Dick, 2002; Capone & McGregor, 2004; Goldin-Meadow, 1998; Iverson, 2010; Iverson & Thelen, 1999; Kent, 1984; McNeill, 1992; Volterra, Caselli, Capirci, & Pizzuto, 2004), the facilitatory effect of gestures on word learning (Acredolo & Goodwyn, 1985, 1988; Capone & McGregor, 2005; Goodwyn & Acredolo, 1993; Goodwyn, Acredolo, & Brown, 2000; Namy, Acredolo, & Goodwyn, 2000; Weismer & Hesketh, 1993; though see Johnston, Durieux-Smith, & Bloom, 2005), the capability to provide prognoses for children with language deficits and autism spectrum disorders as a function of their early gesture use (Brady, Marquis, Fleming, & McLean, 2004; Fenson, et al., 1994; Shumway & Wetherby, 2009; Smith, Mirenda, & Zaidman-Zait, 2007; Thal, 1991; Thal & Bates, 1988; Thal & Tobias, 1992, 1994; Thal, Tobias, & Morrison, 1991; Watt, Wetherby, & Shumway, 2006), the ability to enhance speech intelligibility for adults with dysarthria (Garcia & Cannito, 1996; Garcia, Cannito, & Dagenais, 2000; Garcia & Dagenais, 1998), and the facilitation of word recall for children (Pine, Bird, & Kirk, 2007), typical adults, (Beattie & Shovelton, 2006; Morrel-Samuels & Krauss, 1992; Ravizza, 2003) as well as adults with aphasia (de Ruiter, 2006; Feyereisen, 2006; Hanlon, Brown, & Gerstman, 1990; Marshall, 2006; Miller, 2006; Pashek, 1997; Power & Code, 2006; Raymer, Singletary, Rodriguez, Ciampitti, Heilman, & Rothi, 2006; Richards, Singletary, Rothi, Koehler, & Crosson, 2002; Rose, 2006; Rose & Douglas, 2001; Rose, Douglas, & Matyas, 2002; Scharp, Tompkins, & Iverson, 2007). Although it is tempting to impose explanatory power upon the relationship of speech and gesture for such stimulating topics, it is first necessary to evaluate the basic assumption that the two systems are indeed integrated. Moreover, even if these tantalizing statements regarding the relationship of speech and gesture ring true, there are virtually no data on the mechanism of their interaction. To be

sure, experimental investigations of this still mostly anecdotal relationship of speech and gesture are in their infancy and are not usually theoretically motivated. The existing literature is constituted primarily of studies that are observational and do not directly manipulate the variables of interest. Furthermore, disparate gesture classification schemes and design flaws are pervasive in this literature. Thus, many of the basic tenets regarding the relationship between speech and gesture are tenuous and in need of both expanded theoretical postulation and empirical scrutiny.

Despite the lack of systematic research regarding the relationship of speech and gesture, it is generally accepted that the two processes are interactive during the production of communication. Specifically, the apparent synchronous temporal relationship of speech and gesture is cited as premier evidence of the interactive nature of speech and gesture production (e.g., Goldin-Meadow, 1999; Iverson & Thelen, 1999; Krauss, Chen, & Gottesman, 2000; McNeill, 1992). Yet, not only is there limited empirical investigation of the relative timing of speech and gesture, but there are also few predictive and testable models that encourage systematic exploration of specific points of interaction as well as the temporal consequence of these two production systems. The common observation is that gesture and speech roughly occur at similar times during communication. Remarkably, in the past two decades of gesture research this statement has most often been simply left at that.

Even though many individuals have observed that speech and gesture are produced in tight temporal synchrony, the mechanism responsible for this potential synchronization as well as factors that may affect the degree of time separating two actions have not been elucidated. In fact, it is not at all clear that speech and gesture truly synchronize in a predictable manner or if it is merely a common observation that gesture and speech roughly occur at similar times during

communication due to independent but temporally similar processes. In order for speech and gesture to synchronize in some uniform manner, there must be interaction between the speech production and gesture production systems. However, where in the speech and gesture systems this interaction occurs, or if there is any interaction at all, is unclear.

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 (McNeill, 1985; 1992). However this hypothesis leads to nebulous points of interaction, thus making it impossible to make predictions regarding the effects of specific speech and gesture variables upon the timing of the other movement. Recently de Ruiter (1998; 2000) proposed a model of gesture production with an explicit point of interaction with the speech production system. In short, de Ruiter's Sketch model (see Figure 1) is an extension of Levelt's (1989) model of speech production (see Figure 2) 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 or altering the timing of movement by perturbing the execution of speech or gesture movements, do not influence the timing of gesture.

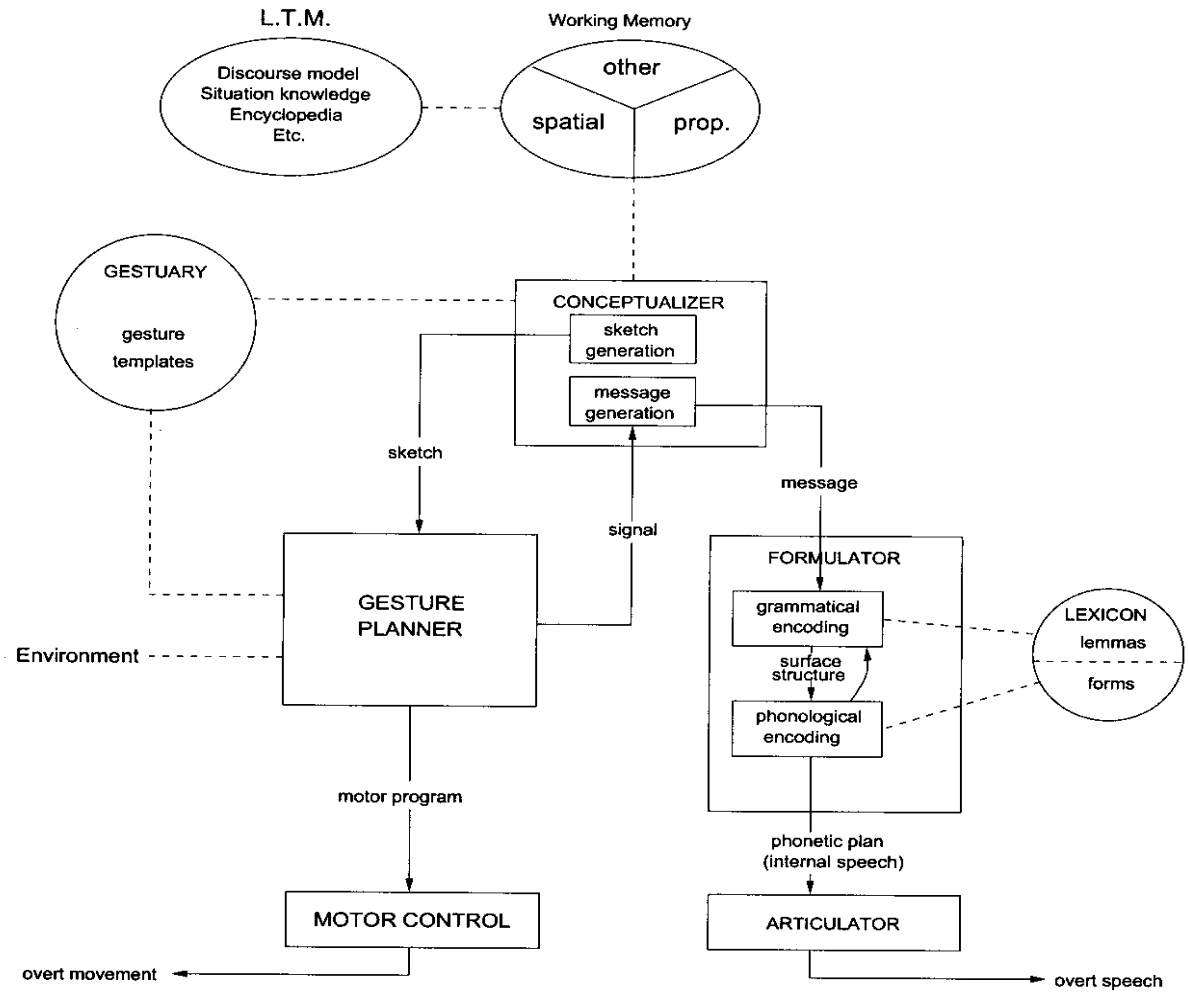


Figure 1 The Sketch model. Adapted from "Gesture and Speech Production," by J.P. de Ruiter, 1998, Unpublished doctoral dissertation, Katholieke Universiteit, Nijmegen, Germany, p. 16.

The prediction of the current study, counter to de Ruiter, is that the degree of synchrony between speech and gesture does indeed differ as a function of processes below the level of the conceptualizer. The first alternative hypothesis is that gestures temporally align with prosodically stressed syllables due to interaction between gesture planning processes and phonological encoding processes within the formulation stage of speech production. The second alternative hypothesis is that increased synchrony of prosodically prominent syllables and gestures results from temporal entrainment of oral and manual movements. The present experiments focused on the effects of prosodic prominence and speech perturbation upon the degree of temporal synchronization between speech and deictic gestures in order to experimentally test these hypotheses.

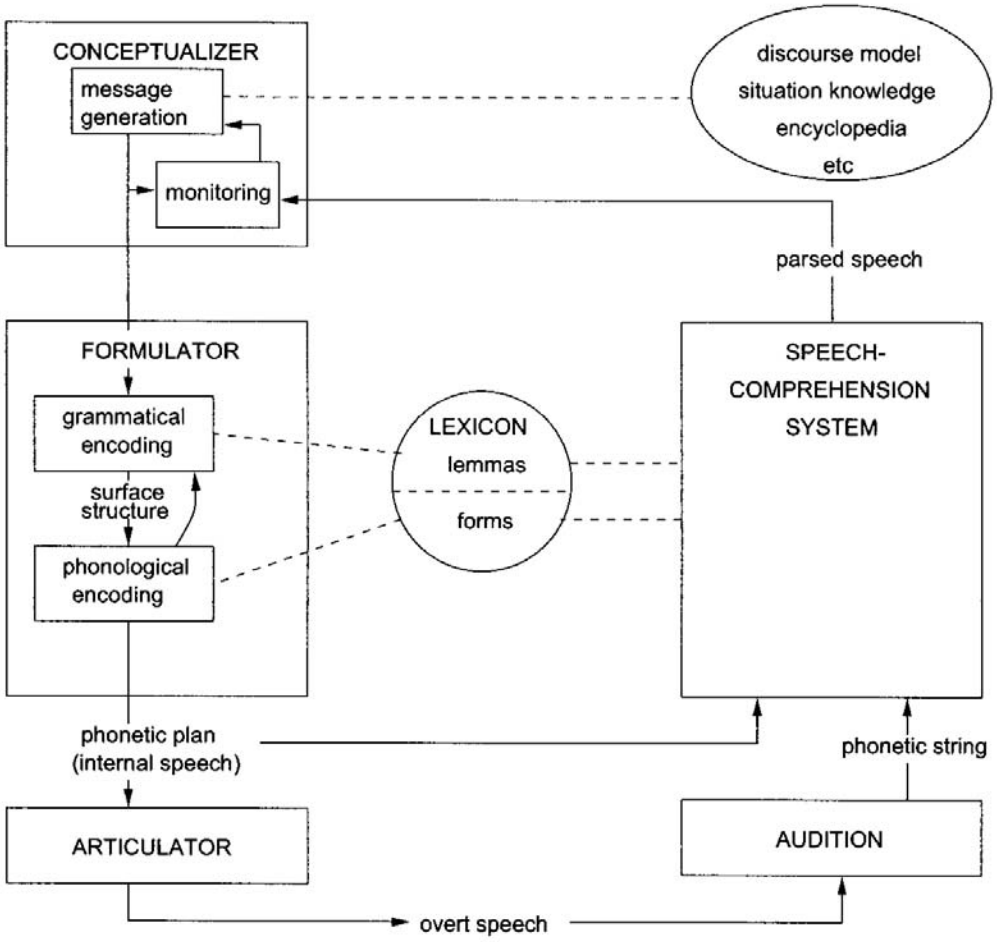


Figure 2 A blueprint for the speaker: Levelt's (1989) model of speech production as illustrated in de Ruiter (1998). Adapted from "Gesture and Speech Production," by J.P. de Ruiter, 1998, Unpublished doctoral dissertation, Katholieke Universiteit, Nijmegen, Germany, p. 9.

As one might expect, because experimental study of the gesture-speech linkage is only budding, the existing literature on the effect of prosodic stress and perturbation of speech on the temporal synchrony of speech and gesture is equivocal. This work is thought-provoking but far from definitive. First, there is evidence to suggest that gestures temporally align with prosodically stressed syllables. Second, observational work implies that the temporal parameters

of a gesture change to remain synchronized with the corresponding speech signal, even when the speech signal is perturbed.

Though the majority of investigations found that gestures and prominent syllables tend to co-occur (Bull & Connelly, 1985; de Ruiter, 1998; Loehr, 2004; McClave, 1998; Nobe, 1996; Rochet-Capellan, Laboissière, Galvàn, & Schwartz, 2008; Yasinnik et al. 2004), the psychological reality of this assertion is disputable for several reasons. First, McClave (1994) demonstrated that beat gestures occurred with unstressed syllables just as frequently as with stressed syllables. Also, de Ruiter (1998) found that lexical stress did not affect the timing of the corresponding gesture for Dutch speakers. In addition to conflicting findings in the literature, abundant design and methodological concerns limit the validity of the findings. With the exception of de Ruiter (1998) and Rochet-Capellan and colleagues (2008), investigators have not directly manipulated speech and/or gesture when investigating the effects of prosodic stress upon synchronization. Most often, examiners perceptually judged the acoustic signal and visually identified the elusive boundaries of gesture in frame-by-frame video analysis. While de Ruiter utilized an ultrasound system to record gesture signals and Rochet-Capellan and colleagues used an infrared tracking device to record gesture and speech movements, only Rochet-Capellan et al have measured the specific temporal parameters of the unfolding gesture relative to the accompanying speech signal. It is also important to point out that the only investigators to directly manipulate the stimuli to elicit predictable prosodic stress have done so only with Dutch (de Ruiter, 1998) and Brazilian Portuguese (Rochet-Capellan, et al., 2008) speakers and with a three word phrase and one of four bisyllabic nonword responses, respectively. Certainly, it is difficult to identify a predictable effect of prosodic stress upon the timing of gesture given such imprecision in studies that employed natural speaking contexts and constrained spoken responses

in the studies that employed an experimental paradigm. Provided the available data, the observation that gestures occur roughly at the time of a prosodically stressed syllable during spoken language production remains correlational relationship at best.

Even if gestures temporally co-occur with prosodically prominent syllables, as the existing literature implies but de Ruiter argues against in his Sketch Model, 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 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). It is indeed possible that gesture processes interact with speech processes at the level of the phonological encoder in the Formulator, as posited in the first alternative hypothesis. However, it is also possible that gesture aligns with prosodically prominent syllables due to temporal entrainment of the oral and manual motor systems at a lower level in the production systems. de Ruiter himself postulated that a "phase locking mechanism at the level of lower level motor planning, instead of a higher level synchronization process" (1998; p. 61) may be responsible for gesture synchronizing with prosodically strong syllables given his unpredicted finding that gesture apices aligned with contrastively stressed syllables.

Although de Ruiter (1998) argues for a possible interaction at the level of motor planning, one could conjecture that in fact the interaction is at the level of motor programming. There is an abundance of theory and experimentation on the topic of motor control and the role of feedback in open versus closed loop systems which will not be reviewed within this document. However, it is necessary to clarify the specific level of motor processing that it essential for the research questions. Adhering to a linguistic-focused model such as the Sketch

Model (1998) and Levelt's (1989) model of speech production provides ample information regarding linguistic and phonologic processes but does not describe motor processing in adequate detail. For that reason, along with the alternative hypothesis that an interaction may not only occur at the level of phonological encoding but also at a "lower-level" of motor processing, forces one to meld a linguistic model with a model of speech production from a motor perspective. Anita van der Merwe's (1997) four-level framework of speech sensorimotor control (Figure 3) is a well-specified model of speech motor control. Her model is founded in current neurophysiological data and reflects the transition from historical view of two stages of motor production (i.e., motor programming and execution) to a three-stage view (i.e., motor planning, motor programming, and execution).

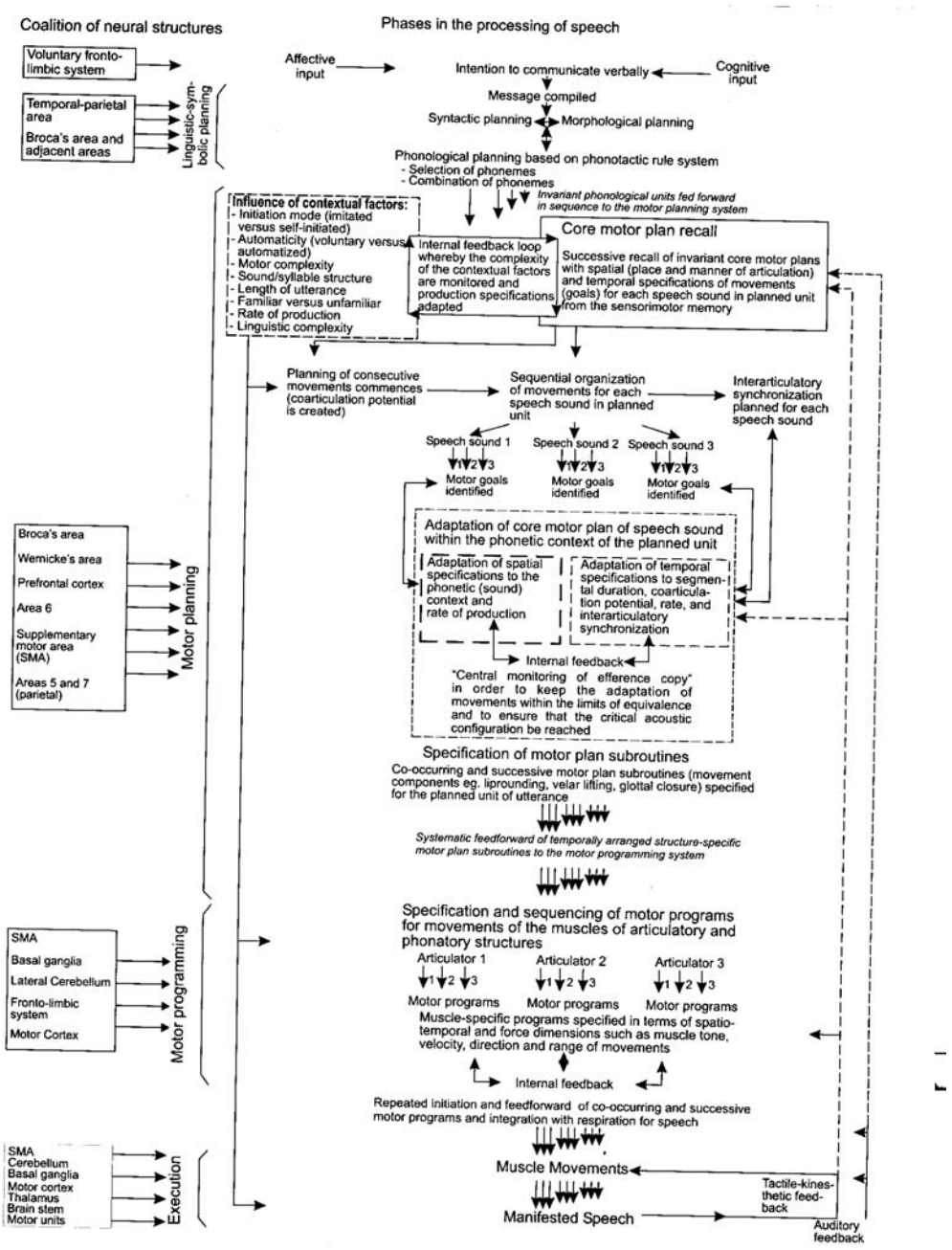


Figure 3 Four-level framework of sensorimotor control of speech production. Adapted from "A theoretical framework for the characterization of pathological speech sensorimotor control" by A. van der Merwe, 1997, In M. McNeil (Ed.), *Clinical Management of Sensorimotor Speech Disorders. New* York, NY: Thieme Medical Publishers, Inc, p. 8.

Even though it is quite possible that speech and gesture interact at more than one level of motor processing, the temporal perturbation of speech via DAF manipulates processing at the motor programming level. In contrast to a motor plan which specifies the generalities of a movement including the basic trajectory of movement, the sequence of movement, and so on, the motor program specifies the spatio-temporal and force dimensions of the movement. A second facet of the motor program is the "sensory feedback can be utilized to change or update a program should the need arise" (p. 11). One form of sensory feedback that can be integrated by the motor program is auditory information, such as that manipulated in this investigation. Another advantage of van der Merwe's sensorimotor model of speech production is that it purposefully mirrors current opinion of motor control in general (e.g., limb motor control). She states the "interface between preplanned motor programs and real-time updating based on sensory input, therefore, seems to be intrinsic to the motor programming of movement, including speech movements" (p. 11). Thus, the proposed entrainment of the speech and gesture systems is rooted in the level of motor programming for both the speech and manual movements. However, it is important to note that this investigation did not explicitly dissociate the three stages of motor processing. Although the current hypothesis is that speech perturbation affects the level of motor programming, perturbation and entrainment of speech and gesture also potentially involves the level of motor planning and/or execution. Systematic dissociation of these motor processing levels relative to the relationship of speech and gesture remains a topic for future research.

If temporal entrainment of the speech and gesture movements at the level of motor programming prevails, 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 (Levelt, Richardson, & La Heij, 1985) and measured the effects on the timing of speech. Likewise, McNeill (1992) is the only individual to directly perturb speech production and report the resultant effects on the execution of gesture by recording the qualitative effects of DAF. These experiments, along with observations of the synchronization of gesture during speech dysfluencies produced by adults and children who stutter (Mayberry & Jaques, 2000; Mayberry, Jaques, & DeDe, 1998) and ad hoc analyses of speech errors produced by typical adults (de Ruiter, 1998) suggest that speech and gesture remain synchronized even when the timing of one of the movements is halted in some way. However, a systematic investigation of the temporal synchronization of speech and gesture following speech perturbation has yet to be conducted.

Yet, these prior experiments offer preliminary evidence that the oral and manual movements associated with speech and gestures may be temporally entrained. It is further hypothesized that entrainment of the two motor behaviors results from the coupling of gesture movements to the rhythmic production of prominent syllables. This hypothesis was first put forth by Tuite in 1993, though in vague terms. Tuite's hypothesis will be amalgamated with work by Cummins and Port (e.g., Port, 2003) which, in line with dynamic systems theory, proposes that neurocognitive oscillators produce pulses that act as attractors for other behaviors such as speech production. Specifically the oscillator pulses attract "perceptually prominent motor events, like vowel onsets or taps of a finger" and "the phase of the internal system is

adjusted so that the perceptually salient event is synchronous with the oscillator pulse" (Port, 2003, p. 607). Thus, the second alternative hypothesis asserts that speech is a rhythmic oscillator that entrains manual gesture movements at these oscillator pulse points.

The general purpose of this project was to explore the theoretical temporal relationship of speech and gesture as a function of (i) contrastive pitch accent (i.e., present or absent), (ii) syllable position (i.e., first and second postion), and (iii) speech perturbation (i.e., 200 ms auditory delay or no auditory delay) upon the degree of temporal synchrony of speech and deictic (i.e., pointing) gestures produced by typical, young, English-speaking adults. Two experiments were conducted to address this purpose.

Synchrony was measured as an interval from gesture apex to vowel midpoint (GA-VM). Vowel midpoint was chosen as the acoustic dependent variable for two reasons. First, vowel midpoint incorporates duration which is frequently proposed as the most consistent acoustic correlate of prosodic stress (Fry, 1955; Sluijter & van Heuven, 1996; van Kuijk & Boves, 1999; Wouters & Macon, 2002). Second, the finding that prosodic stress influences the vowel/nucleus of a syllable to a much greater degree than the on- and offset of a syllable is reflected in the choice of vowel midpoint rather than a measure of rime, syllable, or word duration (Adams & Munro, 1978; Fry, 1955; Greenberg, Carvey, Hitchcok, & Chang, 2003; Suijter & van Heuven, 1996; Tuller, Harris, & Kelso, 1982; Tuller, Kelso, & Harris, 1982; Turk & White, 1999; van Kuijk & Boves 1999). A recent investigation completed by Krahmer and Swerts (2007) provides strong support for choosing vowel duration as the acoustic dependent variable in this particular paradigm as well. Ten Dutch speakers were asked to produce a single utterance, *Amanda goes to Malta*, with either no pitch accent, pitch accent on the second syllable of *Amanda* or pitch accent on the first syllable of *Malta.* The participants were also instructed to produce a beat gesture,

head nod, or eyebrow raise along with the pitch accent 50% of the time. Vowel duration was significantly greater for syllables that were pitch accented compared to those that were not, regardless of position within the utterance or word.

Gesture apex was chosen as the gesture dependent variable also for two reasons. First, it is a single time moment when the gesture reaches its point of maximum extension and measuring a precise moment in time is conducive for an interval-based measure of temporal synchronization. Second, a gesture apex is thought to be synonymous to a gestural stroke which holds the semantic information of the gesture. In addition to the gesture apex to vowel midpoint (GA-VM) interval, the total gesture time (gesture onset to offset) and gesture launch (gesture onset to apex), and gesture return (gesture apex onset to gesture offset) will be measured in Experiment 2. These variables parallel the observations made by previous researchers' studies of the effect of speech perturbation upon the timing of gesture (de Ruiter, 1998; Mayberry & Jaques, 2000; Mayberry, Jaques, & DeDe, 1998; McNeill, 1992).

1.3 SPECIFIC AIMS, RATIONALE, EXPERIMENTAL QUESTIONS, AND HYPOTHESES

Experiment 1:

Specific Aim:

To assess the influence of (i) contrastive pitch accent, (ii) syllable position, and (iii) their interaction on the degree of synchrony between the apices of deictic gestures directed toward a visual display and vowel midpoints of syllables produced within corresponding carrier phrases.

Rationale:

There are four motivating factors for choosing contrastive pitch accent as an independent variable. First, this study attempted to disambiguate previous research that reported an effect of prosodic prominence upon the temporal synchronization of speech and gesture, though with increased control of confounds present in the literature to date. Second, any effect of a prosodic variable, in this case pitch accent, upon the timing of gesture is counter to de Ruiter's Sketch Model (1998; 2000), which asserts there is no communication between the gesture and speech systems below the level of the conceptualizer (Levelt, 1989). Third, pitch accent is notoriously difficult to identify with certainty within spontaneous speech (e.g., Bolinger, 1972); therefore,

contrastive pitch accent was chosen as the prosodic dependent variable so that the pitch-accented syllable may be reliably identified in each response. Finally, because a pitch-accented syllable is assigned greater prominence than all other stressed syllables within a given intonational phrase, the paradigm was developed to measure the greatest potential effect of prosody upon the temporal synchronization of speech and gesture.

Syllable position was manipulated to discriminate between an effect of pitch accent and an effect of word onset. It is well documented in the literature that gesture onset tends to precede the onset of its lexical affiliate (Bernardis & Gentilucci, 2006; Butterworth & Beattie, 1978; Chui, 2005; Feyereisen, 1983; Krauss et al., 1996; 2000; McNeill, 1992; Morrel-Samuels & Krauss, 1992; Ragsdale & Silvia, 1982). Hence, the hypothesis that the *onset* of gesture occurs prior to the onset of the lexical affiliate was tested in these experiments. Also, it is possible that gestures synchronize with the onset of their lexical affiliates instead of with prosodically prominent syllables. These two main effects are not dissociable for syllables in the first syllable position. Additionally, an interaction effect between prosodic stress and syllable position may exist if there is interaction between the phonetic plan of the Formulator and the gesture planner since both stress assignment and lexical access occur within the Formulator. That is, pitch-accented syllables in the initial position should have greater synchrony with gestures than all other syllables if both syllable position and prosodic stress play a role in the synchrony of gesture and speech.

Contrastive pitch accent and syllable position were manipulated within compound word pairs imbedded within utterances (i.e., seven words total for each utterance). The compound words were produced within controlled carrier phrases and spoken by the participants while simultaneously pointing to a corresponding picture of the item. The task was tightly constrained

in this experiment in an attempt to isolate the effect of prosodic stress from other acoustic-phonetic and lexical effects as well as potential cognitive processes that could affect the temporal relationship of gesture and speech. The utterances are long and varied in content to encourage more natural suprasegmental characteristics than single word or short phrase productions. 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. Thus, it is anticipated that acoustic vowel duration would be longer for syllables with pitch accent compared to the same syllables without pitch accent for Experiments 1 and 2. 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, thus eliminating the potential confounding effect of metrical stress (i.e., lexical stress).

Experimental Questions and Hypotheses:

1) **Do the apices of deictic gestures synchronize with the vowel midpoints of pitch-accented syllables in compound words spoken within carrier phrases more than the same syllables without pitch accent regardless of word position?**

H_0^1: There is no significant difference between the mean GA-VM interval for syllables produced with contrastive pitch accent and the mean GA-VM interval for the same syllables produced with no contrastive pitch accent.

2) **Do the apices of deictic gestures synchronize with the vowel midpoints of the first syllable more than the second syllable of compound words spoken within carrier phrases regardless of pitch accent assignment?**

H_0^2: There is no significant difference between the mean GA-VM interval for syllables in the first syllable position and the mean GA-VM interval for syllables in the second position.

3) **Is there a significant interaction between pitch accent and syllable position upon the degree of synchronization between the apices of deictic gestures and the vowel midpoint of pitch-accented syllables of compound words spoken within carrier phrases?**

H_0^3: There is no significant interaction between pitch accent (i.e., presence and absence of pitch accent) and syllable position (i.e., first and second position of compound word pairs) on the mean GA-VM interval.

4) **Are the vowel durations of syllables with contrastive pitch accent longer than for the same syllables produced without contrastive pitch accent?**

H_0^4: The vowel durations of syllables with contrastive pitch accent are not significantly longer than for syllables produced with neutral pitch accent.

Experiment 2:

Specific Aim:

To assess the influence of (i) contrastive pitch accent, (ii) syllable position, (iii) speech perturbation via DAF, and (iv) their interaction during the production of utterances by typical adults on (a) the degree of synchrony between deictic gestures and speech and (b) the individual temporal parameters of deictic gestures (i.e., total gesture time, gesture launch time, and gesture return time).

Rationale:

The primary goal of Experiment 2 was to explore the mechanism responsible for temporal synchronization of prosodic stress and gesture. Accordingly, one core component of Experiment 2 differed from the first experiment; the perturbation of speech. An auditory delay of 200 ms was imposed on 50% of the experimental trials to breakdown the temporal fluidity of the spoken production. The rationale for choosing DAF as an independent variable was both theoretical and practical.

While a finding of increased speech-gesture synchrony for accented syllables compared to non-accented syllables is a contribution to existing research and indicates that the speech and gesture systems must interact at some level, this finding alone does not indicate where synchrony is generated or why synchrony occurs, only that it does. If we adhere to de Ruiter's Sketch Model (1998, 2000) and the parallel speech production model (Ferreira, 1993; Levelt, 1989), then the interaction of the two systems potentially occurs at the level of the phonological encoder where prosody is generated. Yet this explanation seems to fall short. If an interaction exists at the level of the phonological encoder, specifically the Prosody Generator, it is still not clear why gestures synchronize with prominent syllables rather than the onset of the lexical item.

It can also be hypothesized that the potential interaction is not at the level of the phonological encoder, but instead at a lower level or motor programming within the speech system that is not encompassed by a modular, top-down, linear, linguistic-based model such as those models proposed by Levelt (1989) and de Ruiter (1998, 2000). The implication is that the manual gesture and oral speech movements are temporally entrained according to tenets of dynamics systems theory. According to such a view, neurocognitive oscillators produce pulses that act as attractors for the most perceptually prominent points of motor behaviors, such as finger taps or vowel onsets (Port, 2003; Tuite, 1993). In this case, it is proposed that the speech and gesture system are two internal, coupled oscillators and the vowels of pitch-accented syllables act as attractors for the apexes of concurrent manual gestures. According to this view, gestures synchronize with prominent syllables not only as an intrinsic coordination of motor behaviors but also to increase salience of, and therefore attention paid to, the prominent syllable or lexical item by the communication recipient (Jones & Boltz, 1989; Large & Jones, 1999). Greater details of this rationale are provided within the review of the literature.

One way to investigate the potential entrainment of speech and gesture is to perturb one of the systems and examine the resultant effects on the other system, though few investigators have examined the effect of perturbation upon the temporal synchronization of speech and gesture (de Ruiter, 1988; Levelt et al., 1985; Mayberry & Jaques, 2000; Mayberry, Jaques, & DeDe, 1998; McNeill, 1992). While these studies presented intriguing observations of synchronized speech and gesture even in the face of perturbation, no one has conducted a systematic study of the influence of speech perturbation upon the quantitative parameters of gesture. Additionally, this is the first study that attempted to examine the mechanism of speech-gesture synchronization.

Manipulating the presence and absence of an auditory delay allows one to perturb the speech system 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. Although there are other ways to perturb the speech system such as introducing a physical barrier to speech (Abbs & Gracco, 1984; Folkins & Zimmerman, 1981; Gracco & Abbs, 1985; Kelso & Tuller, 1983; Kelso, Tuller, Vaikiois-Bateson, & Fowler, 1984; Shaiman, 1989; Shaiman & Gracco, 2002), anesthetizing the articulators (e.g., Ringel, 1970) utilizing a metronome to alter the rate of speech (e.g., Cummins & Port, 1998), eliciting speech errors (e.g., de Ruiter, 1998), or observing the spontaneous dysfluencies of individuals who stutter (Mayberry & Jaques, 2000; Mayberry et al., 1998), DAF was selected as the variable of choice not only because of its well-documented effects upon the timing of speech (e.g., Howell & Archer, 1984) but also because of the use of DAF in a rudimentary, though influential, study of DAF upon the execution of gesture conducted by McNeill (1992).

DAF is described by Pfordresher & Benitez (2007, p. 743) as "a constant time lag (that) is inserted between produced actions (e.g., piano keypress) and the onsets of auditory feedback events (e.g., the onset of a pitch)". There are two seemingly oppositional effects of DAF upon speakers. DAF can increase fluency for individuals who stutter (e.g., Harrington, 1998; Kalinowski, Stuart, Sark, & Arnson1996). Conversely, DAF causes a breakdown of fluency in typical speakers characterized by decreased speech rate, prolonged voicing, increased speech errors (e.g., phoneme exchanges), increased vocal intensity, and increased dysfluencies (e.g., prolongations and part-word repetitions) (e.g., Burke, 1975; Howell & Archer, 1984; Stuart, Kalinowski, Rastatter, & Lynch, 2002). Though the perturbation caused by DAF "clearly reflects temporal coordination of actions and sound" (Pfordresher & Benitez, 2007, p. 743) and involves the temporo-parietal regions of the cortex (Hashimoto & Sakai, 2003), the mechanism of the auditory perturbation is not clear at this time.

Regardless of the underlying mechanism, the consistent finding that DAF causes lengthened articulatory durations was taken advantage of in the current experiment. Accordingly, the time required to produce each response utterance was expected to be longer when an auditory delay is present relative to when there is no delay. Mean time error (MTE) per trial was measured to validate the effect of DAF upon the temporal execution of speech production. For instance, 17 speakers similar to those that will be enrolled in the current experiment produced an average of approximately 4 syllables per second when presented with a 200 ms auditory delay compared to approximately 6 syllables per second when reading the same passage without an auditory delay (Stuart, et al., 2002, p. 2238). MTE is a measure often used to reflect the expected lengthened duration to complete a spoken word production under the influence of DAF compared to NAF conditions. Elman 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, 1983). MTE is measured in milliseconds and the greater the MTE measurement, the time difference between DAF and NAF conditions. Likewise, the temporal variables associated with the production of the corresponding deictic gesture were also predicted to be different for DAF conditions compared to normal auditory feedback (NAF) conditions. It is anticipated that MTE would be positive, indicating longer utterance duration for DAF trials compared to NAF trials. Additionally, the mean utterance durations for the two conditions were compared using a dependent samples t-test for each participant to test the prediction that speech rate is reduced under the influence of DAF.

The dependent variables for measuring the temporal parameters of the deictic gestures during DAF and NAF conditions consisted of total gesture time (ms) and gesture launch time (ms) (i.e., time from gesture onset to gesture apex). These variables were chosen to measure a potential temporal disruption of gesture that may result due to the disruption of the speech system. In other words, if there is no lower level entrainment of the speech and gesture systems, then the deictic gesture should be executed no differently for DAF and NAF conditions. Alternatively, if the gesture movement is affected by the speech movement timing, then gesture launch time and total gesture time would be longer for DAF trials relative to the NAF trials. A significant difference between the mean GA-VM intervals for DAF and NAF conditions would indicate that the two systems are not entrained during this task. The rationale for this prediction is that the timing of gesture would be the same in both tasks although the onset of the spoken affiliate would be delayed in the DAF condition, thus lengthening the GA-VM interval.

Lastly, the rationale for selecting a 200 ms auditory delay was based upon the ubiquitous finding that this duration yields the most consistent breakdowns in the temporal execution of speech produced by typical adults (Finney & Warren, 2002; Marslen-Wilson & Tyler, 1981; Stuart, et al., 2002), perhaps due to the fact that 200 ms is approximately the length of an average syllable.

Experimental Questions and Hypotheses

1) **Do the apices of deictic gestures synchronize with the vowel midpoints of pitch-accented syllables in compound words spoken within short utterances more than the same syllables without pitch accent regardless of word position?**

H_0^1: There will be no significant difference between the mean GA-VM interval for syllables produced with contrastive pitch accent and the mean GA-VM interval for the same syllables produced with no contrastive pitch accent.

2) **Do the apices of deictic gestures synchronize with the vowel midpoints of the first syllable more than the second syllable of compound words spoken within short utterances regardless of pitch accent assignment?**

H_0^2: There is no significant difference between the mean GA-VM interval for syllables in the first syllable position and the mean GA-VM interval for syllables in the second position.

3) **Is the mean gesture apex to vowel midpoint interval for DAF conditions significantly different that in NAF conditions?**

H_0^3: There is no significant difference between the mean GA-VM interval for the DAF condition and the mean GA-VM interval for the NAF conditions.

4) **Is there a significant interaction between pitch accent, syllable position, and DAF upon the degree of synchronization between the apices of deictic gestures and the vowel midpoint of pitch-accented syllables of compound words spoken within short utterances?**

H_0^4: There is no significant interaction between pitch accent (i.e., presence and absence of pitch accent) and syllable position (i.e., first and second position of compound word pairs) on the mean GA-VM interval.

5) **Is total gesture time greater when an auditory delay of 200 ms is imposed while producing short utterances compared to when there is no auditory delay?**

H_0^5: There is no significant difference between the mean total gesture time accompanying utterances produced with an auditory delay and the mean total gesture time during narratives produced without an auditory delay.

6) **Is gesture launch time greater when an auditory delay of 200 ms is imposed while producing short utterances compared to when there is no auditory delay?**

H_0^6: There is no significant difference between the mean gesture launch time accompanying utterances produced with an auditory delay and the mean gesture launch time during narratives produced without an auditory delay.

7) **Are the durations of utterances spoken under the influence of DAF longer than the durations of the same utterances produced without DAF?**
   H_0^7: The durations of utterances in the DAF condition are not significantly longer than the utterances in the NAF condition.

8) **Are the vowel durations of syllables with contrastive pitch accent longer than for the same syllables produced without contrastive pitch accent?**

   H_0^8: The vowel durations of syllables with contrastive pitch accent are not significantly longer than syllables produced with neutral pitch accent.

2.0 LITERATURE REVIEW

Before discussing the temporal relationship of speech and gesture, a brief overview of gesture and prosody is provided. Next, the key hypotheses and models of gesture production are discussed with their assertions regarding speech/gesture synchronization. An integration of these hypotheses/models of gesture production with models of speech production that incorporate a stage of phonological encoding, followed by a discussion of how the predictions could be explained from a dynamic systems perspective is then presented. Lastly, variables that are hypothesized to affect the temporal synchronization of speech and gesture are reviewed. Investigations of the role of prosodic prominence and perturbation on the temporal synchronization of speech and gesture are the primary focus of this critical review.

2.1 GESTURE IDENTIFICATION AND CLASSIFICATION

2.1.1 What are Gestures?

One of the first methodological obstacles confronted in gesture literature is a discrepancy of taxonomy. Gestures are typically defined as arm and hand movements that are temporally coordinated with speech (Goldin-Meadow, 1999; McNeill, 1992). Gestures do not include conventional and rule-bound movements found in sign languages. On the contrary, the majority of gestures are arbitrary and variable in form. In addition, gestures do not include nonverbal elements such as facial expressions or self- and object-touching movements. Most often in the literature, gestures consist only of manual movements and not extraneous movements of the head (i.e., nodding) or other body parts, though some investigations of gesture do include other types of body movements such as leg, foot, head, and torso movements.

2.1.2 What are the types of gestures?

In general there are two broad classifications of hand movements that accompany speech. There are those gestures that *do not* relate to the semantic meaning of the verbal message and those that *do* relate to the semantic meaning of the verbal message. These will be referred to as marking movements and meaningful movements, respectively. The various classification labels, citations, and examples for these gesture types are described in Tables 1 and 2.

The most frequently studied gesture types are those in the meaningful movement category such as deictic and iconic gestures. Deictic gestures are pointing gestures that can be used to point to both concrete and abstract referents. Deictic gestures have a relatively fixed manual construction in Western culture which consists of extending the index finger with the other three fingers folded back to the palm. Deictic gestures are also the first type of gesture to be used by children. Although the present investigation focuses on the use of deictic gestures, other types of gestures are described to provide background information for the literature review.

Iconic gestures are movements that carry some semantic meaning related to the accompanying spoken message. A similar gesture type is a metaphoric gesture. Metaphoric gestures carry more abstract semantic meaning in contrast to iconic gestures that "represent body movements, movements of objects or people in space, and shapes of objects or people...concretely and relatively transparently" (Goldin-Meadow, 2003, p. 7). Yet the distinction between degree of iconicity for iconic and metaphoric gestures is ambiguous. As Krauss and Hadar state (1999, p. 100), simply because the distinction is widely accepted does not make it useful. For that reason, iconic and metaphoric gestures are often collapsed into a single

category and are referred to as representational or lexical gestures. Consequently, in this document both iconic and metaphoric gestures are defined as representational gestures. Representational gestures are idiosyncratic and have no fixed form or movement. A single movement of a representational gesture may hold multiple meanings as well.

It is also important to note that while a representational gesture is related to the semantic content of the accompanying speech, the gesture need not have a one-to-one correspondence to a single word. In fact, it is common for a representational gesture to offer additional information that is not present in the speech stream (e.g., spreading hands wide in front of you to indicate the concept of large while stating *he gave me a present*). Likewise, an individual can use a deictic gesture to point to indicate information that is not present in the speech stream (e.g., pointing to the door while stating *he went that way*). Conversely, a gesture can directly *match* the information communicated in the speech and be redundant with the spoken message (e.g., pointing towards the ground while stating *the elevator went down*).

Marking gestures are less often investigated due to the assumption that they do not carry semantic meaning and are therefore less integrated with the speech system. Marking gestures are most often referred to as beat gestures. Traditionally these gestures are thought to mark the rhythm of an utterance (Efron, 1941; Ekman & Freisen, 1972; McNeill, 1992). However, as Feyereisen (2000, p. 150) states "evidence of their connections with prosody in normal or brain-damaged subjects is still lacking". Beat gestures are composed of short, quick movements and are the last type of gestures to emerge in development due to the presumed connection to higher-level discourse formulation (McNeill, 1992). Although marking gestures are thought to correspond with the rhythm of speech, they are not the focus of the present project because of the difficulty of controlling their presence in discourse. The goals of the present

project will be extended to more natural spontaneous speech and a variety of gestures in the future.

Table 1 *Gesture Taxonomy: Marking Movements*

Gesture Description	Gesture Type	Citation
Short and fast movements of the hands that mark the rhythm of speech or emphatically mark a lexical item or topic change.	Beats	McNeill, 1992
	nondepictive speech markers	Rime & Schiaratura, 1991
	batons	Efron, 1941; Ekman & Freisen, 1972
	motor	Krauss, Chen, & Gottesman, 2000

Table 2 *Gesture Taxonomy: Meaningful Movements*

Gesture Description	Gesture Type	Citation	Example
Movements of the hand or finger that point toward an object that is either concrete or abstract.	Deictic	Ekman & Friesen, 1969; McNeill, 1992; Rime & Schiaraturan, 1991;	Pointing to a ball while requesting. Pointing in the direction the event being described.
Movements that reflect the content of the verbal output. Often indicates qualities of an object, person, or action such as movement, size, shape, and position.	Iconic	McNeill, 1992; Rime & Schiaratura, 1991	Rotating finger while describing a spinning motion.
	symbolic	Acredolo & Goodwyn, 1988; 2002	Using an opening and closing motion of the hand in reference to a bird's beak.
	illustrators	Cohen, 1977; Ekman & Friesen, 1969	
	representational	Iverson, Capirci, Longobardi, & Caselli, 1999	
	conversational	Krauss, Dushay, Chen, & Rauscher (1995)	
Movements that reflect some abstract concept in the accompanying speech. They are image based, but abstract in nature.	Metaphoric	McNeill, 1992	Using a cupped hand as a presentation of a question (McNeill, 1992).
Movements that may or may not accompany speech but provide information to a listener based upon a shared, symbolic meaning of the movement.	Emblem	Goldin-Meadow, 2003	Placing an extended index finger in front of your puckered lips to communicate "be quiet"
	Symbolic	Rime & Schiaratura, 1991; Acredolo & Goodwyn, 1988	
Movements that reflect either the concrete or abstract content of the accompanying verbal output.	Lexical	Krauss, Chen, & Gottesman, 2000; Krauss & Hadar, 1999	Include examples for both iconic and metaphoric gestures.

2.1.3 What are the components of gesture?

Once a manual movement is identified as a gesture, the segments that comprise a gesture can also be coded. The initiation and termination of a gesture can be identified but other points within a gesture can be coded as well. The *gestural stroke* is the portion of the gesture that corresponds to the most meaningful part of the movement and is also characterized by the greatest extension of movement (McNeill, 1992, 2005). Before and after a gestural stroke, the manual movement may pause, presumably to retain synchronization with the lexical affiliate (Krauss et al., 1996; McNeill, 1992). These are referred to as the *pre-stroke hold* and the *post-stroke hold*, respectively. A gestural stroke is obligatory while the gestural holds are not. A gestural retraction phase, when the hand returns to a rest position, completes the movement to a termination point.

One other gestural segment, a *gesture apex*, has been coded in recent investigations. A gesture apex is roughly synonymous with a gestural stroke but differs in the measurement procedure. An apex is coded according to movement parameters such that the point of maximum extension of the hand or fingers is considered the apex. If the apex is held, one can also measure the initiation, hold time, and termination of the apex. The gesture apex is the dependent variable of gesture for this project.

2.2 OVERVIEW OF PROSODY

2.2.1 What is prosody?

Providing a comprehensive discussion of the prosodic characteristics of speech is well beyond the scope of this manuscript. However, it is important to provide some background information to clarify the role of prosody as it pertains to the objectives of the current project. Prosody encompasses the suprasegmental elements of speech, such as stress, rhythm, rate of speech, pauses, melody, and intonation (e.g., Crystal, 1975; Kent, 1997; Lehiste, 1970). Prosody is the reason that a single utterance can be produced in a countless number of ways and is the source of the saying *it's not what you say, but how you say it*. Prosody can provide cues for the emotional characteristics of an utterance and give an indication of the communication act of the utterance (e.g., declarative statement vs. question). The prosodic characteristics of speech are not confined to a single phonologic or syntactic unit and therefore prosodic variables can be found within units of syllables, words, and utterances.

Because prosody breaks speech down into smaller components and prosodic cues can act as perceptual cues, prosody is also a fundamental construct for spoken word recognition for infants, children, and adults (e.g., Gerken, Juszyk, & Mandel, 1994; Grosjean & Gee, 1987;

Hirsh-Pasek, Nelson, Jusczyk, Cassidy, Druss, & Kennedy, 1987). Children also develop the ability to control the prosodic characteristics of speech at an early age, even before they master control of segmental elements (Crystal, 1979; MacNeilage & Davis, 1993; Snow, 1994). Prosody can also affect speech development. For instance, less salient syllables (i.e., weakly stressed syllables) are often omitted by young children (Gerken, 1991; Schwartz & Goffman, 1995; Snow, 1998).

All anatomical structures involved in speech production, (i.e., respiratory, pharyngeal, laryngeal, nasal, oral) are also involved in the modulation and production of prosody. As such, there are a variety of ways to measure prosody, though no variable consistently reflects prosodic changes in every production. Furthermore, many variables are proposed to correlate to perceptual judgments of prosody, though there is much disagreement in the literature regarding the reliability and validity of these measures. Most often, the percept of prosody is composed of quantifiable acoustic correlates including variations in fundamental frequency (f0), amplitude, and segment/syllable duration and pause times (Emmorey, 1987; Kent, 1997; Kent & Read, 1992). Additionally, prosody can be measured physiologically using kinematic variables such as jaw displacement and velocity profiles of lip and jaw movement (Dromey & Ramig, 1998; Goffman & Malin, 1999; McClean & Tasko, 2002, 2004; Schulman, 1989) and even rib cage displacement (Connaghan, Moore, Reily, Almand, & Steeve, 2001). Likewise, almost all breakdowns in communication, whether they are higher-level deficits such as language impairments in children and aphasia in adults or lower-level deficits such as dysarthria, are accompanied by some element of abnormal prosody and the corresponding percept of unnatural speech.

Prosody is a component of all levels of the speech and language system. For instance, parameters of prosody have been associated with the level of discourse (e.g., emotional intonation patterns), syntax (e.g., phrase marking), phonology (e.g., stress assignment), and articulation (e.g., rate of speech) (Emmorey, 1987; Hird & Kirsner, 1993; Levelt, 1989). Because prosody interfaces with both linguistic and phonologic levels of processing, prosodic variables provide a unique opportunity to examine the point of translation between higher-level linguistic processes and lower-level phonologic and articulatory processes. Most importantly for the current discussion, the accessed prosodic structure representation is hypothesized to be the conduit between higher and lower levels of speech and language planning. As Wheeldon (2000) states, "the claim is that, following the generation of the syntactic structure of an utterance, an intervening prosodic structure is generated that groups words into prosodic units and that it is these prosodic units which guide the generation of the phonological form of the utterance" (p. 249). Thus, the generation of prosodic structure is critical in converting syntactic representations to executable speech.

Though what are the actual units of prosodic structure? The past two to three decades have resulted in a surge of interest prosodic theory in the field of linguistics. A number of individuals have led the way on developing theories in relation to prosodic structure (e.g., Beckman & Edwards, 1994; Halle & Vergnaud, 1987; Hayes, 1986, 1989, 1995; Gussenhoven, 2004; Liberman & Prince, 1977; Nespor & Vogel, 1986; Pierrehumbert, 1980; Selkirk, 1980, 1984, 1986, 1995, 1996). However, there are discrepancies within the linguistic literature regarding the number and type of prosodic units. The prosodic constituents proposed by Elisabeth Selkirk (1984) are adopted here given that these constituents are also incorporated within the models of speech production (Ferreira, 1993; Levelt, 1989) that will be employed in

this investigation. For a thorough tutorial on differing viewpoints on the units of prosodic structure please refer to Shattuck-Hufnagel and Turk (1996).

Prosodic constituents are ordered in a hierarchical structure according to a Strict Layer Hypothesis. The Strict Layer Hypothesis asserts that a prosodic unit is composed of one or more units that are in the domain immediately subordinate to that level. The constituents proposed by Selkirk include the utterance (U), intonational phrase (IPh), phonological phrase (PPh), prosodic word (PWd), foot (ω), and syllable. Refer to Figure 4 for an example of the layers of the prosodic hierarchy.

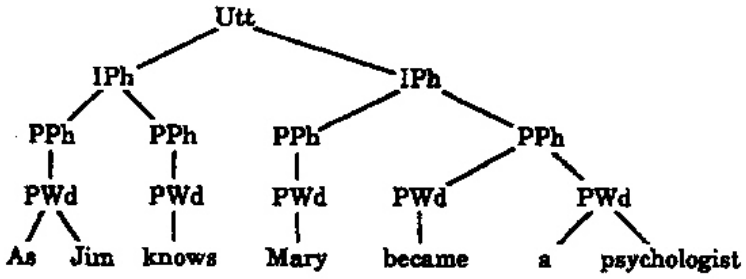


Figure 4 Prosodic hierarchy. Adapted from "Creation of prosody during sentence production" by F. Ferreira, 1993, *Psychological Review, 100*, p. 236.

The dominating unit of the prosodic hierarchy is the utterance (U) which is made up of one or more intonational phrases (IPh). Intonational phrases are thought of as idea units and are defined by prosodic boundary cues such as pausing, pitches resetting, and syllable lengthening (Gerken & McGregor, 1998). As we will discuss later, pitch accent is assigned at the level of the intonational phrase. The subordinate unit to an intonational phrase is a phonological phrase

(PPh). Phonological phrases are most similar to the structure of syntactic phrases (e.g., noun phrases and verb phrases), though they often do not have a one-to-one correspondence to syntactic boundaries. Selkirk summarizes the composition of phonological phrases according to the X-max algorithm (Selkirk, 1986). The X-max algorithm states that the syntactic structure is scanned and the information up to and including the right bracket of a syntactic phrase is included in a phonological phrase.

Prosodic words (PW) are the next level of the prosodic hierarchy. In the simplest terms, a prosodic word can be thought of as a lexical (i.e., content) word plus any adjacent grammatical morphemes (i.e., function words). For example, the phrase *the ocean* forms only one prosodic word. Metrical stress is assigned at the level of the prosodic word. There are also constraints for the stress patterns assigned to a prosodic word. These constraints are reflected in the foot (ω) structure of the prosodic word, which is the next level of the prosodic hierarchy. In American English, metrical feet are either trochaic or monosyllabic. A trochaic foot consists of a strong syllable followed by a weak syllable and a monosyllabic foot consists of a single syllable which can be either weak or strong. Linguists have posited that iambic feet (i.e., a weak syllable followed by a strong syllable) are not permissible in American English but that it is permissible for some syllables to be unfooted. An unfooted syllable descends from a prosodic word rather than a foot. Furthermore, prosodic words in American English can have only two unfooted syllables, one on either edge.

Consequently, syllables are the smallest and final constituent of the prosodic hierarchy, though even a syllable can be broken down into its phonetic components. The syllable is comprised of an onset, nucleus, and coda (i.e., the final phoneme of the syllable). Together the nucleus and coda are referred to as a rime. The nucleus is obligatory while the onset and coda

are not. As discussed above, each individual syllable can be footed or unfooted and can be weak or strong in metrical stress assignment. Each syllable can also be pitch-accented within an intonational phrase. Now let's turn to discuss prosodic stress in more detail.

2.2.2 What is prosodic prominence?

At the outset, it is important to note that there is a distinction to be made between prosodic prominence and intonation. Bolinger (1958) was among the first to propose a greater specification of intonation across an utterance. Bolinger and other contemporary linguists (e.g., Cruttenden, 1997; Gussenhoven, 2004; Hayes & Lahiri, 1991; Ladd, 1996; Liberman, 1975; Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988) discuss this distinction at length, even though terms like prominence, stress, accent, focus, rhythm, and intonation are still often used interchangeably both in research and clinical arenas. The difference between prominence and intonation is discussed given this frequent confusion of terminology.

Prosodic prominence is typically characterized by stress. Most often, stress is thought of as a dichotomy, i.e., a syllable is either stressed (i.e., strong) or unstressed (i.e., weak). Several types of stress have been posited including metrical (i.e., lexical) stress, sentential stress, and contrastive stress (i.e., emphatic stress, also termed accent). A syllable can be assigned stress at a lexical level as metrical stress or at a higher level of prominence as pitch accent.

Prosodically prominent syllables are temporally organized to create a rhythmic structure for speech production. That is, stressed and unstressed syllables are successively arranged in such a way that a rhythm (i.e., meter) of speech is attained. Languages vary with respect to the

ways in which sequences of syllables are assigned stress value. Specifically, English has been described as a stress-timed, isochronous language, in which stress tends to be distributed evenly across periodic intervals in accordance with a preferred metrical pattern of alternating strong (S) and weak (W) syllables (Abercrombie, 1967; Hayes, 1985; Lehiste, 1977; Liberman & Prince, 1977; Nespor & Vogel, 1989; Pike, 1945; Selkirk, 1984).

In contrast, intonation is the melody of an utterance rather than the rhythm of an utterance. An utterance can be produced multiple times with the same stress and rhythm pattern, but with very different intonation patterns. The primary function of intonation is to convey communicative information such as providing the distinction between a declarative statement and a question (Ladd, 1980, 1996) and relaying the emotion associated with an utterance. Levelt states that intonation "expresses a speaker's emotions and attitudes" and "is a main device for transmitting the rhetorical force of an utterance, its obnoxiousness, its intended friendliness or hostility" and "signals the speaker's intention to continue or to halt, or to give the floor to an interlocutor" (1989, p. 307). Also in contrast to prominence, intonation is related to changes of pitch rather than temporal changes of phonetic segments.

With that being said, let's return to discussing two specific types of prosodic prominence, stress and pitch accent. Syllables are stressed by increasing the effort involved in producing the syllable. Hence, stressed syllables are correlated to changes in duration, amplitude, and/or fundamental frequency of a syllable (Lehiste, 1970). More recently, differences in spectral tilt between stressed and unstressed syllables were also measured (Sluijter, 1995; Sluijter & van Heuven, 1997). Additionally, vowel quality differs for stressed and unstressed syllables. Stressed syllables are composed of full vowels compared to unstressed syllables that are usually composed of reduced, centralized vowels. These acoustic changes lead to an increase in

perceptual salience or prominence. As stated earlier, no single acoustic variable directly and consistently corresponds to the percept of stress, though durational measures tend to be the best acoustic predictor of stress. Research has demonstrated that durations of stressed syllables are significantly longer (65-100%) in duration than unstressed syllables (Fry, 1955; Kent & Reed, 1992; Liberman & Prince, 1977; Rusiewicz, Dollaghan, & Campbell, 2003). Stress also tends to be acoustically realized within the nucleus (i.e., vowel) of a syllable (Edwards, Beckman, & Fletcher, 1991; de Jong 1991, 2004; Harrington, Fletcher, & Roberts, 1995; Summers, 1987).

Individual lexical items hold intrinsic metrical stress (i.e., lexical stress). Metrical stress is "a structural, linguistic property of a word that specifies which syllable in a word is in some sense stronger than any of the others" (Sluijter & van Heuven, 1996, p. 2471) and is an "abstract feature on the level of the lexicon" (van Kuijk & Boves, 1999, p. 96; see also Kent & Read, 2002). The dominant stress pattern of English is a strong-weak stress pattern, i.e., a trochaic pattern. An example of a trochee is *puppet*. Conversely, an iambic stress pattern consists of a weak syllable followed by a strong syllable (e.g., *baboon*).

A second type of prosodic prominence is pitch accent. Unlike metrical stress, which is associated with individual words, pitch accent is assigned to syllables at the intonational phrase level. It is important to note, that like stress and rhythm, pitch accent assignment varies considerably across languages. Only pitch accent in American English is summarized here.

Pitch accent is often discussed as a prominence marker as well as a landmark for the creation of an intonational contour. The aims of the current project involve the former function of pitch accent. Ferreira (1993, p. 238) nicely summarizes the definition of pitch accent as a "general term to describe the presence of some sort of prosodic prominence on an element of a sentence". She continues to state that "a pitch accent affects the likelihood of an intonational-

phrase boundary" (p. 238). Additionally, pitch accent "expresses the prominence of a concept, the interest adduced to it by the speaker, or its contrastive role" (Levelt, 1989, p. 307). Also, pitch accent "is a phonetic feature, with measurable correlates in production, acoustics, and perception…in principle, each word, including monosyllabic function words, can be accented" (Van Kuijk & Boves, p. 96; see also Kent & Read, 2002). Hence, a syllable can be stressed or unstressed but then can also be accented or unaccented, though typically only stressed syllables are accented in spontaneous conversation. Pitch accent, by definition, is characterized by either a quick rising or falling of pitch, though few researchers have actually conducted acoustic analyses of the phonetic realization of pitch accent (e.g., Grabe, 1998; Neijt, 1990). In fact, there is not only evidence that accented syllables have increased vowel durations (Beckman & Cohen, 2000; Cooper, Eady, & Mueller, 1985; de Jong, 2004; Turk & White, 1999), but also that pitch accent is associated with increased rime duration, even when pitch accent is placed on a normally unstressed syllable (Sluijter & van Heuven, 1995). Thus, duration is an acoustic correlate of both metrical stress and pitch accent. Campbell summarizes, "durational lengthening serves as a cue to stress and focus (i.e., accent) marking in speech" (2000, p. 322).

There are numerous pitch accent intonational patterns such as H* (*high*), L* (*low*), H*L (*high-low*), and L*H (*low-high*). In these notations, the asterisked tone is the pitch-accented tone and it may be preceded or followed by a second rising or falling tone. A number of factors come into play when designating a syllable as a pitch accented syllable. Most often it is a stressed syllable that is given pitch accent. There is usually only one syllable that is accented in an intonational phrase (Beckman, 1986; Hayes, 1995; Gussenhoven, 1991, 2004; Shattuck-Hufnagel, 1995). Another variable that determines whether or not a syllable is assigned pitch accent is the context and focus of the utterance. Pitch accent is more likely to be assigned to new

or contrastive lexical items in an utterance rather than known lexical items. When a sentence is produced, the greatest amount of prominence will be given to the pitch accented lexical item, regardless of the metrical structure of the remainder of the utterance. As Levelt (1989, p. 305) states, "pitch accent will overrule everything else" during spoken language planning and execution. Pitch accent is not to be confused with idiosyncratic emotional changes in f0 such as an increase in f0 associated with excitement. However, it is true the subjective distribution of pitch accent makes the empirical study of the phenomenon difficult, especially for the acoustic correlates of pitch accent in spontaneous conversational speech. In fact, this individual and context dependent assignment of pitch led Bolinger to write his 1972 document aptly titled, *Accent is Predictable (If You're a Mind Reader)*. As a result, it is critical that the location of pitch accent be controlled during empirical study.

Consequently, that is why contrastive pitch accent was chosen as the specific independent variable for the current project. Contrastive pitch accent is also commonly referred to as contrastive or emphatic stress. Ellis Weismer and Hesketh (1998, p. 1445) state that the function of contrastive stress is to act as a "focusing device, with its placement within the utterance being more dependent upon situational or pragmatic factors than grammatical or semantic factors". A lexical item is given contrastive accent when it is novel or in opposition to the context or line of discourse. For instance, Speaker 1 may ask, *did the running back score the touchdown?*, while Speaker 2 clarifies, *no, the QUARTERBACK, scored the touchdown*. In this case the contrast is between the types of players. Pitch accent is assigned to designate the lexical item as more important because of its contrastive role. Manipulation of contrastive conditions allows for predictable assignment of pitch accent.

2.2.3 Pitch Accent and the phonological encoder

The concept of phonological encoding was defined by Levelt (1989) as the process by which a lemma is translated into a phonetic plan as a string of pronounceable syllables via morphological, metrical, and segmental representations of the lemma. Levelt also states that the duration of syllables is set at this stage. A great deal of interest has been generated on the role of phonological encoding in recent years, though the many responsibilities of the phonological encoder are not of relevance for this project. This discussion focuses solely on the creation of pitch accent within Prosody Generator of the phonological encoder. This is one of the few components of speech production that Levelt theorizes about which span a unit larger than a prosodic word. However, his theory pitch accent assignment across an utterance was largely unsubstantiated. Fortunately for the purposes of this series of experiments as well as for consistency with the models described previously, Ferreira (1993) has applied data from her work with prosody to broaden Levelt's 1989 model to span the planning of multiword utterances. Moreover, her study was largely based on contrastive pitch accent protocols. Ferreira's study and the mechanism of contrastive pitch accent placement according to Ferreira and Levelt are jointly reviewed.

Ferreira (1993) conducted four experiments that demonstrated that prosodic structure is a psychological reality in sentence production and that the processing of prosodic structure is independent of the processing of phonetic segments. Of these four experiments, two employ a contrastive stress paradigm which consisted of undergraduate students (n=10 in each experiment) silently reading a short sentence on a card and then reading it aloud. Each of the ten sentences

were produced both in a *neutral* condition without contrastive stress and in a *prominent* condition with contrastive stress. The participants were instructed to emphasize the capitalized word in the prominent condition and were also provided a leading question that provided additional cause to emphasize the target item. An example sentence pair is "The crate contains the missing book" compared to "The CRATE contains the missing book". Findings included an increase in duration of both the target word and following pause in the prominent condition relative to the neutral condition. Ferreira also manipulated intrinsic word duration in the second contrastive stress paradigm so that a word with a vowel of long duration (e.g., *mouse*) was compared to vowel of short duration (e.g., *cat*) in both prominent and neutral conditions. Manipulation of intrinsic word duration yielded the finding that "word and pause durations trade off so as to respect the total duration of the interval allocated to a word" (p. 245). That is, word duration was increased in the prominent conditions regardless of intrinsic words duration, though the following pause duration actually decreased in the prominent condition when intrinsic word duration was long.

Ferreira (1993) elegantly translated these findings to provide greater specification to Levelt's (1989) model of speech production. As mentioned earlier, Levelt specifies the production processes for single lexical items (i.e., prosodic words) to a greater degree than for sentential contexts. According to Levelt (1989) prominence markers, in our case pitch accent for the purpose of contrastive emphasis, are first designated within the conceptualizer during planning of the preverbal message prior to retrieval of a particular lemma for that discourse representation. A prominence marking is converted to a pitch accent and related acoustic parameters in later processing within the Formulator.

Prosodic features are planned specifically within the Prosody Generator within the phonological encoder within the formulator (see Figure 5). Levelt states that the Prosody Generator is a "processing component that computes, among other things, the metrical and intonational properties of the utterance" (p. 365). The Prosody Generator is responsible for the creation of the prosodic hierarchy constituents (e.g., prosodic words, phonological phrases, etc.) and the metrical grid for these constituents across the entire utterance. Extra beats are assigned to a lexical item if it is designated as having pitch accent after its metrical stress is retrieved from the lexicon. Ferreira (1993) modifies Levelt's proposal to further specify that a metrical grid "is constructed for the sentence to represent its overall stress and timing pattern and to reflect the changes in its metrical pattern brought about by the sentential context:" (p. 247).

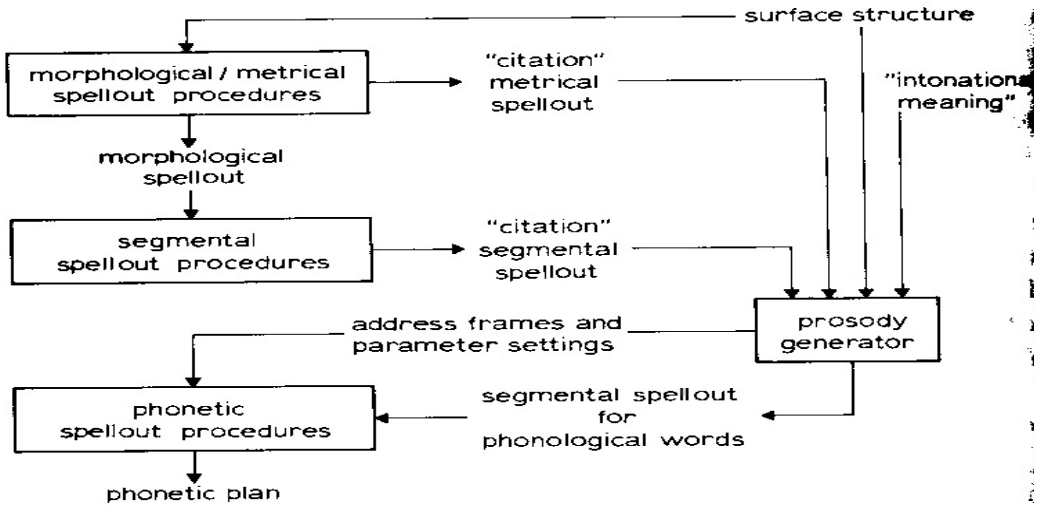


Figure 5 Stages of phonological encoding according to Levelt (1989). Adapted from Levelt, W.J.M. (1989). "Speaking: From intention to articulation" by W.J.M Levelt, 1989, Cambridge: MIT Press, p. 158.

Stress and timing across an utterance are affected by a number of elements according to metrical grid proposals originally laid out by Selkirk (1984) and as adapted by Levelt (1989) and Ferreira (1993). Let's utilize one of Ferreira's examples for discussion purposes. Her example utterance is *the girls argued*. First, the level of stress on the metrical grid is dependent upon metrical stress assignment. In our example, the first word is a function word and does not receive metrical stress, but rather is thought to cliticize with an adjacent content word. The second word, *girls*, is a one-syllable content word which does have metrical stress. The final word, *argued*, is a two-syllable content word that is a trochee with metrical stress with first syllable metrical stress assignment.

	x	x	
x	x	x	x
x	x	x	x
the	girls	ar	gued

According to Ferreira's findings (1993), the second element that affects stress and timing across an utterance is prosodic constituent boundaries. That is, words in the final positions of prosodic constituents (e.g., prosodic words, phonological phrases, intonational phrases, utterances) receive additional stress. The boundaries in our example can be notated as the following (PWd=prosodic word; PPh=phonological phrase; IPh=intonational phrase; Utt=utterance):

((((the girls) PWd) (argued) PWd) PPh) IPh) Utt

An extra beat is added for each boundary. In the case of multisyllabic words, extra stress is added to the already metrically stressed syllable. The first word has no prosodic constituent boundaries, the second word has two boundaries and the final word has four boundaries. The grid below reflects the alterations that result.

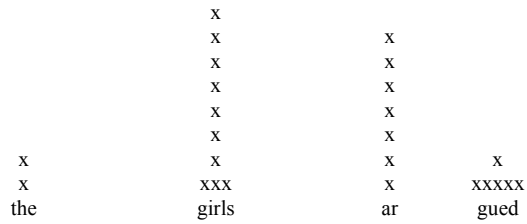
		x	
		x	
	x	x	
	x	x	
	x	x	
x	x	x	x
x	x	x	x
the	girls	ar	gued

Ferreira also proposes that prosodic constituent boundaries affect the assignment of silent demibeats. Selkirk (1984) initially postulated that silent demibeats were assigned at the right boundary of prosodic, rather than syntactic, constituents. Silent demibeats correlate to durational timing features in contrast to stress assignment indicated by beats. Hence a beat *x* is added vertically, and a demibeat *x* is added horizontally for every boundary. For multisyllabic words, the silent demibeats are added to the final syllable, regardless of its metrical stress assignment due to the acoustic-phonetic phenomenon of syllable final lengthening. The modified grid is as follows:

		x	
		x	
	x	x	
	x	x	
	x	x	
x	x	x	x
x	xxx	x	xxxxx
the	girls	ar	gued

Pitch accent is realized by placing additional stress and lengthened segment duration for the target word above and beyond the stress of any other word in the utterance according to the pitch-accent prominence rule (Levelt, 1989; Selkirk, 1984). Even though it would seem that

some look-ahead mechanism is necessary to assure that the syllable is designated as having more prominence than later syllables, Levelt proposes that pitch accent can be assigned incrementally like all other processes in his speech production model. He states that once stress is assigned to the pitch accented syllable, then all subsequent syllables will be assigned stress that is less than that level. Ferreira does not address this issue. However, the assignment of stress and timing based upon metrical stress and prosodic boundaries may be difficult to reconcile with incremental assignment of pitch accent prominence if much greater prominence is placed on a later word due to the number of boundaries it has in comparison to early words. This is the case in our example where the final word has two more beats and demibeats than the preceding word due to prosodic constituent boundaries. For instance, if contrastive accent is placed on the second word of our example, *girls*, then greater prominence is placed on that word than any other word. In order to be consistent with Ferreira's model of prosody processing in utterances we will assume that prosodic constituent boundaries determine the initial metrical grid but that a look-ahead mechanism is necessary to assign greater stress to an accented word when the most prominent syllable is in a later incremental position. In other words, the metrical grid for an utterance is constructed according to the metrical structure accessed in the lexicon, modified given the prosodic constituent boundaries, then the additional beats required to make the pitch accented word the most prominent of the utterance are assigned. Therefore, if contrastive accent is placed on *girls* in reference to the question, *Who was arguing?*, the grid would look like:



Thus, de Ruiter's model for gesture production processing was adapted using the parallel speech production component of the model to incorporate Ferreira's findings and hypotheses of the generation of prosodic prominence across an utterance. de Ruiter's hypothesis that gesture and speech production processes do not interact below the level of the conceptualizer remains the same following this modification. Though, the enhancement of the processes responsible for the assignment of pitch accent allows for greater validity and subsequent explanatory worth for the manipulation of prosody within the present investigation.

This summary has provided a cursory and admittedly over-simplified review of prosody and prosodic prominence. Notably, these distinctions between different types of prosodic prominence, as presented within this section, rarely are made in the literature pertaining to the effect of prosody upon the timing gesture. Also, the construction of rhythm and assignment of pitch accent across an utterance has not been incorporated into any model of gesture production. In fact, few theorists have considered the effect of prosodic prominence at all upon the production of gesture and opposing views on the role of prosody exist among their hypotheses. The next section presents these hypotheses/models of gesture production and their statements about the temporal relationship of speech and gesture.

2.3 THEORIES OF GESTURE PRODUCTION

In order to complete a systematic examination of the relative timing of gesture and speech, it is also essential to approach the research questions within a theoretical framework, rather than relying on anecdotal observations. However, rarely have researchers accomplished this goal. To be fair, this is most likely due to the absence of testable models of gesture production to use as a foundation for their research questions. Until fairly recently, theories of gesture production have consisted of nonspecific statements and assumptions in contrast to a more desirable predictive model with explicit points of interaction between the speech and gesture systems and subsequent effects upon the temporal expression of gesture in relation to the speech signal. Indeed, both types of models, those with stringent predictions and those with more general hypotheses bring something to scientific progress and innovation. In this section, a total of five theories and models of gesture production are reviewed. This is not an exhaustive review of every theory or model of gesture production in the literature. There are a number of other thought-provoking accounts of the production and purpose of gesture that are not included in this review. The majority of these authors conjecture that gesture and speech are completely independent processes (e.g., Butterworth & Beattie, 1978; Butterworth & Hadar, 1989; Feyereisen & deLannoy, 1991; Hadar, 1989; Hadar, Wenkert-Olenik, Krauss, & Soroker, 1998) and others are

concerned only with the shared linguistic formulation processes of speech and gesture (e.g., Kita & Özyürek, 2003; Morsella & Krauss, 2004). Accordingly, only those theories and models that are dependent upon the tight temporal synchrony of speech and gesture are presented. The first model, the Sketch Model (de Ruiter, 1998, 2000) provides the foundation for the null hypothesis of the current investigation. Following the discussion of the Sketch Model, a brief synopsis of the Growth Point Theory (McNeill, 1985, 1987, 1992, 2000, 2005), the Rhythmical Pulse Hypothesis (Tuite, 1993), the Facilitatory Model (Krauss, Chen, & Chawla, 1996; Krauss, Chen, & Gottesman, 2000), and the Entrained Systems Model (Iverson & Thelen, 1999) is presented.

2.3.1 Sketch

de Ruiter's (1998; 2000) Sketch Model, is perhaps the best-specified, though least tested model of gesture production to date. According to the Sketch Model, the primary purpose of gesture is to communicate to the recipient, similar to McNeill's stance. In contrast to McNeill, the model does not rule out the ability of gesture to enhance lexical retrieval and formulation processes for the speaker, as Krauss and colleagues counter (Krauss, Chen, & Chawla, 1996; Krauss, Chen, & Gotttesman, 2000). de Ruiter is among the first to address the need for specific points of interaction between gesture and speech production along with the consequences upon the synchronization of the two. In spite of the headway made by de Ruiter, there are both theoretical notions from others in the gesture literature as well as sparse, empirical evidence to suggest that certain predictions of de Ruiter's Sketch model should be challenged.

The Sketch Model is an extension of Levelt's (1989) model of speech production (Figure 2). The reader is encouraged to review Levelt's comprehensive text for a complete account of the model. In brief, Levelt's model is a stage-based, information-processing model that assumes that speech production proceeds through several stages beginning with creating communicative intentions and accessing representations in long-term and working memory for single word encoding. The communicative message is generated at the conceptualization stage via spatial, propositional, and kinesthetic representations. The resulting preverbal message is then encoded grammatically and phonologically with access to the lexicon at the formulation stage. Also during the formulation stage, the Prosody Generator assigns metrical patterns to the lexical items and determines the distribution of stressed syllables and assigns the phonetic correlates of pitch accent. This step in the formulation stage is critical for the predictions of Experiment 1. Finally, the motoric execution of the subsequent phonetic plan is completed at the articulation stage. Speech production planning is incremental according to Levelt and the constructed unit is a *phonological word* which can be composed of more than one lexical item (i.e., clitics along with their associated content words), though the model does consider some aspects of multiword utterance generation such as pitch accent assignment.

According to de Ruiter, gesture production is completed in parallel to speech production, with a specific point of interaction between the two systems (Figure 1). The conceptualizer is the only segment of the speech production system that is linked to the gesture system. de Ruiter asserts that gestures originate in the conceptualizer by accessing spatiotemporal information from working memory for the gesture while propositional information is accessed for the preverbal message. The output of the conceptualizer is the preverbal message, which is sent to the formulator and the *sketch*, which is sent to the *gesture planner*. The information that is

encoded in the sketch differs depending on the type of gesture that is to be produced. He summarizes the encoded information for a deictic gesture, the gesture of interest for this investigation, is "a vector in the direction of the location of the referent…(and) a reference to the appropriate pointing template in the gestuary" (2000, p. 295). A *gestuary* stores templates for gestures and gesture conventions such as those that are important for deictic gestures. The gestuary will hold a template for the conventualized pointing movement (i.e., extended index finger with other fingers retracted). However, information such as the direction of movement, speed of movement, position of the hand in space, and even which hand is used is dependent upon many online factors. Therefore, the template is just that, "an abstract motor program" (p. 296). It is modified given the semantic and physical context in which the deictic gesture is produced. The sketch is sent to the gesture planner after the gesture template is retrieved from the gestuary and encoded within the sketch.

The gesture planner then constructs a motor program for the deictic gesture after accessing the motor template in the gestuary and information about the speaker's environment (e.g., which hand is free, how large is the gesture space, location of the referent). The gesture planner sends the motor program for the deictic gesture to the lower level motor control modules(s) (2000, p. 297) which then executes the movement.

This model was chosen as the theoretical foundation of this research not only because it poses explicit points of interaction between the speech and gesture systems, but also because the model is the only one to provide hypotheses regarding the temporal relationship of speech and gesture production in adults. de Ruiter (2000) rightly acknowledges the daunting task of reliably determining the temporal boundaries of gestures as well as classifying the relevant lexical/conceptual units to utilize as the reference points for temporal measures of synchrony.

He states that gestural onsets are "roughly" synchronized with the onset of "conceptual affiliate" (p. 291). With consideration of this potential caveat, de Ruiter puts forth several hypotheses regarding the synchronization speech and gesture. First, the model predicts that gestural onset will precede the onset of the conceptual affiliate due to the assumption that "the preverbal message is sent to the formulator only after the gesture planner has finished constructing a motor program and is about to send it to the motor-execution unit…once the motor program for the gesture has been constructed, the gesture planner will send a message to the conceptualizer specifying when the generation of speech can be initiated" (p. 299).

The Sketch Model also makes predictions regarding the synchronization of speech and multiple phases of gesture, specifically pre- and post-gestural holds. A pre-gestural hold occurs when a manual movement begins then halts for the conceptual affiliate's onset prior to executing the gestural stroke. According to the model, the gesture sketch can be sent to the gesture planner with subsequent initiation of the gesture while the preverbal message is sent to the Formulator after the sketch is sent. Once the preverbal message moves on to the Formulator, the conceptualizer alerts the gesture planner to continue with the execution of the gestural stroke. It is not clear from the model's predictions though whether the stroke then also precedes the conceptual affiliate since the signal to resume gesture stroke execution occurs prior to grammatical and phonological encoding in the formulation stage of speech production. A post-gestural hold occurs when the "hand remains motionless after the stroke has been completed until the related speech has been fully produced (de Ruiter, 2000, p. 299). According to the model, the gesture planner only receives a signal from the conceptualizer to retract the gesture after completion of the preverbal message. Again, the precise relationship between gestural onset and conceptual affiliate execution is not clear from the model. de Ruiter purposefully

remains vague on the specific temporal parameters of the speech-gesture relationship because of the aforementioned difficult prospect of identifying affiliates and reliably measuring the on and offsets of gesture and speech, as well as a lack of empirical investigations to guide theoretical hypotheses.

A key tenet of the Sketch Model is that there is no communication between the speech and gesture systems below the level of the conceptualizer. de Ruiter states that "the model does not permit representations active in the formulator to influence the timing of gesture" (p. 305) and there is no feedback from the Formulator available to the gesture planner. Consequent to this prediction, prosodic stress should have no effect upon the timing of the gesture. In fact, de Ruiter explicitly states that "lexical stress or pitch accent are therefore predicted to have no effect on speech/gesture synchronization" (pp. 305-306). Likewise, de Ruiter also claims that perturbation of speech cannot affect the timing of gesture execution because this is below the level of the conceptualizer. Therefore, once the gesture is initiated it cannot be altered by changes in the speech signal such as a hesitation, speech error, change in speech rate, etc. Again, he explicitly claims that "once the gesture stroke has started to execute, it can no longer be interrupted" (p. 306). The current project aims to test these two predictions.

While de Ruiter is one of few theorists to consider the role of prosodic stress on the timing of gesture and speech, he is the only to posit that prosodic stress does *not* affect the timing of gesture. For instance, McNeill's phonological synchrony rule from his seminal work (1992) states simply the opposite such that the stroke of the gesture synchronizes with the most prominent syllable of the accompanying speech. Similarly, Tuite's hypothetical relationship between gesture and speech production is built upon the notion that gestures synchronize with prominent syllables. Two other notable theories presented by Krauss and colleagues (1996;

2000) and Iverson and Thelen (1999) discuss the importance of speech/gesture synchronization, though not in reference to prosodic variables. A brief review of those theories of gesture production as they relate to temporal synchronization and the Sketch Model follows.

2.3.2 Growth Point

de Ruiter's hypothesis regarding the purpose of gesture parallels that of David McNeill's (1992) which repeatedly contends that the goal of a gesture is to communicate. McNeill's work is often referred to as the Growth Point Theory based upon his postulation that "the growth point is the speaker's minimal idea unit that can develop into a full utterance together with a gesture" (1992, p. 220). McNeill's work is the most frequently cited theory, though a lack of testable predictions diminishes its validity. Nevertheless, McNeill's (1985, 1987, 1992, 2000, 2005) hypotheses have generated great interest in gesture and inspired the surge of empirical research of the relationship of gesture and speech conducted in the past two decades. Another positive consequence of McNeill's Growth Point Theory is the spawning of more thorough and predictive models such as de Ruiter's Sketch Model. The basic premises of McNeill's work are reviewed in order to provide the reader with a historical perspective of the issues as well as to observe the pertinent similarities and differences with the Sketch Model particularly in reference to the temporal synchronization of speech and gesture.

McNeill (1985, 1987, 1992, 2000, 2005) asserts that speech and gesture are part of a fully integrated system with the combined purpose to communicate a speaker's underlying mental representations. He states that "gestures exhibit images that cannot always be expressed in

speech, as well as images the speaker thinks are concealed…speech and gesture must cooperate to express a person's meaning." (1992, p. 11). McNeill's theory further stresses that gestures are manifestations of a person's inner "thought" process and that gestures are "the person's memories and thoughts rendered visible" (p. 12). A critical hypothesis is that gesture and speech are tightly linked at all possible levels of the speech formulation and speech production process.

The support for this hypothesis is primarily anecdotal and mostly based upon individual observations of adults and children during cartoon narration tasks, most often a Sylvester and Tweety cartoon (see Figure 6). An example is of a man saying "and he bends it way back" while making an arcing motion with his hand. This example is frequently cited as evidence for the communicativeness of gestures and evidence for an integrated gesture and communicative system.

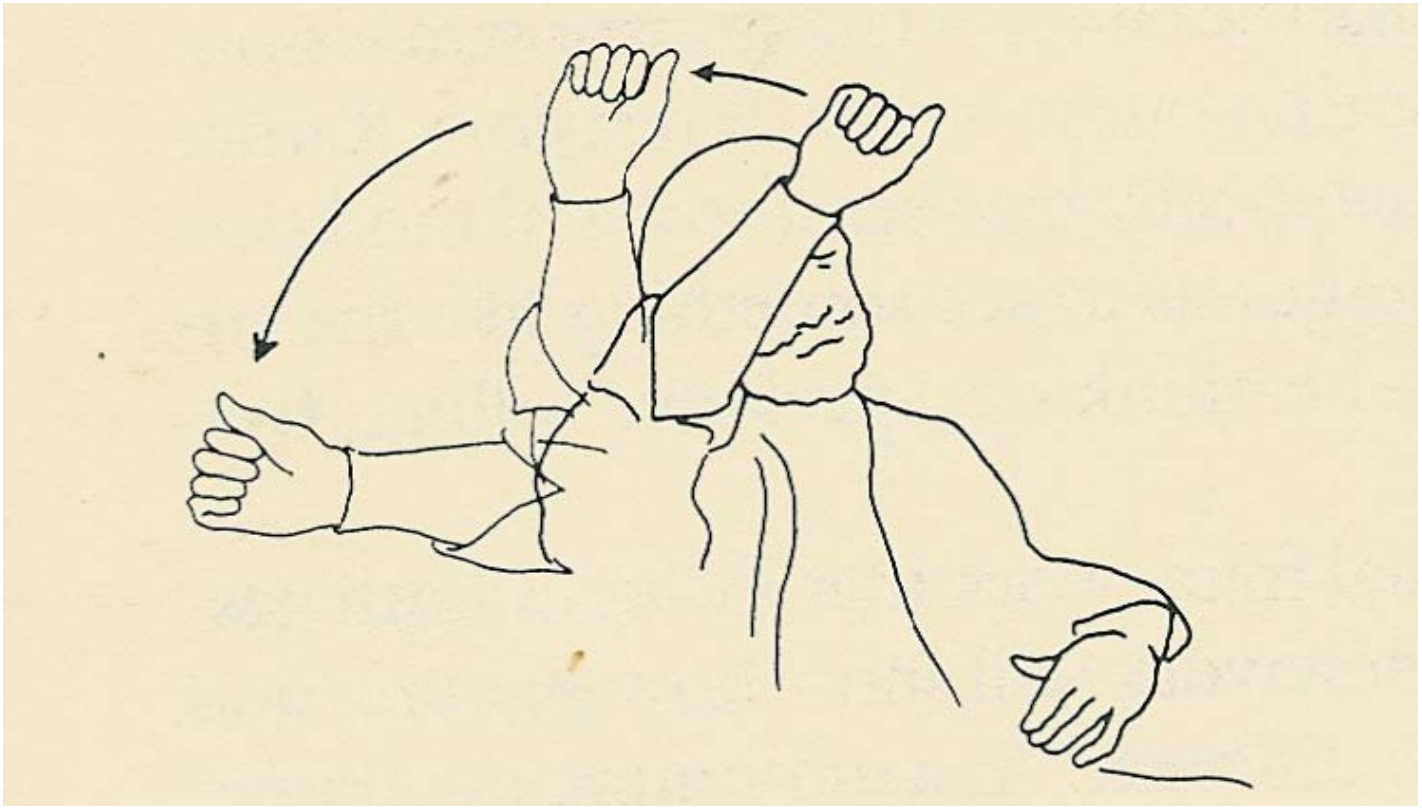


Figure 6 Sylvester and Tweety cartoon narration example. Adapted from "Hand and Mind: What Gestures Reveal about Thought," 1992, D. McNeill (Ed.), Chicago: The University of Chicago Press, p. 13.

McNeil (1992) provides five arguments for the proposed integrated system (pp. 23-24):

1. Gestures occur only during speech.
2. Gestures and speech are semantically and pragmatically coexpressive.

3. Gestures and speech are synchronous.
4. Gestures and speech develop together in children.
5. Gestures and speech break down together in aphasia.

Upon first glance the arguments seem intuitive and plausible. Yet, the theory is lacking virtually any support from empirical investigations and the arguments are imprecise and are insufficiently operationalized.

The third argument is particularly relevant for the current project. McNeill (1992) emphasizes the tight coexpressive synchrony of speech and gesture, though synchrony is never properly defined or quantified. Of the three phases of gesture production, preparation phase, stroke phase, and retraction phase, the Growth Point Theory notes that the stroke phase is the fundamental portion of the gesture and most important for the synchronization to speech. Also, McNeill rightly acknowledges the potential differences in temporal synchrony measurements based upon different gesture phase dependent variables. He states, "the synchrony rules refer to the stroke phase: anticipation refers to the preparation phase…it is only the stroke of the gesture that is integrated with speech into a single smooth performance, but the preparation for the stroke slightly leads the coexpressive speech" (p. 26). This distinction between gesture phases will also be important in our later discussion of equivocal findings of empirical studies of the temporal relationship of gesture and speech.

McNeill hypothesizes three "synchrony rules": the Phonological Synchrony Rule, Semantic Synchrony Rule, and Pragmatic Synchrony Rule. The Semantic Synchrony Rule posits that speech and gesture present the same semantic meaning, or "idea unit", simultaneously while the Pragmatic Synchrony Rule posits that speech and gesture serve a shared pragmatic function. The Semantic and Pragmatic Synchrony Rules are not applicable for this project. The

Phonological Synchrony Rule does relate to the current research questions and is in stark contrast to de Ruiter's hypothesis regarding prosodic stress and gesture timing. The Phonological Synchrony Rule represents the temporal properties of the relationship of speech and gesture and posits that the gestural stroke either occurs with or ends at the "phonological peak syllable" of the speech. However, the Growth Point Model neither defines how to determine the "phonological peak syllable" nor what is considered a unit of speech within which to classify prosodic stress.

There is no doubt that the Growth Point Theory has sparked considerable interest in the potential relationship between speech and gesture. Also, the hypotheses articulated by McNeill are certainly accepted by many. On the other hand, the hypotheses regarding the temporal synchrony of gesture laid out by McNeill are not supported by controlled empirical inquiry and are also not built within an information-processing model framework to provoke investigation that could support or refute specific points of interaction between gesture and speech production. The next hypothesis of our discussion is even more poorly specified than the Growth Point hypothesis, though it is the hypothesis of gesture production that is most founded on the relationship to prosodic stress.

2.3.3 Rhythmical pulse

de Ruiter's prediction that prosodic stress does not affect the timing of gesture not only conflicts with McNeill's (1992) Phonological Synchrony Rule, but also the prediction is contradictory to

work by Tuite (1993) that proposes an interaction of gesture and speech at an unspecified lower level of motor processing in the speech production process. Tuite's (1993) Rhythmical Pulse Hypothesis asserts that gesture and speech are linked prosodically and that gesture and speech originate from a *kinesic base*. Tuite argues that this kinesic 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. Regardless of the type of gesture, the Rhythmical Pulse Hypothesis theorizes that the gestural stroke "tends to coincide with the nuclear syllable of the accompanying tone group" (p. 100; Figure 7). Tuite's Rhythmical Pulse Hypothesis is novel in that he considers the temporal relationship of gesture and speech as related to a motoric interaction of gesture and speech. Even though Tuite does not directly situate his hypothesis within a dynamic systems perspective, the notions of rhythmical pulses and pulse peaks is very much in line with work in the dynamic systems literature (Barbosa, 2001, 2002; Cummins & Port, 1996; Jones & Boltz, 1989; Large & Jones, 1999; Lashley, 1951; O'Dell & Nieminen, 1999; Port, 2003). Tuite's hypothesis is integrated with such work in a subsequent section of this document.

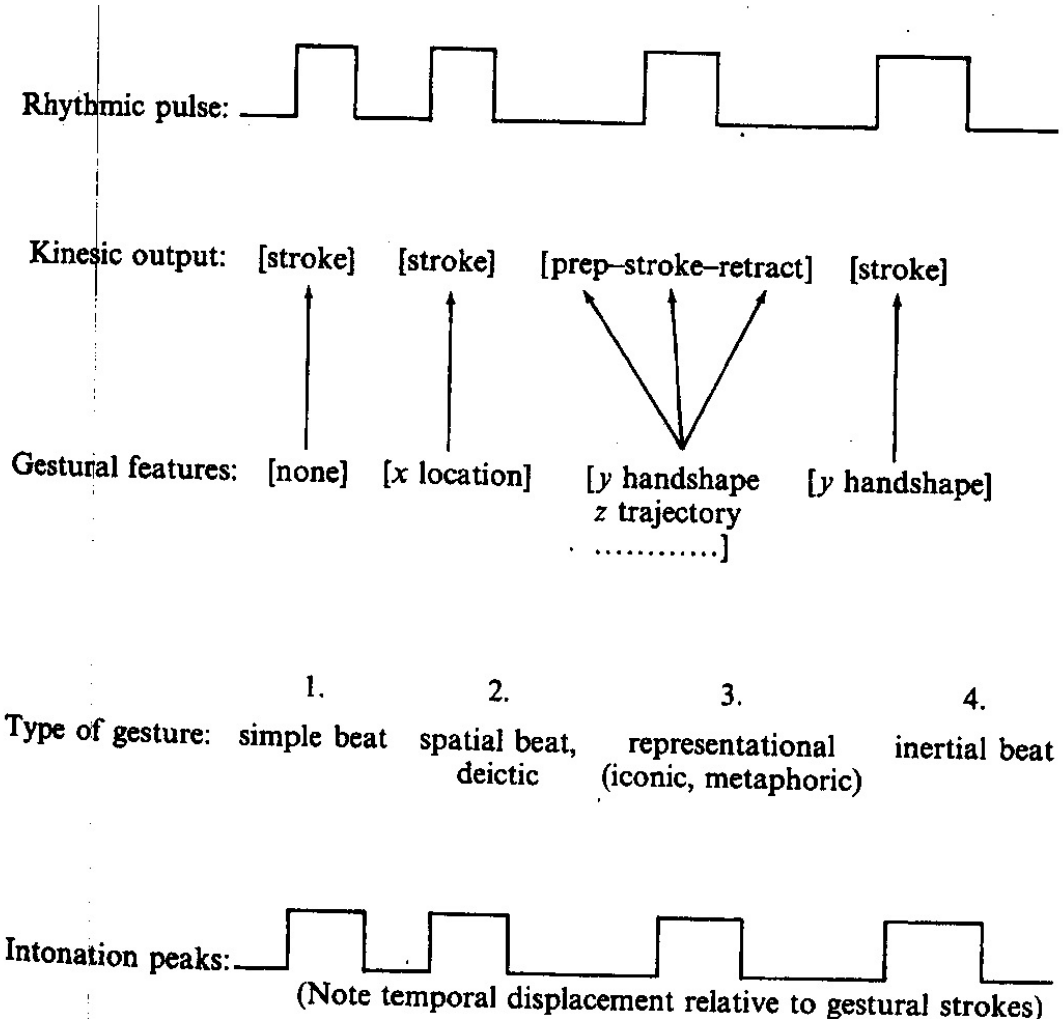


Figure 7 Rhythmic pulse alignment across speech and gestures. Adapted from "The production of gesture" by L. Tuite, 1993, *Semiotica, 93,* p. 99.

The Rhythmical Pulse Hypothesis also shares similarities with McNeill's model. Similar to the Growth Point model's Phonological Synchrony Rule, The Rhythmical Pulse Hypothesis postulates that the gestural stroke coincides with the intonation peak of the associated lexical item and that gestures occur more or less rhythmically in time. However, Tuite's proposal encompasses beat, deictic, and lexical gestures manual gestures as well as non-manual

manifestations of the rhythmical pulse such as head, leg, or foot movements. Although Tuite's hypothesis is similar to McNeill's work in regards to the role of prosody upon gesture/speech synchronization, the Rhythmical Pulse Hypothesis contradicts McNeill's work in regards to the purpose of gesture. According to Tuite, gesture is "production-centered rather than reception-oriented" and "the activity of gesture primarily occurs for the 'benefit' of the speaker/gesturer and not for the listener" (Tuite, 1993, p. 94). Indeed, Tuite's conception is remarkably vague, but nonetheless an interesting origin of inquiry of the temporal relationship of speech and gesture and warrants further expansion, especially if data generated by the current study supports the concept of motor coordination of speech and manual movements during communication. Before discussing the integration of these models with models of speech production. A brief review of two other models of gesture production and their predictions is offered regarding the temporal parameters of gesture in relation to the accompanying speech signal.

2.3.4 Facilitatory

de Ruiter is not the first or only author to utilize Levelt's (1989) model of speech production as a scaffold for a model of gesture production. Robert Krauss and others (1996, 2000) presented a model of gesture production in which the sole purpose of gesture is to assist the speaker in lexical retrieval, rather than to provide communicative information to the recipient. Hence, this model will be referred to as the Facilitatory Model for ease of discussion. Krauss and colleagues (1996, 2000) propose that gestures originate *prior* to the conceptualization stage of speech production by the spatial/dynamic feature selector and are linked to visual images in working

memory. In other words, both gestures and words are retrieved from a spatial and dynamic mental representation.

The Facilitatory Model specifies gesture production as a three-stage process as shown in Figure 8. The three stages are spatial/dynamic feature selection, motor planning, and motor system execution. After a mental representation is activated in memory, the spatial/dynamic feature selector identifies and specifies spatial and dynamic features such as direction, contour, size, shape, speed, etc. The spatiodynamic features are then fed into the motor planner. The description of the motor planning process is extremely vague and simply states that the abstract features are translated into a motor program which provides the commands for motor execution of the gesture. This motor program is then executed and is monitored by the kinesic monitor.

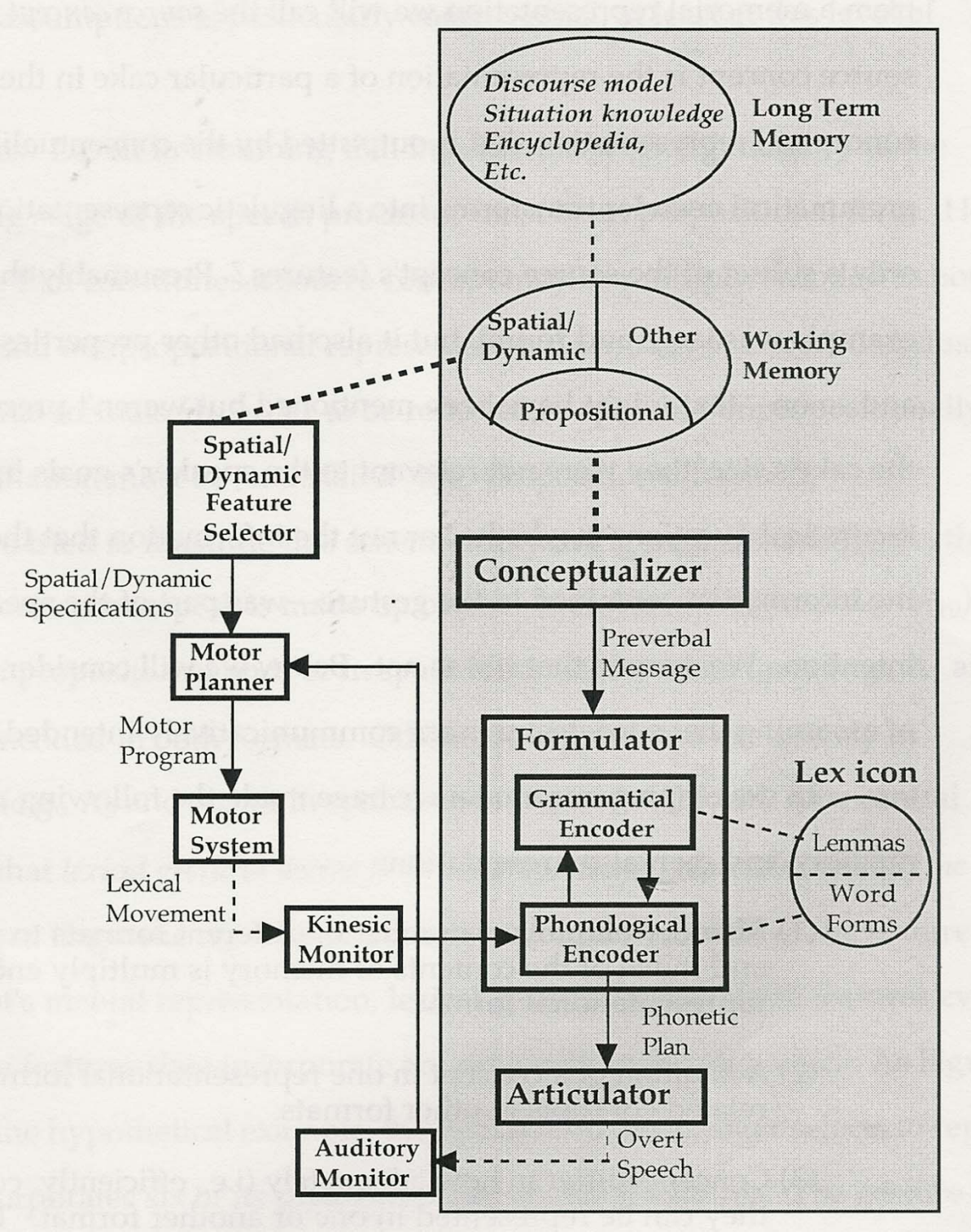


Figure 8. A facilitatory model of the speech-gesture production system. The model is a modification of Levelt's (1989) speech production model. Adapted from "Lexical gestures and lexical access: A process model," by R.M. Krauss, Y. Chen, and R.F. Gottesman, 2000, In Language and Gesture, D. McNeill (Ed.), New York: Cambridge University Press, p. 267.

All of the above stages occur separately but in parallel to the speech formulation and conceptualizing processes. It is not until after the gesture is executed that the kinesic monitor provides input to the phonological encoder, which is embedded in the formulation stage of speech production. The input to the phonological encoder is said to include features of the original representation in motoric form. The features themselves are not necessarily the features of the spoken lexical items since the gesture production began prior to conceptualization. However, in this model the input from the gesture system to the phonological encoder is thought to help in facilitating lexical retrieval because of cross-modal priming. Krauss and colleagues propose that a gesture is terminated when the acoustic signal of the lexical target is heard by the speaker.

Like all gesture production hypotheses, there is little empirical work to support or refute the Facilitatory Model. One piece of evidence that Krauss and colleagues (1996, 2000) cite is the temporal relationship between gesture and speech production. In contrast to McNeill's assertion that production of gesture coincides with the production of the spoken word associate, Krauss and others claim that gestures precede their lexical affiliate. According to Morrel-Samuels & Krauss (1992), lexical gestures precede their lexical affiliate by an average of 0.99 seconds (range of 0 to 3.75 seconds). The argument then follows that the gesture must be initiated prior to the conceptualization stage of speech production in order to be executed prior to execution of the spoken lexical item and allowing for possible lexical retrieval enhancement. This argument also is in opposition to de Ruiter's postulation that gesture originates within the conceptualizer.

As previously stated, the Facilitatory Model is similar to the Sketch Model since it also an extension of Levelt's (1989) model of speech production. Additionally, both the Sketch Model and Facilitatory Model predict that gesture can be beneficial to the speaker. However, there are also many discrepancies between the two models. As already pointed out, the Krauss et al. (1996, 2000) hypothesize that gesture is initiated *prior* to the conceptualization stage versus the Sketch Model that hypothesizes that gesture originates *within* the Conceptualizer. Hence, in de Ruiter's model, the Conceptualizer is responsible for generating the gestural "sketch" as well as the preverbal message that is then sent to the Formulator for grammatical and phonological encoding. Also in contrast to the Sketch Model, gestures are accompanied by lexical affiliates (i.e., a single lexical item) rather than conceptual affiliates. Another difference between the two models is that the Sketch Model accounts for all gesture types with the exception of beats (i.e., iconics, deictics, emblems), however the Facilitatory Model makes predictions only regarding lexical (i.e., iconic) gestures. Most importantly for our purposes, the Facilitatory Model neither accounts for the role of prosody nor perturbation in the temporal synchronization of gesture and speech. Therefore, although the Facilitatory Model is similar to the Sketch Model in its construction it fails to offer relevant predictions for the current experiments. The final theory that is discussed, the Entrained Systems Theory also does not incorporate prosodic stress or perturbation as pertinent variables. Nonetheless, it is briefly summarized for sake of complete review of relevant hypotheses of speech/gesture temporal synchronization. Furthermore, the premise of the theory will be incorporated with a hypothesis temporal entrainment of the speech and gesture systems developed in a later section.

2.3.5 Entrained systems

Iverson and Thelen (1999) proposed the final gesture production theory of this review. Iverson and Thelen propose an entrained system of speech and manual movements that is rooted in dynamic systems theory. For discussion purposes, this theory as the Entrained Systems Theory. Like the four preceding theories, Iverson and Thelen's postulation is reliant upon the assumption that speech and gesture are temporally synchronous and are part of a unified system. In contrast to the Sketch Model, Growth Point Theory, Rhythmical Pulse Hypothesis, and Facilitatory Model which do not consider the development of communication and motor processes in children, the intention of Iverson and Thelen is to explain the co-emergence of vocal and manual movements from infancy to toddlerhood.

Iverson and Thelen state 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 in common communicative intent" (p. 36). Figure 9 conveys the four phases of the developmental progression of the rhythmic entrainment of the speech production system to the manual gesture system. The 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. Conversely, when an infant begins to acquire speech and language, gestures most often precede their spoken affiliates and the two effectors are entrained to a lesser degree. The authors hypothesize that motor control of the hands is relatively

more stable and the preferred mode of communication rather than the vocal apparatus secondary to the novel and effortful process of speech processing and execution at that particular point in development.

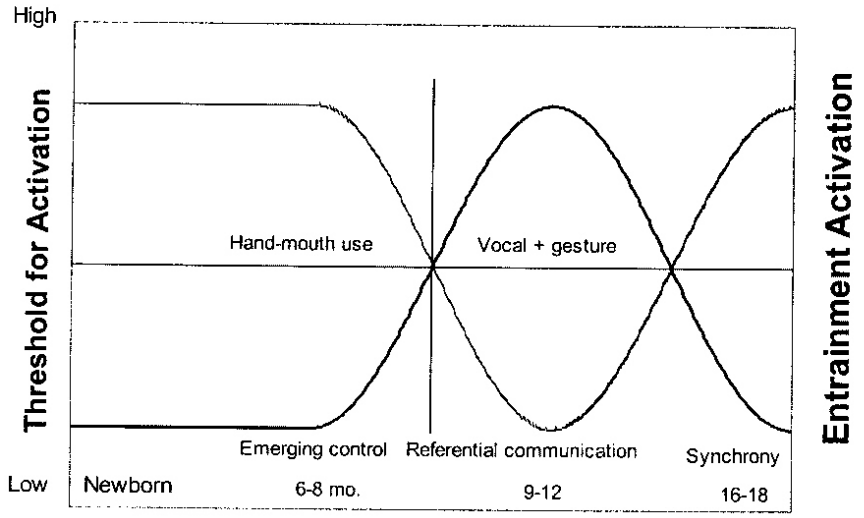


Figure 1. Threshold and entrainment activation levels in the oral–manual system during the first two years.

Figure 9 Four phases of entrainment in the first two years. Adapted from "Hand, mouth and brain: The dynamic emergence of speech and gesture," by J.M. Iverson & E. Thelen, 1999, *Journal of Consciousness Studies, 6,* p. 31.

Iverson and Thelen (1999) propose four early phases of infant development that consist of a loose coupling of hand-to-mouth movements. These initial linkages are thought to reflect co-occurring manual and mouth movements during the first six months of life. At six to eight months of age, an infant begins to gain control of the manual and vocal systems. The authors state this is the time period that vocalization and arm/hand movements are entrained through rhythm. At this age, rhythmic movements of the arms and hands are often accompanied by canonical babbling (Iverson, Hall, Nickel, & Wozniak, 2007). They further state that it is the entrainment of rhythmic manual movements that will later act as the vehicle for mandibular

oscillations to be overlaid with babbling and later assist infants to gain even greater control over the individual systems. The third phase of the progression reflects further developments of motor control and is termed a time of "flexible coupling". This phase occurs prior to an infant's first birthday and is associated with an increase in the independent use of gesture from speech and vice versa. During this time, gestures (i.e., pointing gestures) are used as the predominant communication modality. The Entrained Systems Theory asserts that the threshold for gestures is lower and the activation for gestures is higher than for vocalization due to the infant having greater control of the limbs and hands compared to the oral articulators.

The fourth and final phase of the Entrained Systems Theory occurs around 16 to 18 months of age when speech and gesture begin to converge and synchronize once again. This emergence of synchronous speech and gesture accounts for the presumed tight temporal synchrony of gestures and speech that is so often referred to in the literature. Iverson and Thelen (1999) postulate that as speech becomes more practiced and less effortful, the activation will heighten and the threshold will lower for speech production. The activation then "has the effect of capturing gesture and activating it simultaneously" (p. 35) which then manifests as a synchronous and entrained production of speech and gesture. Balog and Brentari (2008) indeed found evidence for this fourth phase of the Entrained Systems theory. They investigated the co-occurrence of rising and falling intonation contours in the vocalizations of children between the ages of 12 and 23 months of age in a 30-40 minute play exchange with their mothers and examiner. Older children (18 to 23 months) were found to synchronize their intonational patterns and nonverbal body movements more often than younger children (12 to 17 months). The activation and entrainment proposed for toddlers is theorized to continue throughout one's lifespan.

In short, the Entrained Systems Theory relies upon the belief that "the 'stroke' (or active phase) of the gesture is timed precisely with the word or phrase it accompanies" (p. 35). Although the model does not predict the precise temporal relationship of speech and gesture or the precise speech variable that is proposed to coincide with the gestural stroke, the coupling of movements implies a simultaneous activation and production of movement similar to Tuite's Rhythmical Pulse Model and the Growth Point Model. On the other hand, this theory is novel since it is the only to make predictions about the simultaneous development of the speech and gesture production systems. Though, the authors make no explicit statements about what affects the synchronization of speech and gesture produced by older children and adults. Iverson and Thelen also are among the only theorists to postulate a motoric level of interaction of the speech and gesture systems rather than a conceptual, lexical, or phonological level of interaction. Similarly, they are the only theorists to structure their postulations within a dynamic systems framework.

Table 3 *Summary of Gesture Production Theories*

	Citation(s)	Basic Premise	Temporal Synchrony
Sketch	de Ruiter, 1998, 2000	-Gesture originates in the conceptualizer where spatio-temporal information is accessed from working memory and sent as a "sketch" to the gesture planner which then constructs a motor program for subsequent motor execution. -This model is also an extension of the Levelt (1989) model of speech production. -Gestures do not have lexical affiliates, they have "conceptual affiliates" that can span phrases rather than words. -The model includes iconic, deictic, emblem gestures, not beat gestures. -Gesture aids the speaker and the conversational recipient.	-Gestural onset is "roughly" synchronized with the onset of the "conceptual affiliate". -de Ruiter rightly acknowledges the difficult task of reliably determining the temporal boundaries of gesture as well as the relevant lexical affiliate of which to assess their shared temporal parameters. -The preverbal message is sent to the formulator after the gesture planner creates the associated gesture's motor program. -Lexical stress and pitch-accent are clearly stated as having no affect upon the synchronization of speech and gesture. -After the initiation of gesture execution, it cannot be interrupted.
Growth Point	McNeill, 1985, 1987, 1992, 2000, 2005	-Speech and gesture form a shared system which expresses a speaker's mental representations to a listener. -They are tightly linked at all levels of production. -Gestures primary function are to communicate to the conversational recipient.	-The synchronization of speech and gesture is emphasized as evidence for an integrated system. -The stroke is the most important phase of the gesture for synchrony. -The gestural stroke begins with or ends at the phonologic peak syllable. -Provides three rule of synchronization; (1) Phonological Synchrony Rule, (2) Semantic Synchrony Rule, and (3) Pragmatic Synchrony Rule.
Rhythmical Pulse	Tuite, 1993	-Gesture and speech are linked prosodically and originate from a kinesic base which is either a gestural stroke or intonational peak of the spoken phrase. -Gesture can be a manual or nonmanual movement.	-The gestural stroke coincides with the nuclear syllable of the spoken "tone group". -The gestural stroke can also precede the lexical affiliate secondary to the increased processing time required for speech production.
Facilitatory	Krauss, Chen, & Chawla, 1996; Krauss, Chen, & Gottesman, 2000; Krauss & Hadar, 1999	-Gestures facilitate lexical access via cross-modal priming. -Model is an extension of Levelt's (1989) model of speech production. -Input from the gesture system to the phonological encoder is thought to help enhance lexical retrieval because of shared spatiodynamic features that originated in working memory. -The model only includes lexical (iconic) gestures. - Gestures aid the speaker not the conversational recipient.	-The onset of gesture precedes the lexical affiliate. -Gesture must be initiated prior to the conceptualization stage of speech production in order to be executed prior to execution of the spoken lexical item and allowing for possible lexical retrieval facilitation.
Entrained	Iverson & Thelen, 1999	-Gesture and speech originate as a coupled system in the earliest stages of development. -Vocal and motor behaviors are entrained rhythmical movements that progress to a tight temporal synchrony of speech and gesture.	-The gestural stroke is timed precisely with its lexical affiliate. -There is no specific variable proposed the gestural stroke co-occurs with (i.e., stressed syllable; onset of semantic affiliate, etc.) -Children go through periods of vocal-motor coupling that are more or less entrained as a function of effort versus automaticity.

Table 4 *Temporal Synchrony: Predictions of Models and Theories*

	Onset of gesture relative to speech	Prosodic effect
Integrated	-Stroke begins with or ends at the phonologic peak syllable -Preparation movement of the gesture precedes speech	Yes
Facilitatory	-Gestural onset precedes lexical affiliate	Not stated
Sketch	-Gesture *roughly* synchronizes with onset of conceptual affiliate	No
Rhythmical Pulse	-Gestural stroke co-occurs with or precedes the nuclear syllable of a "tone group"	Yes
Entrained	-Gestural stroke timed precisely with affiliate	Not stated

2.3.6 Gesture production theory summary

Thus, contradictory hypotheses regarding the effect of prosody upon the temporal synchrony of speech and gesture abound (Table 3). de Ruiter stands alone among contemporary theorists in his postulation that gestures are *not* affected by prosodic stress. Others have either hypothesized increased synchronization of prosodic stress and gestural stroke (McNeill, 1992; Tuite, 1993), or have not made explicit predictions regarding this potential relationship (Iverson & Thelen, 1999; Krauss et al., 1996, 2000). de Ruiter's hypothesis regarding the cessation of interaction between the speech and gesture production systems below the level of the Conceptualizer yields a predicted null effect of prosodic stress upon gesture timing. In other words, there is no predicted

interaction between the levels of motor processing of the production mechanisms (i.e., phonological encoding, articulation, motor execution, etc.) and the production of gesture.

Specifically, de Ruiter states that because the Formulator and points below in the speech production system do not affect gesture timing and vice versa, then variables such as prosodic stress, lexical access, and the interruption of speech cannot affect the timing of gesture production. Yet, not only do other theories (Krauss et al. 1996; 2000; McNeill, 1992; Tuite, 1993) predict that these variables actually *do* affect the temporal relationship of gesture and speech (Table 4), but there are also existing data that point to such variables affecting the synchronization of speech and gesture (Bull & Connelly, 1985; de Ruiter, 1998; Loehr, 2004; Mayberry & Jaques, 2000; Mayberry, Jaques, & DeDe, 1998; McClave, 1998; McNeill, 1992; Morrel-Samuels & Krauss, 1992; Nobe, 2004).

2.4 INTERACTION OF LINGUISTIC, SPEECH, AND MANUAL PROCESSES

I have provided considerable detail regarding de Ruiter's hypothesis that processes within the Formulator do not affect the timing of gesture. Likewise, the process of assigning prosodic stress and accent to a syllable has been described at length as well as the common prediction that gestures align with prosodically stressed syllables, perhaps due to interaction between the Formulator and the Gesture Planner. Conversely, it is also possible that the interaction between the speech and gesture production systems is actually not at the level of the phonological encoder

but rather at the level of motor programming. Expressly, speech and gesture may synchronize due to some form of entrainment of the speech and manual motor systems rather than because of 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, as the reader recalls from the earlier discussion, hypotheses such as those proposed by McNeill (1992), Tuite (1993), and Iverson and Thelen (1999) posit interactions of the speech and gesture systems below the level of phonological processing. Specifically, it is hypothesized that the speech and gesture systems are temporally entrained.

No doubt, there is inherent semantic and visual-spatial information encoded in not just the speech stream, but also the gesture movement. Therefore, the unified, multimodal production of manual gesture and speech offers a unique opportunity to investigate the amalgamation of linguistic processes with both speech and manual motor behaviors. Nevertheless, the convergence of linguistic, speech, and manual processes is rarely contemplated not only from a gestural perspective but from other viewpoints as well. Historically, each of these entities is studied as independent phenomenon. However, there is increasing interest in and evidence for the interaction of manual processes with speech and language processes from behavioral, neuropsychological, and neurophysiological data. A brief discussion of some of these findings is presented to provide the reader with a comprehensive rationale for this investigation.

2.4.1 Shared neuroanatomical substrates

Theorists of language evolution have proposed the hypothetical progression of fine manual movements to intentional language as well as the hemispheric lateralization of oral and manual movements in our ancestors and modern-day humans as evidence for shared motor processes of the two systems (Arbib, 2005; Corballis, 2002; 2003; 2010; Gentilucci & Dalla Volta, 2007; Lieberman, 1998). Additionally, significant changes of neural architecture during evolution suggest that spoken language may have stemmed from manual gestures. A major location of cortical change is Broca's area (see Gentiluccia & Dalla Volta, 2007 for a review). This area in the left inferior frontal gyrus is typically associated with ventral premotor cortex (area F5) in primates bilaterally (see Figure 10). Extensive investigation has found that F5 is a sight for "mirror neurons" in monkeys. The purpose of these mirror neurons is to activate both when the monkey is doing an action as well as simply viewing an individual performing the same action. Studies have found that F5 only responds for hand and mouth movements, such as grasping, sucking, and facial gestures. In fact, Ferrari and colleagues (2003) have further isolated neurons of F5 to include neuronal populations for ingestive mouth movements separate from communicative mouth movements in addition to the already well established manual mirror neuronal populations. It follows that this area of the cortex may have enabled our ancestors to develop manual and facial movements simultaneously, potentially for a common goal (i.e., communication). It also follows that this area serves an imitative purpose for motor learning and performance, thus allowing one to go from being an observer to a performer (e.g., Studdert-Kennedy, 2002).

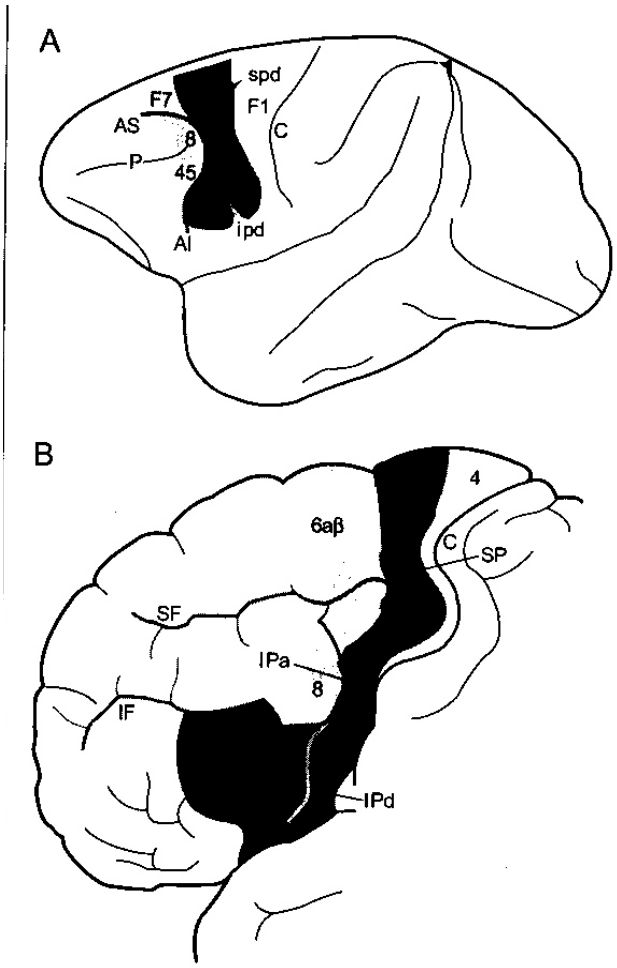


Figure 10 Primate cortical area 5 and human Broca's area. Adapted from "Language within our grasp," by G. Rizzolatti & M.A. Arbib, 1998, *Trends in Neuroscience, 21,* p. 190.

An analog mirror neuron area has been found in Broca's area as well as the superior temporal sulcus and primary motor cortex in modern humans (Avikainen, Forss, & Hari, 2002; Grèzes, Armony, RoI& Passingham, 2003). Not only is there activation in these areas during the observation and production of speech movements, but similar to F5 in monkeys, the area also is active for the observation and production of finger and hand movements (Heiser, Iacoboni, Maeda, Marcus, & Mazziota, 2003; Rizzolatti & Arbib, 1998). Likewise, there is neuroimaging

data that demonstrates excitability of Broca's area during the execution of hand and arm movements (Bonda, Petrides, Frey, & Evans, 1994; Schlaug, Knorr, & Seitz, 1994), as well as activation of the hand motor cortical area of the language dominant hemisphere during reading, but not of the leg motor cortex (Meister, Boroojerdi, Foltys, Sparing, Huber, and Topper, 2003). Lastly, Gentilucci and colleagues have demonstrated that the size of grasped objects affects the size of mouth opening and the amplitude of syllable production (Gentilucci, Benuzzi, Gangitano, et al., 2001; Gentilucci, Santunione, Roy, & Stefanini, 2004). These results were in line with previous data that reflected excitability of the hand motor area during speech tasks (Seyal, Mull, Bhullar, Ahmad, & Gage, 1999; Terai, Ugawa, Enomoto, Furubayashi, Shiio, Machii, et al., 2001; Tokimura, Tokimura, Oliviero, Asakura, & Rothwell, 1999). Meister et al. interpreted this finding as support for "the hypothesis that there are phylogenetically old connections between these two regions which evolved during the evolution of speech" (p. 406). Indeed, there is ample evidence for multiple shared neuroanatomical substrates for the sequential and temporal aspects of speech, language, and manual processes in humans such as lateral perisylvian cortex, supplementary motor cortex, premotor cortex, cerebellum, and Broca's area (e.g., Erhard, Kato, Strupp, et al., 1996; Fried, Katz, McCarthy, et al., 1991; Grabowski, Damasio, & Damasio, 1998; Krams, Rushworth, Deiber, et al., 1998; Ojemann, 1984; Peterson, Fox, Posner, et al., 1989).

2.4.2 Developmental parallels

Others have turned to the shared ontogenetic roots of linguistic, speech, and language tasks and have asserted that the co-emergence of arm/hand movements and speech/language developmental milestones and the expression of similar endogenous rhythmic patterns across systems. For example it is often accepted that rhythmic hand banging emerges at the approximate time of canonical babbling of repeated consonant-vowel [CV] syllables) is evidence of the development of an integrated verbal and manual communication system, at least in early development (Bates & Dick, 2002; Capone & McGregor, 2004; Goldin-Meadow, 1998; Iverson & Fagan, 2004; Iverson, 2010; Iverson, Hall, Nickel, & Wozniak, 2007; Iverson & Thelen, 1999; Kent, 1984; McNeill, 1992; Volterra, Caselli, Capirci, & Pizzuto, 2004). Many have asserted that the co-emergence of arm/hand movements and speech and language milestones is evidence of the development of an integrated verbal and manual communication system (Balog & Brentari, 2008; Bates & Dick, 2002; Capone & McGregor, 2004; Goldin-Meadow, 1998; Iverson & Thelen, 1999; Kent, 1984; McNeill, 1992; Volterra, Caselli, Capirci, & Pizzuto, 2004). The integration of manual and verbal communication has been stated to have possible linguistic roots (Balog & Brentari, 2008; Bate & Dick, 2002; Capone & McGregor, 2004; Goldin-Meadow, 1998; McNeill, 1992) and motor roots (Iverson, Hall, Nickel, & Wozniak, 2007; Iverson & Thelen, 1999; Kent, 1984). Yet, the crucial question is whether the parallel development of gesture and speech, especially in the toddler years and beyond, is coincidental or if the communication system is truly integrated across manual and verbal modalities. Gentilucci and colleagues (Gentilucci & Dalla Volta, 2007; Gentilucci, Stefanni, Roy, & Santunione, 2004) do

suggest that gesture and speech are integrated via a shared mirror neuron system that assists in communication development in children given their findings that relatively greater effects were found in children than adults in their investigation of syllable production while observing large and small fruits being brought to a mouth.

The emerging gesture, linguistic, speech, and manual milestones are organized in Table 5 modified from Bates & Dick (2002, p. 294) and Iverson & Thelen (1999, p. 30). The message that this information sends is that it certainly does seem likely that manual, linguistic, and oral/speech milestones are being reached in a sequential and correlated manner. However, this does not constitute causation or concrete evidence that the systems are integrated. One can only speculate at this time that manual and speech systems are coupled in development and that linguistic units are related to both manual and oral movements.

Table 5 *Gestural, Speech, and Linguistic Milestones*

Age	Gesture	Linguistic	Speech/Oral	Manual
Newborn			Sucking, crying, vegetative sounds	Hand to mouth/reflexive grasping
6-8 months			Cooing, sound play, rhythmic canonical babbling	Reaching, rhythmic waving and rhythmic hand banging
8-10 months (9-14 months for Iverson & Thelen, 1999)	deictic gestural routines	Word comprehension	Variegated babbling, first word use following gesture use, gestures and speech do not share referents, gesture and speech are temporally asynchronous	First gestures, gestures precede words, fine motor finger skills improve
11-13 months	*lexical*-recognitory or gestural names ritualized requests	Word production/naming		
12-18 months (16-18 months for Iverson & Thelen, 1999)	*lexical*-iconic	Increased word learning	Vocabulary is increasing	Fine motor finger control continues to improve

2.4.3 Concomitant deficits

It is well-established that individuals often exhibit concomitant deficits following stroke, traumatic brain injury, and other organic etiologies. However, more subtle concomitant motor deficits have been described for other disordered populations such as children with language

impairment (Bishop, 2002; Bishop & Edmundson, 1987; Corriveau & Goswami, 2009; Gross-Tsur, Manor, Joseph, & Shaleve, 1996; Hill, 2001; Hill, Bishop, & Nimmo-smith, 1998; Johnston, Stark, Mellits, & Tallal, 1981; Powell & Bishop, 1992). Hill completed a review of the specific language impairment literature (2001) and concluded that language impairments and motor impairments do indeed coexist and that there is a shared underlying mechanism for both deficits. Bishop and colleagues (Bishop 1990; 2002; Hill, Bishop, & Nimmo-Smith, 1998; Powell & Bishop, 1992) have also asserted that children with specific language impairment (SLI) frequently exhibit manual motor impairments even in the absence of diagnosed speech or other motor disorders based upon empirical investigations. While it is anticipated that individuals with neurological infarcts are likely to exhibit speech, language, and manual deficits due to the close proximity of their cortical regions, it is not at all obvious why children with language impairments also may exhibit differences of coordination or timing of manual movements.

An extensive review of the literature by Hill (2001) indicated a high rate of comorbidity of SLI and limb motor skill. Hill's motivation in completing the review was to determine whether SLI was specific to language or if these children also exhibit deficits that are seemingly unrelated to linguistic factors. The purpose of the review was to identify co-morbid relationships, not to identify why deficits of language and limb control may coexist.

A more recent investigation of motor deficits and language impairment was completed by Bishop (2002). Bishop and colleagues completed a number of twin studies attempting to determine a genetic etiology for specific language impairment. Subsequent to their investigation of language skills, they began to see a common occurrence of motor deficits with these children compared to their age matched peers. In a series of two experiments, the motor skills of children with language impairments and children with both speech and language impairments were

compared to the motor skills of their peers. It was predicted that children with only speech deficits would have greater motor impairment than children with speech deficits along with language impairment. These children with speech and language impairment were predicted to exhibit significantly more motor impairments than children with only language impairment and children with no history of communication disorders.

All children completed a tapping task to assess motor skill in experiment one. The dependent variable was number of thumb taps in 30 seconds on a tally counter held in the palm of the hand. As predicted, the three disordered groups performed significantly slower than the resolved, unaffected, and control groups. Furthermore, a trend of slower performance was observed with the individuals with a speech deficits, either alone or with co-occurring language impairment to perform more slowly than children with SLI. However, this trend failed to reach significance.

Experiment 2 was completed two to three years following experiment 1. Of the original twin sample, 37 pairs with at least one twin with a history of SLI participated along with a group of twins from the general population (51 MZ pairs, and 49 DZ pairs). The motor task was changed in experiment two to a peg moving paradigm, Annett's Peg-Moving task. The participants moved ten pegs as quickly as possible from the back to front row of a pegboard. There were three trials for each hand, in alternating succession though the performance was again combined across all trials, regardless of hand. The correlation of performance between hands for the new sample was r=0.832 (n=200). The average peg moving rate was calculated in pegs per second to provide a more even distribution of the data for comparison purposes. The speech and language impaired group performed at significantly slower rate than children developing typically. Again there was a nonsignificant trend of the speech impaired children

performing slower than the typical children. The results of experiments one and two demonstrated that children with speech deficits are also likely to be slower than their peers on manual motor tasks.

Even more recently and more relevant for the current project, Corriveau and Goswami (2009) explored a potential underlying mechanism regarding the fairly consistent findings 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 researchers also examined the children's motor dexterity via a peg insertion task but only found a significant group difference in the number of pegs inserted using the non-dominant hand. Further, Corriveau and Goswami conducted multiple regression analyses which revealed that although the performance on the pegboard task was not correlated to language and reading abilities, performance on metronome rhythmic entraining task was "related to all measures of language and literacy" (p. 127). Thus, the authors posit a deficit of auditory rhythmic processing in children with SLI. They further hypothesize that a deficit of auditory rhythmic processing could result in difficulties with the segmentation and subsequent language that occurs based upon prosodic cues.

Corriveau and Goswami's (2009) study does not parse the relative contribution of the perception and production of rhythm, but certainly exemplifies the need for additional study on the entrainment of children and adults with communication disorders. The results of the experiments in this literature base also signal the need for more controlled investigation of motor

task performance of children with not only speech deficits but also language impairments as well. The relevant implication for this line of work is that linguistic and manual movements not only display shared processing in typical development and adult use, but also in the breakdown of language, such as that exhibited by children with language deficits. Thus, the concomitant manual and language deficits are indicative of interactive mechanisms, particularly for temporal processing of language and manual functions.

The present investigation seeks to complement this literature by not only understanding mechanism underlying the potential rhythmic processing deficits of individuals with speech and language disorders understood, but also the potential therapeutic benefit of motor entrainment tasks. This line of thought is in agreement with Corriveau and Goswami's (p. 129) statement:

> Although this idea may appear speculative, it is of note that patients with movement disorders such as Parkinson's disease can be helped by auditory rhythms (e.g., Thaut et al., 1996). If auditory rhythms can be used to rehabilitate motor problems, then there is some plausibility in the reciprocal idea that motor rhythms might be able to help in the development of better auditory rhythmic sensitivity in children with auditory rhythmic sensitivity in children with auditory processing problems and poor language.

There is potential to not only facilitate language skills via improved auditory rhythmic sensitivity but also to affect the speech production processes of children and adults with speech sound production disorders via the interaction and possible entrainment of speech and manual movements (Garcia & Cannito, 1996; Garcia, Cannito, & Dagenais, 2000; Hammer, 2006; Hustad & Garcia, 2005). For instance, Garcia and Cannito (1996) found that intelligibility was enhanced for a single speaker with dysarthria when an individual only listened, not viewed, the participant's speech with gestures compared to when they spoke without gestures. This finding was found specifically for a condition in which the number of beat gestures, those gestures

thought to be most likely to entrain to the rhythm of speech, also increased. The authors postulated that the temporal and prosodic features of the speech of the individual with dysarthria may have been aided by the accompanying beat gestures. Garcia, Dagenais, and Cannito (1998) later examined the acoustic parameters of this speaker's productions and found that the utterances produced with gestures also had relatively shorter interword-intervals implying a possible increase in natural prosodic intervals. However, this hypothesis has yet to be fully explored.

2.4.4 Facilitating effects of manual movements

In somewhat incongruous contrast to impaired manual task performance exhibited by those with language deficits, manual movements may actually facilitate word retrieval and word learning for individuals with language deficits as well as typical adults and children. There is increasing interest regarding the effects of gesture and other manual movements such as finger tapping on the recovery of word retrieval processes of adults with aphasia. For instance, Hanlon, Brown, and Gertsman (1990) found that pointing with the right hand led to improved naming performance for individuals with nonfluent aphasia more so than pointing with the left hand. There were no facilitatory effects of pointing in the confrontation naming tasks for the participants with fluent aphasia. Paskek (1997) not only demonstrated immediate benefits of gesture for word retrieval but also sustained improvement six months post-training for a case study of an individual with severe Broca's aphasia and apraxia of speech. The participant was trained to label sixty pictures that also had a corresponding representational gesture. Pashek's

results showed that naming was better in gesture+speech condition relative to speech only condition. Similar to Hanlon, et al., the long-term facilitatory benefit of the gesture+speech condition held only for the right hand. Additionally, Rose and Douglas (2001) conducted study with six individuals with aphasia. Their research also demonstrated a facilitatory effect for gestures during object naming; however, the effect was only for iconic gesture types, not pointing, cued articulation, or visualization cues. Again, a recent surge of interest on this topic has commenced. The reader is encouraged to review the discussions of the utility of gestures for aphasia treatment by Rose (2006) and Raymer (2007) and these other recent references (de Ruiter, 2006; Feyereisen, 2006; Power & Code, 2006; Raymer, Singletary, Rodriguez, Ciampitti, Heilman, & Rothi, 2006; Richards, Singletary, Rothi, Koehler, & Crosson, 2002; Rose, 2006) for more information.

Despite the recurring finding that gesture helps to facilitate word retrieval for adults with aphasia and the use of manual cueing for children with autism (e.g., McLean & McLean, 1974), the manual modality has rarely been experimentally manipulated as a cue for word learning and recall for children with language impairments. In fact, Ellis Weismer and Hesketh's (1993) study of the role of prosodic and gestural cues on word learning by children with specific language impairment (SLI) remains the only of its kind. Eight typically developing children and eight children with specific language impairment in kindergarten participated in the study. Novel words were trained using an "outerspace creature" named Sam. In the visual cue condition, an iconic gesture was produced simultaneously with the novel word. The novel words were nonsensical (e.g., *pod, gi, wug*) but were meant to convey spatial concepts such as *beside* and *on top of*. To test to the children's retention of the novel words, they were asked, *Where is Sam?*. As expected the children with SLI were more inaccurate then the children with typical

language skills across all conditions, except for the visual cue condition for which they performed the same. Thus, the iconic gestures increased word learning for both children with and without language impairments.

Others have described improved word learning and retrieval performance via gestural cues for typically developing children. The belief that iconic gestures can enhance language development has even prompted the recent and overwhelming interest in the use of "baby signs" with infants (e.g., Acredolo & Goodwyn, 1990). While it is often implied that one of the advantages of teaching signs to hearing infants is to speed and increase language skills of young children, the validity of this claim remains in dispute (see Johnston, Durieux-Smith, & Bloom, 2005). However, the call for controlled, theoretically-based research on this topic, particularly given the public's interest in the utility of signs with infants, has slowly spawned such investigations. Recently, Capone & McGregor (2005) completed a study of novel word learning with and without the presence of iconic gestures by toddler-aged children. They manipulated the use of iconic gestures in learning of labels six novel objects by nineteen toddlers, ages 27 to 30 months. These objects were purchased from a kitchen supply store, though the labels were actually nonsense words. Each iconic gesture either referred to the object's shape or function. As predicted, the children were able to most accurately retrieve the object labels that were trained in the experimental condition. The authors speculate that "perhaps our gestures drew attention to an important aspect of the word learning problem (shape, function, or both), thereby reinforcing salient semantic content of the spoken language" (p. 1478). Later Capone (2007) expanded that the results were consistent with an associationistic account with the lexical-semantic system consisting of a "distributed neural network of auditory, visual, tactile, proprioceptive, olfactory, and/or gustatory features (i.e., information nodes)" (p. 735). While it

appears evident that manual gestures do enhance lexical retrieval for adults with aphasia, children with language impairments, and typically developing children, the mechanism of this facilitatory effect is yet unknown.

Perhaps one of the most intriguing investigations of the facilitatory effect of manual movements on word retrieval is Ravizza's (2003) study of the effect of "meaningless movements" (i.e., finger tapping) upon lexical retrieval with college-aged adults. In short, she demonstrated that word retrieval was performed with greater accuracy in the finger tapping condition in a tip-of-the-tongue (TOT) paradigm. The stimuli were composed of 200 definitions taken from previous TOT paradigms (e.g., *a professional mapmaker-cartographer*; *a bottle designed to be carried in one's pocket-flask*). Experiment 1 lasted one hour and consisted of 70 definitions while Experiment 2 lasted 2 hours at consisted of 165 definitions. Participants were instructed to read the definition and then to type in the lexical item immediately if they knew the answer. If the answer was unknown, the computer would ask if they were in a TOT state which was defined as "a feeling that one 'knows' a word but is not able to articulate it at the present time" (p. 611). If the participant claimed to be in TOT state, they were asked to rate how likely they were to recognize the word on a 5-point scale. Then they were given the definition again for 30 seconds. At this time, dependent upon group assignment, the participants either sat still or were to tap both index fingers on the table at their own pace. Participants in both groups were told to depress two foot pedals to restrict foot movements while the no movement group also were required to depress finger keys as well. If the participant recalled the word during this period, they were instructed to type it in. Results established that lexical access occurred with greater accuracy in the finger tapping condition. Ravizza hypothesizes that "movements somehow boost activation levels of lexical items …that are insufficiently primed when people

are in a TOT state" (p. 612). These findings suggest that it is not just the shared semantic properties of gesture and speech that potentially facilitates lexical retrieval, but that the basic movement of the hand or finger may interact with the speech and language system to improve word learning and word recall.

2.4.5 Dual-task paradigms

Investigators have also studied the interactions between linguistic, speech, and manual tasks by employing concurrent task paradigms to compare the tradeoffs of duration and accuracy for each variable (Chang & Hammond, 1987; Dromey & Bates, 2005; Dromey & Benson, 2003; Hamblin, 2005; Hiscock & Chipuer, 1986; Kinsbourne & Cook, 1971; Kinsbourne & Hiscock, 1983; LaBarba, Bowers, Kingsber, & Freeman, 1987; Peters, 1977; Seth-Smith, Ashton, and McFarland, 1989; Smith, McFarland, & Weber, 1986; Thornton & Peters, 1982). The rationale for conducting a concurrent task (i.e. dual task) paradigm is to examine the allocation of resources between two or more processes given the well-established fact that single tasks are performed faster and with greater accuracy than when performed concurrently with a second task. There is a distribution or allocation of resources across task processes because there is only a limited pool of resources. Additionally, it is hypothesized that the more similar two tasks are to one another, the more interference that results as a result of shared or proximal neuroanatomical structures. According to Dromey and Benson, "if the left hemisphere is occupied with communication, it has been reasoned that performance of the right hand, which it

controls, would deteriorate more than the left in a concurrent speaking and manual task" (p. 1235).

Dromey and colleagues (2003; 2005) have carried out the most recent and the well-controlled studies of the concurrent effects of linguistic, cognitive, and motor tasks. However, these studies are not without caveats. The greatest potential confound is the lack of a manipulation of complexity of task and the inability to assess the contribution and fluctuation of attention to task performance. Still, a brief discussion this work is provided information on the tradeoff of speed and accuracy of speech production during the co-production of a manual task.

The purpose of the first investigation (2003) was to "compare three different types of distractor tasks to evaluate their influence on speech movements…each type of distractor was anticipated to require different processing resources, which might then result in different effects on motor performance" (p. 1235). There was a linguistic and a cognitive task that consisted of noun to verb generation and mental arithmetic, respectively. The motor task required the twenty participants to put three washers on a bolt and then tighten nut on it. A strain gauge system was used to measure lip and jaw kinematics. The participants recited *Mr. Piper and Bobby would probably pick apples* along with a pacing beep spaced every 3 seconds. Each individual recited this sentence 15 times in isolation and simultaneously with each of the concurrent tasks. Results showed that even though the manual motor task was not cognitively demanding, there was a significant difference in lower lip (LL) displacement and velocity in the speech only compared to the speech+motor task. Because there were significant difference of LL displacement and velocity, the authors stated that the processing demands were enough to change the speech performance of the subjects. According to Dromey and Benson, these differences result from undershooting their articulatory targets during the tasks because they

have more than one motor demand. Other studies found the concurrent motor tasks have an effect on the finger movements but not on the speech and attributed this to speech "winning out" in the hierarchy of importance (Smith, McFarland, and Weber, 1986). However, there are other studies that found opposite results in that there were effects for both (Chang & Hammond, 1987; Kelso, Tuller & Harris, 1983).

Dromey and Bates (2005) followed up the 2003 manuscript with a similar investigation of three different types of concurrent tasks (i.e., cognitive, linguistic, and visuomotor tasks) along with a speech task which required the participants to recite *Peter Piper would probably pick apples* 15 times. The general methodology was the same as in the earlier experiment. However, the motor task was changed to a visuomotor task. A moving target was viewed on a computer screen and each of the 20 participants attempted to track and click on the target as often as possible. Again, the motor task interfered with the speech task. Lower lip + jaw displacement and utterance duration were significantly reduced for the combined speech and visuomotor task compared to the speech only task. One explanation given for these two effects was possible temporal entrainment of the speech and manual systems. That is to say, individuals were instructed to track a fast-moving target with a hand movement which then in turn increased the rate of speech and reduced amplitude of articulator movement.

In fact, the finding that temporal rate and variability are highly correlated across different effectors is not unique to this experiment. For instance, as finger tapping rate increases, speech rate increases (Franz, Zelaznik, & Smith, 1992; Kelso, Tuller, & Harris, 1981; Klapp, Porter-Graham, & Hoifjeld, 1991; LaBarba et al., 1987; Smith, McFarland, and Weber, 1986). These data are indicative of synchronization as a result of temporal entrainment across the speech and

manual systems. Let us now turn to a discussion regarding temporal entrainment and dynamic systems as it relates to speech and upper limb motor processing.

2.5 DYNAMIC SYSTEMS THEORY AND TEMPORAL ENTRAINMENT

The topic of temporal entrainment and dynamic systems theory has been alluded to throughout this manuscript. As stated previously, simple top-down modular processing accounts that focus on linguistic processes such as models presented by de Ruiter (1998, 2000) and Levelt (1989) fail to distinguish the mechanism by which gesture synchronizes with prominent syllables. The model of sensorimotor speech motor control presented by van der Merwe (1997) allows for greater specification of the various levels of motor processing, most importantly for our purposes, the level of motor programming. Yet, this model does not provide information regarding the interface of speech with other motor systems at the level of motor programming. In order to flesh out this issue, one must turn to a nonlinear account of motor behavior like concepts encased by dynamic systems theory. Even though the notion of a motor program seems at odds with the tenets of dynamic systems theory, there are indeed similarities, particularly between the motor programs and coordinative structures. Schmidt and Lee (1999) summarize that, "in both the motor program and coordinative structure concepts, the many degrees of freedom in the musculature are reduced by a structure or organization that constrains the limbs to act as a single unit…also, both notions involve the tuning of spinal centers, corrections for errors

in execution, and freedom of the executive level from the details of what occurs at the lower levels in the motor system" (p. 155).

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., Thelen & Smith, 2002) to inanimate properties like biological and chemical processes (e.g., Lorenz & Diederich, Telgmann, & Schutte, 2007). In this section, a brief history and overview of dynamic systems theory is provided and the evidence for coordinated temporal processing is described. Again, the application of nonlinear dynamics is vast. Hence, this section focuses specifically on the relevance of the topic to speech and upper limb movements.

2.5.1 History and Overview of Dynamic Systems Theory

Nonlinear science has roots in Chaos theory in physics and chemistry (see Gleick, 1987) and was later extended to the development of dynamic systems theory. The attraction of these concepts for the explanation of human behavior is based on the fact that many, if not most, human actions and cognitive processes display nonlinearity and considerable variability. Lashley's (1951) groundbreaking manuscript was among the first to describe the nonlinearity of complex biological systems in contrast to the traditional view of examining linear processing. Likewise, Bernstein (1967) advanced the call for an alternative to traditional top-down processing accounts in his description of the *degrees of freedom problem*. Bernstein conjectured that executive

control was not necessary in the planning and execution of movement. Conversely, he stated that groups of muscles acted together as functional synergies to complete a motor goal since there are potentially countless ways to actually attain each motor goal. The functional grouping of muscle groups are described as "*coordinative structures* that are constrained to act as a unit in a motor task (and) behave as limit-cycle oscillators, that is, they have a preferred frequency and amplitude of oscillation, and they return to their preferred state after perturbation" (Smith, McFarland, & Weber, 1986, p. 471). Thus, the major foundations of dynamic systems theory were in place, namely the concepts of nonlinearity, elimination of central control, and coordination of structures.

There are several major points of divergence between a dynamic systems perspective and a traditional view of linear behavior. Perhaps the greatest of these is the eradication of a central command executive. In contrast, the structures involved in a dynamic system are self-organizing. That is, the system has a "natural tendency to perform certain patterns and to switch into patterns that are more efficient under certain parameter conditions…without conscious involvement associated with volitional control" and "suggest that patterns of coordination emerge *spontaneously* from the interaction of the available degrees of freedom of the system" (Schmidt & Lee, 1999, p. 220). Another important component of a dynamic system is stability, or lack thereof. A dynamic system, by definition, is a system that is always in a state of flux. The system has preferred patterns or attractor states, but those states can change under the influence of other parameters, both internal and external to the system.

The behavior of interest is often described according to its position within a *phase space* and a given moment of that behavior is described according to its *relative phase*. 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 (Kelso, 1984), finger-oscillation (Kelso, 1984), finger tapping (Tuller & Kelso, 1989), and pendulum swinging (Schmidt, Shaw, & Turvey, 1993). Kelso's (1984) classic study of finger closure is summarized to provide a quick exemplar (see Schmidt & Lee, 1999 for review). The results of these other cited studies are very similar.

If one opens and then closes the index finger and thumb bilaterally at a rate of one "pinch" every one second, the time of opening and closing of each hand would be performed simultaneously even though there is no instruction to do so. When two effectors are in synchrony, they are said to be *in-phase* and have a relative phase angle of 0˚. An in-phase pattern is the naturally preferred pattern. Conversely, one can also perform the bilateral opening and closing movements in opposition to one another with relative ease. For example, when the left fingers are opening, the right fingers are closing. This pattern is identified as *anti-phase* and the relative phase angle is 180˚. Not only is there a natural tendency for a complex system to be in-phase or anti-phase, but variability of a system can lead to stability without conscious control. Thus, the system is self-organizing. For instance, if one performs the anti-phase finger closing task, but then increases the rate of pinching gradually there will be increased variability (i.e., decrease of stability) as the rate increases then an intriguing switch from anti-phase to in-phase coordination.

It is thought that certain patterns of coordination, especially in-phase coordination, maximize the efficiency of the coordinative structures performing the behavior. Each of the hands is performing a rhythmic oscillatory movement. Therefore, each of these coordinated structures is termed an *oscillator* and the coordinated right and left hands are described as

coupled oscillators. These terms apply not to just this example but to all examples of two or more effectors that perform oscillatory movement patterns and are temporally coordinated. Thus, *time is fundamental to entrainment*.

The oscillators are said to be *entrained* when they influence one other mutually to "produce a single coordinated behavior, synchronous in time and space" (Iverson & Thelen, 1999, p. 28). Similarly, Thelen and Smith (2002, p. 304) iterate "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". Yet another clear definition comes from Clayton, Sager, and Will (2004) who define 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 (2009) use the more descriptive term of "pulse-based rhythmic entrainment" (p. 4) to describe this phenomena. This term will be utilized in later parts of this manuscript particularly because it encompasses the hypotheses regarding the entrainment of speech and manual movements posited but Tuite (1993) and Port (2003).

The idea of entrainment actually was first presented 70 years ago by von Holst (1937, 1939, 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 when they are produced in isolation. Smith et al. (1986, p. 471) further specify that "two oscillations are entrained if they occur at the same frequency or at harmonically related frequencies and have a consistent phase relationship". A few examples of oscillators that can

entrain one another are the) and finger tapping along with metronome beats (e.g., Aschersleben, 2002), repetitive phrase production along with metronome beats (Cummins & Port, 1998) and finger tapping and repetitive syllable production (Kelso, Tuller, Harris, 1981; Smith et al., 1986).

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 are share a pulse-based entrainment (Bluedorn, 2002; Clayton, Sager, and Will, 2004; Merker, Madison, and Eckerdal, 2009). That is, that the pattern of one behavior (e.g., marking beat gestures) is co-occurs at certain points in time with the cycle of another behavior (e.g., pitch accented syllables in the speech stream). Bluedorn provides 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 consisitent relationship with each other. (2002, p. 149)

Entrainment can be external, such that a rhythmic behavior of one species is entrained to that of another. Examples are limitless but include fireflies synchronizing illumination, synchronization applause patterns, and even parrots "dancing" to music (Schachner, Brady, Pepperberg, & Hauser, 2009). Tapping (e.g., Correiveau & Goswami, 2007) or repetitive phrase production along with a metronome is another example of external entrainment (Cummins & Port, 1998). Entrainment can also be internal, 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. Another example is the rhythmic synchronization of movements of the right and left arms (Kugler & Turvey, 1987). As reviewed

in an earlier section, Iverson and Thelen (1999) conjectured that gesture and speech may also be oscillators that can be internally entrained. This leads us to the crux of the current research.

2.5.2 Dynamic systems theory and speech

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 & Byrd, 2000; Saltzman & Munhall, 1989; Tuller, Harris, & Kelso, 1982; Tuller, Kelso, & Harris, 1982, 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 & 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, Kelso, & Bunz, 1985; Saltzman & 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, Jeannerd, MacKenzi, & Marteniuk, 1991; Prablanc, O'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, Turrell, Fleury, Teasdale, Lamarre, & Martin, 1999; Komilis, Pelisson, Prablanc, 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, Richardson, La Heij, 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 & Zimmerman, 1981; Kelso & Tuller), applying a mechanical load to an articulator, most often the mandible or lower lip, either in a consistent or transient manner (Abbs & Gracco, 1982, 1984; Folkins & Abbs, 1975; Gracco & Abbs, 1985, 1986; Kelso & Tuller1984; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Shaiman, 1989; Shaiman & Gracco, 2002), or by removing afferent information from the motor system by using local anesthetics like Xylocaine and nerve blocks (Kelso & Tuller, 1983; see Abbs, Folkins, Sivarajan, 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, Hampson, & Johnson, 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, Freedland, Larson, & Hain, 1998), masking noise (e.g., Ringel & Steer, 1963; Kelso & Tuller, 1983), and the feedback of the auditory signal with the presence of a delay (e.g., Howell & Archer, 1984; Stuart, Kalinowski, Rastatter, & Lynch, 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 & Archer, 1984; Stuart, Kalinowski, Rastatter, & Lynch, 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.

2.5.3 Speech and manual coordination

As discussed earlier, there are abundant behavioral, neuropsychological, and neurophysiological data to suggest that there is indeed an interface between speech and manual movements. Likewise, there is evidence from studies embedded in dynamic systems theory that also supports a shared processing of speech and manual motor patterns, particularly for rhythmic movements. Many have investigated the influence of rhythmicity on the coordination of bimanual or limb movements as previously described. Also, there is growing interest in the effect of external rhythms imposed by a metronome in speech cycling and synchronous speech tasks (Cummins, 2002a, 2002b; Cummins & Port, 1998; Port, Tajima, & Cummins, 1998; Tuller & Kelso, 1990). Such research aims to find the underlying harmonics of the percept of speech rhythm. Although there is a long way to go, the evidence suggests that individuals prefer simple harmonic ratios (*i.e.,* 1:1, 2:1, and 3:1) not just for limb and hand movements, but also for production of prominent prosodic features of speech. This work has prompted others to generate models of rhythmic speech production based on the notion of coupled oscillators (Barbosa, 2001, 2002; Port, 2003; O'Dell & Nieminen, 1999). Gracco (1994, p. 24) summarizes by stating that "any centrally generated rhythmic mechanism for speech and any other flexible behavior must be modifiable by internal and external inputs" and that the "rhythmic mechanism is the foundation for the serial timing of speech movements".

Still others have taken this line of questioning a step further. Since it is proposed that (1) speech movements can be coupled to an external stimulus and (2) that manual movements can be coupled to each other, to external stimuli, and to other limb movements, then it can be further hypothesized that speech and manual production are potentially rhythmically coordinated and entrained to each other. There is limited experimental and theoretical work on this topic, though what exists is intriguing and holds promise for future investigations such as the one at hand.

Perhaps the most interesting aspect of the dynamic systems perspective for the current investigation is the concept of coordination across effectors. The argument is that there are effector-independent representations, particularly for the temporal parameters of motor behaviors that can transfer across different systems. Typically, experiments have examined the transfer of motor learning from one side to the other (e.g., Vangheluwe, Suy, Wenderoth, & Swinnen, 2006) and from arm to leg transfer or vice versa (e.g., Keele, Jennings, Jones, Caulton, & Cohen, 1995; Kelso & Zanone, 2002).

Another line of research that was already briefly addressed focuses on the shared temporal parameters of movements across different motor systems, most often the frequency and amplitude of oscillatory movements. Kelso and Tuller (1984) describe this notion of *stable temporal patterning* such that "the temporal stability often takes the form of a phase constancy among cooperating muscles as a kinematic parameter is systematically changed; we believe that this invariant temporal structure is a fundamental 'signature' of coordinated activity, including, perhaps, the production of speech" (p. R931). The search for a so-called temporal invariant has been arduous. Nevertheless, there is evidence for a systematic change of temporal patterning of one oscillator in relation to the other (Munhall et al. 1985; Franz, Zelaznik, & Smith, 1992;

Kelso, Tuller, & Harris, 1981; Klapp, Porter-Graham, & Hoifjeld, 1991; LaBarba et al., 1987; Ostry et al., 1987; Ostry, Cooke, & Munhall, 1987; Smith, McFarland, and Weber, 1986).

In fact, Kelso, Tuller, & Harris (1981) claimed that even disparate motor acts like speech and finger movement can be organized as a single coupled unit. Based on this earlier work by Kelso et al., Smith and colleagues (1986) 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 either a comfortable rate or at varying rates. Tapping and syllable repetition occurred at a harmonic ratio of 1:1 when participants were permitted to perform the tasks at their preferred rate. Even though there was not absolute coordination when the participants were instructed to alter the rate 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. Moreover, the authors point out that the lack of practice may have hindered the emergence of absolute coordination in the rate-change conditions. Subsequently, Franz, Zelaznik, and Smith (1992) also demonstrated that the temporal patterns of repetitive movements of the mandible and single syllable repetition were significantly correlated to the repetitious productions of finger and forearm movements. Hence, the limited research on interactive temporal patterns suggests that different motor systems, whether bilateral systems, upper and lower limb systems, or even manual and speech systems can indeed be entrained and affect the temporal parameters of the other.

2.5.4 Speech and gesture entrainment

Which brings us back to the issue at hand, are speech and gesture temporally entrained? Based upon the general principles of dynamic systems theory, the wealth of experimental data on the rhythmic coordination of manual, limb, and/or articulator movements, the interface of temporal patterns shared across manual and speech systems, and emerging theoretical thought on the topic, it certainly is worth exploring this question. However, even those that have initiated the study of the shared temporal patterns of speech and manual activity have only examined simple repetitive syllable production and finger tapping. Secondly, individuals have theorized that there are rhythmic commonalities between hand gestures and speech (Cummins & Port, 1996; Wachsmuth, 1999) and that speech and gesture are entrained (Iverson & Thelen, 1999), perhaps according to prominent moments in the speech signal and manual movement (Tuite, 1993). This is the first investigation that systematically explored the hypothetical entrainment of speech and gesture.

The proposal is that speech and gesture are two internal, coupled oscillators. In contrast to speech or finger movements being entrained to the rhythmic pulse of a metronome, they are self-entrained according to rhythmic pulses. 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". As the reader recalls from the review of Iverson and Thelen's (1999) Entrained Systems Theory of gesture production, they hypothesize that speech and gesture become temporally synchronized in early development by way of rhythmical activity of the arms, hands,

and speech apparatus. The rhythmic coordination of early movements results in entrainment of the manual and speech systems. Correspondingly, Tuite (1993) suggests that speech and gesture are synchronized according to their rhythmical pulses. According to this view, a gestural stroke, or in the case of the present experiment a gesture apex, synchronizes with a prosodically prominent syllable because of a shared kinesic base. Hence, it is logical to integrate these hypotheses of gesture production with a dynamic theory of rhythmic speech production. Port (2003) 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). The notion of periodic pulses has also been applied to work with perception and attention (e.g., Large & Jones, 1999) as well. According to Port, 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 (1993) or Port's (2003) 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., Cummins & Port, 1998) 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.-in order to solve a wide range of problems in complex motor coordination, and in the prediction of environmental events" (p. 609).

By integrating the hypotheses put forth by Port (2003), Tuite (1993), and Iverson and Thelen (1999) we can develop a more elaborative hypothesis of speech and gesture entrainment. It is proposed that indeed speech and gesture are temporally synchronous due to temporal entrainment of the two motor systems as Iverson and Thelen hypothesize. 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. Specifically the most salient portion of the accented syllable (i.e., vowel midpoint) entrains the most salient portion of the gesture (i.e., gesture apex), resulting in the perceived temporal synchronization of speech and gesture. It can be further hypothesized that speech acts as the 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 possible that entrainment of speech and gesture is a "two-way street" (Rochet-Capellan, Laboissière, Galvàn, & Schwartz, 2008). In other words, rather than an accented syllable acting as the attractor for a gesture, the gesture may act as the attractor for accompanying speech production, counter to Port's (2003) hypothesis. Nonetheless, this hypothetical source of entrainment or which rhythm is more "powerful" as Bluedorn (2002, p. 149) puts it, was not tested in this investigation. In future investigations, one could manipulate the timing of gesture (e.g. by altering the visual target or physically perturbing hand movement) to examine whether speech entrains to gesture. In this investigation, only the

speech signal was manipulated; hence, only the hypothesis that speech is the attractor for gesture was tested.

2.6 TEMPORAL SYNCHRONY OF SPEECH AND GESTURE

It is evident from our theoretical discussion regarding gesture production that many have placed great worth on the perceived temporal synchronization of speech and gesture. However, do speech and gesture form an integrative system or is it merely coincidental the two movements roughly co-occur during spoken language production? The answer to this question is unclear given the available literature. Even the definition of synchrony is ambiguous in the vast majority of studies with virtually no identification of precise temporal boundaries in any of the investigations. Almost 20 years later, the following statement by Butterworth and Hadar (1989) in reference to McNeill's (1985) early postulations still rings true in regards to both the theory and empirical work on the temporal synchronization of speech and gesture. They summarize, "he treated the synchrony between a gesture and a section of speech as if it were completely transparent, and his own examples consist of a portion of transcribed speech with a description of the accompanying gesture underneath it…no temporal parameters are specified…the most that can be implied is that there is some temporal overlap in the production of the speech and the gesture" (p. 170).

To be sure, the entire experimental topic of gesture production and its relationship to speech and language processes is just beginning and it is difficult to gather reliable and valid information from the majority of existing data, especially regarding the temporal synchronization

of speech and gesture. Though it is also true that the hypothesized synchronization of speech and gesture is a vital tenet of gesture production models as well as relevant for understanding linguistic and speech processes and the potential interface of the oral and manual motor systems. Thus, synchronization of speech and gesture warrants careful reflection within an empirical paradigm. It is argued that this has not been achieved to date, though existing research does provide a number of thought-provoking findings and methodologies to build upon. Three other hypotheses are presented regarding the temporal relationship of gesture and speech before reviewing the literature that examines the hypotheses that gesture remains synchronized with speech in spite of perturbation and that gesture coincides with syllables with prosodic stress.

2.6.1 Gestures precede and/or synchronize with speech

The most readily predicted and supported finding regarding the temporal synchrony of speech and gesture is that gestures precede or fully synchronize with their lexical affiliates (Bernardis & Gentilucci, 2006; Blake, Myszczyszyn, Jokel, & Bebiroglu, 2008; Butterworth & Beattie, 1978; Chui, 2005; Feyereisen, 1983; Krauss, et al. 1996, 2000; McNeill, 1992; Morrel-Samuels & Krauss, 1992; Ragsdale & Silvia, 1982). That is, no study found that gestures are initiated after their lexical affiliates. However, it is still not clear whether gestures co-occur with the onset of speech or if they commence prior to the speech signal. One of the main reasons for this ambiguity is that inconsistent measurement points have been used for calculating the time interval between the manual and speech movements. For instance, some have measured gestural

onset to speech onset (Bernardis & Gentilucci, 2006; Morrel-Samuels & Krauss, 1992), while others have chosen to measure the interval between the gestural stroke and a point in the speech signal (Chui, 2005; McNeill, 1992; Nobe, 1996). Most recently, Blake et al. (2008) examined whether five to ten year old children would precede and/or synchronize with the onset of a noun phrase during narrative production, regardless of whether they exhibited specific-language impairment or were developing language typically. Indeed, gestures, especially iconic gestures, for both groups either simultaneously co-occurred with the lexical "concept" or directly preceded the spoken production of that lexical "concept". A caveat to this investigation is that the determination of the timing relationship between speech and the lexical affiliate was made via visual and auditory inspection, with only 79% reliability from 18% of the sample.

The project at hand does not examine whether gestures precede or synchronize with the lexical affiliate in general, but how the degree of synchrony between gestures and speech may be manipulated. Nevertheless, the experiments' methodologies will lend insight into whether gestures always occur before or simultaneously with the lexical affiliate. Both the onset and apex of the gesture will be temporally measured, hence allowing for a comparison of these disparate measurement points in the existing literature.

2.6.2 Lexical familiarity

A second prediction regarding speech-gesture synchrony is that the degree of synchrony increases with decreasing lexical familiarity. This prediction stems from Krauss and colleagues'

Facilitatory Model (1996, 2000). Morrel-Samuels & Krauss (1992) completed one of the only studies that examined this prediction. Sixty gesture-lexical affiliate combinations selected from a picture description task were rated by 17 participants as familiar or unfamiliar on a seven-point scale. According to Morrel-Samuels & Krauss (1992), lexical gestures precede their lexical affiliate by an average of 0.99 seconds (range of 0 to 3.75 seconds). The authors also found that the time between gesture onset and voice onset time of the lexical affiliate increased as the familiarity decreased. Morrel-Samuels and Krauss used these data to argue that the gesture must be initiated prior to the conceptualizing stage of speech production in order to be executed prior to execution of the spoken lexical item and allowing for possible lexical retrieval enhancement. However, a caveat of this study was that the difference in synchrony time could be solely attributed to the robust finding that lexical access requires less time for frequent compared to infrequent words. Thus, the planning and production of a gesture could proceed without any interaction with the speech system and the word frequency effect alters the time between gesture production and lexical affiliate production.

A more recent study (Bernardis & Gentilucci, 2006) employed a different approach to the question by examining the effect of meaningful speech and gestures versus pseudowords and gestures upon the temporal synchronization of the co-expression of the speech and gesture in a more controlled paradigm with Italian speakers. Each participant viewed a screen and one of three words (*no, stop, ciao*) and one pseudoword (*lao*) was presented quasi-randomly. One of five responses was required: gesture only, verbal production only, simultaneous production of gesture and speech, verbal production of the word with a meaningless gesture, or gesture of the word with verbal production of the pseudoword. Using kinematic measures of the manual movements and acoustic measures of the speech signal, results indicated that the temporal

relationship of the onset of meaningful gesture and the onset of meaningful speech was significantly different depending on whether meaningful speech and gesture were simultaneously produced or if a pseudoword or pseudogesture was produced. The interval between gesture onset and speech onset (Experiment 1, 547 ms and Experiment 2, 518 ms) significantly differed from the same interval for pseudowords and meaningful gestures (Experiment 2, 693 ms), and words and meaningless gestures (Experiment 1, 375 ms). Although these two investigations (see Table 6) point to a potential effect of ease of lexical access on the temporal production of gesture, additional studies are required that manipulate lexical frequency and minimize working memory differences across conditions. Again, this hypothesis will not be studied in the current project. Though, though lexical frequency is another variable that could be manipulated in future research to test de Ruiter's prediction that processing within the Formulator does not affect the timing of gesture

Table 6 *Temporal Synchrony: Lexical Frequency Findings*

	n	Participant Description	Stimuli	Task	Gesture Types Included
Morrel-Samuels & Krauss (1992)	*n*=17	-9 men; 8 women -Undergraduates	-13 pictures of landscapes, abstract art, buildings, people, and machines -221 narratives with 2,328 hand movements	Picture description	-"Speech-related" -Gestures were hand movements that moved at least 1.5 inches in 0.8 sec
	n=129	77 men; 52 women in groups of 10	-2,328 movements yielded agreed upon 197 lexical affiliates -Final stimulus set was 60 hand movements	-Identified lexical affiliates of the 2,328 hand movements -8 of the 10 people in a group needed to agree on the affiliate	See above
	n=17	9 men; 8 women	-60 lexical affiliates -28/60 were multi-word	Rated lexical familiarity of the 60 lexical affiliates on a 7-point scale	See above
Bernardis & Gentilucci (In press)	*n*=28 14 in Exp. 1 14 in Exp. 2	Right-handed, Italian-speaking 21-24 year olds	-3 real words (no, stop, ciao) -1pseudoword (lao)	-Response to computer generated task -4 blocks with 15 trials each, quasi- counterbalanced and randomized - After each trial, the participant was required to respond in one of four ways as stated in the Condition column.	Emblem gestures and 1 "pseudo"-gesture

Table 6 (continued). *Temporal Synchrony: Lexical Frequency Findings*

	Independent Variables	Dependent Variables	Interval Time Points	Results Summarized
Morrel-Samuels & Krauss (1992)	Familiarity of lexical affiliate	Gesture-speech synchrony	-Gestural onset -Voice onset of lexical affiliate (measured from "slow-motion" analysis of recording	-Gestural onset always precede the onset of the lexical affiliate -Range of synchrony = 0-3.8 sec (*M*=0.99 sec; *SD*=0.83 sec) -The longer the gesture, the greater the synchrony -The more familiar the lexical affiliate, the tighter synchrony between gesture and speech are according to a multiple regression analysis
Bernardis & Gentilucci (In press)	Condition -Accurate gesture -Spoken response of the word -Simultaneous production of: accurate gesture and accurate word -meaningless gesture and accurate word -accurate gesture and pseudoword	Gesture-speech synchrony	-Gestural onset -Speech onset	-Gestural onset always preceded the lexical affiliate even for the pseudoword and pseudogesture - The interval between word onset and meaningless gestures significantly decreased (546.8 vs. 375.2 ms) compared to the interval between pseudowords and meaningful gestures significantly increased (517.8 vs. 693.1 ms) when comparing to the interval between word and meaningful gesture

2.6.3 Development and disorders

One can expect that if little is known about the precise temporal relationship of speech and gesture production in adults then even less is known about the same hypotheses in childhood and in disordered populations. Indeed, this is true. Though virtually no data is available for the synchronization of speech and gesture produced by individuals with communication disorders, there is some evidence to suggest that synchrony increases during development (see Table 7).

The emergence of speech-gesture synchrony is thought to occur around the time of two-word speech. Butcher (Butcher & Goldin-Meadow, 2000) was the first to examine the timeline of the synchronous production of speech and gesture in her dissertation. A total of six children were initially observed during play between the ages of 12 and 21 months and then observed another 5 to 11 times. Speech and gesture became synchronous (i.e., the video frame which held the gestural stroke's apex also included a vocalization) around the time that children were also first observed to produce two-word speech. This finding was replicated by McEachern & Haynes (2004) with more controlled time intervals between observation sessions and more controlled observation procedures than employed by Butcher and Goldin-Meadow's earlier study (2000). Balog and Brentari's (2008) more recent investigations of the relationship of nonverbal body movements and intonation patterns in the vocalizations of 12-23 month old children demonstrated that body movements and speech show signs of synchronization, even before the age of two, particularly for falling intonation contours.

To date, there has been no investigation of the degree of gesture-speech synchrony for almost all disordered populations (e.g., children and adults with language disorders, childhood apraxia of speech, aphasia, Down's syndrome, autism, dysarthria, etc.). The only exception is Mayberry and colleagues' observational work with children and adults who stutter (Mayberry & Jaques, 2000; Mayberry, Jaques, & DeDe, 1998), which will be reviewed in the next section on perturbation. Yet, a number of studies do suggest a number of theoretically, diagnostically, and therapeutically intriguing differences between the production of gestures for many of these populations and those produced by typical children and adults (Attwood, Frith, & Hermelin, 1988; Blake, Myszczyszyn, Jokel, & Bebiroglu, 2008; Caselli, Vicari, Longobardi, et al., 1998; Corriveau & Goswami, 2009; Garcia & Cannito, 1996; Garcia & Dagenais, 1998; Hanlon, Brown, & Gerstman, 1990; Hill, 2001; Hill, Bishop, & Nimmo-smith, 1998; Hustad & Garcia, 2005; Iverson, Longobardi, & Caselli, 2003; Osterling & Dawson, 1994; Pashek, 1997; Raymer, Singletary, Rodriguez, et al., 2006; Rose & Douglas, 2001; Smith, Mirenda, & Zaidman-Zait, 2007; Stefanini, Caselli, & Volterra, 2007; Stone, Ousley, Yoder, et al., 1997; Thal, O'Hanlon, Clemmons, Fralin, 1999; Thal & Tobias, 1992). The temporal relationship between gesture and speech during development and for individuals with speech, language, and other disorders is a premature focus for the present project. Still, one prospective goal of this program of research is to cultivate a long-term research plan that extends the methodology and findings of these experiments to children developing speech and language as well as children and adults with speech and language disorders.

Table 7 *Temporal Synchrony: Developmental Findings*

	n	Participant Description	Stimuli	Task	Gesture Types Included
Butcher (1995) Butcher & Goldin-Meadow (2000)	*n*=6	-3 boys; 3 girls -Initially tested between the ages of 12 & 21 months -Final testing session completed between the ages of 19 and 27.5 months	-Spontaneous play interaction with examiner and/or caregiver (one hour) -The first half hour of the interaction was used as the experimental conversation unless there were less than 100 communicative behaviors in which case the coder continued until the 100 communicative behavior mark. -The length of interaction was 30 to 48 minutes	-The children were observed at variable intervals. -Four of the children were observed approximately every 2 weeks and the other two were tested approximately every 6 to 8 weeks.	-Hand movements were classified as a *gesture* according to 4 criteria. -First, the movement had to be directed toward another individual. -Second, the gesture could not be a self-adaptor and the movement had to be empty-handed. -Third, the movement was excluded it was part of a game (e.g., patty-cake) or ritual act. -Finally, imitated gestures were excluded.
MacEachern & Haynes (2004)	*n*=10	-4 boys; 6 girls -Typically developing children determined via parent report, MacArthur Communicative Development Inventory, Denver Developmental Screening Test, and hearing screening -Initially tested between the ages of 15 and 17 months -Final testing session completed at approximately 21 months of age	-Spontaneous play interaction (one hour) -Children were seen every 30 days (+/- 7 days) for six months. -Thus, there was a total of six sessions for each child	-A one hour spontaneous play sample was videorecorded while the infant interacted with his/her caregiver in a room in the Auburn University Speech and Hearing Clinic. -The caregiver and child were provided with age appropriate toys and several "stations" of toys to elicit different types of play activities such as gross motor play, snacking, reading, etc.	-The children's communicative behaviors were coded as in the Butcher dissertation and Butcher and Goldin-Meadow investigation as gesture alone, speech (verbalization or vocalization) alone, or gesture and speech (See above).

Table 7 (continued) *Temporal Synchrony: Developmental Findings*

	Procedures	Results Summarized
Butcher (1995) Butcher & Goldin-Meadow (2000)	-The interaction was transcribed and then coded for gestures, speech, and gesture/speech combinations. -Gesture and speech were both considered communicative behaviors. -*Vocalizations* were classified as either meaningful or meaningless. Meaningful vocalizations were not necessarily real English words, they were also consistent speech sounds that were used idiosyncratically in reference to an object or event. Meaningless vocalizations were communicative but not used as a consistent reference to an object or event. -*Gestures* were classified according to handshape, movement type, a space in which they were produced. -*Gesture-Speech* relationships were also coded. Gestures were either coded as produced in isolation or with speech. If a gesture was coded as co-occurring with speech, it was designated as with occurring with a meaningless or meaningful vocalization. -The temporal relationship of gesture and speech was examined by measuring the time between the gesture and the speech it accompanied to the nearest video frame with was 1/30 of a second. The point of the gesture that was measured as the referent point was the stroke or peak (furthest movement of the gesture before retraction of the hand). -Speech and gesture were considered synchronous if this point of the gesture was accompanied by the vocalization. -Synchrony was measured between gesture and both meaningful and meaningless vocalizations.	-5/6 participants produced asynchronous gesture-speech combinations at the initial session. -The gesture-speech combinations of the sixth child were synchronous at all sessions. -Temporal synchrony does not seem to manifest until children began to produce two-word speech. -Speech and gesture are not unified in early development. -5/6 children also produced a higher proportion of gestures without speech relative to what is expected for adults. -These children produced between 60-97% of their gestures in isolation compared to adults who produce only 10% of gestures in isolation. -Early on in the one-word speech period, children do not produce gesture+meaningful speech combinations.
MacEachern & Haynes (2004)	-Gestures, vocalizations, and verbalizations were considered communicative behaviors -Gesture-speech combinations in which the gestural stroke or the peak of gesture extension occurred after the vocalization/verbalization were classified as asynchronous -Five dependent variables at each of the 6 test session ages (16-21 months) were analyzed separated with ANOVAs. -These variables were mean number synchronized vocalizations, synchronized verbalizations, single word+gesture complementary combinations, single word+gesture supplemental combinations, and two-word utterances.	-As one would expect, synchronized vocalization and gestures decreased with age as verbalization and gesture synchronization increased. -The children demonstrated vocal-gesture synchrony at the earliest ages tested (15-17 months) leading the authors to conjecture that synchronization of vocalization and gesture begins prior to these ages

2.6.4 Prosodic Prominence

The next hypothesis, that gestures coincide with prominent syllables, is relevant for the present series of experiments. Recall that this hypothesis is pivotal for theories such as McNeill's Growth Point theory and his phonological synchrony rule (1992) but rejected outright by others (de Ruiter, 1998, 2000). 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; 1970; Kendon, 1972; 1980) and that the body moves closely in time with speech (Condon & Ogston, 1966; Dittman & Llewellyn, 1969). Thus, the idea that 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" (McNeill, 1992, p. 15). However, few experimental investigations of the relationship of gestural movements and prosodic features of speech have been conducted to date and nothing is known about the mechanism of this potential temporal relationship.

Yet, in comparison to our first three hypotheses, the hypothesis that gestures coordinate with prominent syllables has actually been the subject of the greatest amount of empirical scrutiny. Even so, there are many limitations in this work. For example, there are inconsistent findings between these studies, though even the investigations that share the affirmation that syllable prominence affects the timing of gesture are difficult to equate given widely divergent stimuli, methodologies, segmentation of speech and

gesture, and temporal boundaries measures. Consequently, the majority of the researchers were more interested in identifying a general co-occurrence of gesture and prominent syllables rather than measuring the degree of synchrony between specific points within the speech signal and the gesture movement. Likewise, the objective of most of these studies was to observe, not manipulate, speech-gesture synchrony.

Birdwhistell (1952; 1970) was the first to cite a relationship between body movement and language, particularly intonation. Birdwhistell proposed that the relationship between body movement and intonation is one in which kinesic stress and linguistic stress coincide. According to Birdwhistell, one can track synchrony of small movements of the arms, hands, face, and so on, referred to as *kines,* and the units of speech. However, Birdwhistell did not make any specific predictions about this relationship. Kendon (1972; 1980) also proposed a relationship between gesture and intonation. Gestural strokes are proposed to coincide or slightly precede stressed syllables. In addition, he observed a tendency for gestural phrase boundaries to coincide with tone groups (i.e., prosodic phrases, phonemic clauses, intonation groups, tone units, intonational phrase). Individuals such as Birdwhistell, Kendon, McNeill, and Tuite hypothesized that prosodic stress and manual gestures synchronize, but only offered cursory anecdotes or observational accounts as support for their assumptions. Empirical study of these hypotheses does exist, though they have yielded conflicting and equivocal findings and are far from optimal in their methods and design.

To date, there have been nine experiments that examined of the effect of syllable prominence upon the timing of gesture (Bull & Connelly, 1985; de Ruiter, 1998, Experiments 1 & 2; Loehr, 2004; 2007; McClave, 1994, 1998; Nobe, 1996, Experiment

4; Rochet-Capellan, Laboissière, Galvàn, & Schwartz, 2008; Yasinnik, Renwick, & Shattuck-Hufnagel, 2004). Each of these studies is summarized in Tables 8, 9, 10, and 11. It is difficult to reconcile the findings from this body of work given the fact that disparate methodologies, classification schemes, and temporal measures were employed. For instance, the type of gestures varied. Bull and Connelly (1985) offered the first account of the relationship of stress and body movement, though included not just arm and hand movements, but movements of the trunk, head, legs, and feet. McClave (1994) examined solely beat gestures, while de Ruiter (1998) and Rochet-Capellan and colleagues (2008) examined only deictic gestures. Nobe (1996) and McClave (1998) coded all representational gestures, Loehr (2004; 2007) coded all gestures including head and eyebrow movements, and Yasinnik et al. (2004) coded gestures according to their movement (i.e., discrete and continuous) rather than their semantic relationship to the lexical affiliate.

Not only do these studies differ in the type of gestures examined, but they also differ in the way in which syllable prominence was classified. Certainly, quantitatively measuring syllable prominence has proven difficult for many because there are different types of prominence (i.e., lexical stress, emphatic stress, nuclear stress, pitch-accent) and both perceptual (i.e., loudness, pitch), acoustic (i.e., duration, intensity, fundamental frequency, pitch changes), and physiologic correlates (e.g., kinematic) of stress as previously described. Therefore, the selection of an independent variable of syllable prominence should be well-motivated and reliably identified.

Table 8 *Speech and Gesture Elicitation and Classification Procedures: Prosody and the Temporal Synchrony of Speech and Gesture in Spontaneous Speech*

Citation	Temporal synchrony as a function of prosody?	Spontaneous	Controlled Speech and/or gestures	Gesture Types	Identification of Prosodic Prominence	Temporal boundaries
Bull & Connelly (1985)	Yes	Yes	-		"Tonic stress"	*Imprecise* All body parts included, no specific boundary markings
McClave (1994)	No	+ Two conversations	-	Beats	Stressed syllables and tone units	*Imprecise* Downpoint of gesture (lowest spatial position) No specific point of measurement within a syllable
Nobe (1996, Experiment 4)	Yes	+ Cartoon narration	-	Representational (iconics, deictics, metaphorics)	Intonation units and corresponding fundamental frequency and intensity peaks	*Imprecise* Collapsed both gestural onsets and strokes in most analyses No details about the measurement point within the peak syllable However, this was one of only two studies to employ acoustical analyses of prosody.
McClave (1998)	Yes	+ Two conversations	-	Representational Beats included in only some analyses	Tone unit nucleus (last major pitch change within a tone unit)	*Imprecise* Gestural preparation and stroke No details about the measurement point within the nucleus syllable
Yasinnik, Renwick & Shattuck-Hufnagel (2004)	Yes	+ Three lectures	-	Discrete and continuous	Pitch-accented syllables using ToBI system	*Imprecise* Co-occurrence of a "hit " (a stop or pause of manual movement) and pitch-accented syllables

Table 9 *Summary of Methodology and Results: Prosody and Temporal Synchrony of Speech and Gesture in Spontaneous Speech*

Citation	Stimuli	Participants
Bull & Connelly (1985)	Dyadic conversation lasting 15-20 minutes	*n*=10 British students
McClave (1994)	Two dyadic one-hour *conversations* about no particular topic.	*n*=4 (two females; two males) no additional information regarding participant demographics
Nobe (1996, Experiment 4)	*Cartoon narration*: Sylvester and Tweety cartoons After watching each of 12 episodes, the participant immediately recalled the events of the cartoon to a naïve listener	*n*=6 (five females; one male) taken from McNeill's existing data native English speakers graduate students
McClave (1998)	Task and participants were identical to McClave's earlier study (1994) Two dyadic one-hour *conversations* about no particular topic	*n*=4 (two females; two males; undergraduates; individuals in conversational groups knew each other well)
Yasinnik, Renwick & Shattuck-Hufnagel (2004)	Portions of a lecture were selected for each participant. The durations of the segments were 5, 7.5, and 1.83 minutes.	*n*=2 American English speaking males and 1 Australian English speaking male

Table 9 (continued) *Summary of Methodology and Results: Prosody and the Temporal Synchrony of Speech and Gesture in Spontaneous Speech*

Citation	Gesture and Speech Segment Classification
Bull & Connelly (1985)	-Gestures were defined as any body movement that emphasized speech -Movements were coded as related to "tonic stress" -No criteria provided for identification of stress
McClave (1994)	-A portion of the conversation that was "densely packed with gesticulations" (p. 47) was analyzed. -"Over" 50 beats were analyzed every 1/30 sec by advancing individual frames of the videorecording. -The onset of a syllable was matched with the presence of absence of manual movement in a given frame. -Tone units were identified according to six criteria outlined by Kendon (1972). -The nucleus of each tone unit was identified as the most prominent syllable. -The description is unclear regarding the distinction between unstressed and stressed syllables. -The downward movement of the beat was also analyzed as a relevant point of movement in relation to stressed vs. unstressed syllables.
Nobe (1996, Experment 4)	-Only a small portion of the representational gestures were "randomly" chosen for analysis (*n*=48). -No information provided regarding how these gestures and accompanying speech were chosen and what the total sample of gestures were. -Furthermore, no information was provided regarding the identification and classification of gestures. -Intonation units were selected based upon being segmented by silent pauses, under one intonation contour, and a single semantic unit. There was no reliability for this identification process. -Peak fundamental frequency and peak intensity were identified within the intonation groups using SoundScope software. No information is provided regarding the analysis process or reliability of the identification process.
McClave (1998)	-As in the earlier study, a portion of the videotapes that were "densely packed with gestures" was chosen for analysis, though other sections were included if deemed "noteworthy" (p. 74) -"Over" 125 gestures from only 3 of the 4 speakers were analyzed frame-by-frame from the videotape -The fourth speaker was excluded because he produced few gestures -Gestures were classified according to McNeill's classification systems and then collapsed as propositional (i.e., representational) gestures or beat gestures. -The preparation, stroke, hold, and retraction for each gesture was identified. -The identification of tone units was the same as the earlier study (see above) -In order to analyze pitch changes, CSL was used to analyze the narrow-band spectrograms and measure the fundamental frequency of syllables in the tone units. -The Tone and Break Indexes (ToBI) was also utilized to identify the pitch contour of the tone group. -The variability of pitch was considered relative to an individual's frequency range. Fro three participants of shift of up to 5 Hz was stable, while for the fourth a shift of up to 8 Hz was stable. -For each syllable, except for the first, the direction of F0 change and direction of gesture was identified.
Yasinnik, Renwick & Shattuck-Hufnagel (2004)	-Gestures were segmented using iMovie software. Using frame-by-frame advancement, gestures were classified as *discrete* (i.e., with an abrupt stop of movement) and *continuous* (i.e., repetitive movements without "hits"). A discrete gesture included hits (i.e., abrupt stop or pause). -Gesture onsets and offsets were marked. The stroke or apex was not measured. -The prosodic structure, including pitch-accents and intonational phrases, of the speech segments was coded using ToBI. -The ToBI files were aligned with the acoustic wave files.

Table 9 (continued) *Summary of Methodology and Results: Prosody and the Temporal Synchrony of Speech and Gesture in Spontaneous Speech*

Citation	Results summarized
Bull & Connelly (1985)	-emphatic movements of the body most often performed by arms, hands, and head -90.5% of "tonic stresses" were accompanied by body movement
McClave (1994)	-No quantitative or statistical results are provided. -All results are merely descriptive examples chosen from the sample. -These examples demonstrated that beats coincided with both stressed and unstressed syllables.
Nobe (1996, Experiment 4)	-77% (37/48) gestures "occurred in" intonation units with coinciding fundamental peak frequency and intensity. -36/37 of these consisted of strokes that occurred prior to the acoustical peaks, the other stroke coincided with the peaks. However, it is also noted that the stroke may have started earlier than the peaks. Thus, the temporal boundaries remain imprecise despite the implementation of acoustic analyses. -The remaining 11 gestures (23%) "occurred in" intonation units with non-coinciding peak F0 and intensity. These gesture "onset/stroke preceded and/or co-existed with, but did not start after, the last primary peak of either F0 or intensity" (p. 55). -6 of these 11 gestures had strokes that "co-existed with" the peak F0, compared to only three that coincided with peak intensity. Thus, it appears that the gesture may be more likely to synchronize with peak F0 than peak intensity. -From these findings, Nobe presents an acoustical peak synchrony rule, which closely parallels McNeill's phonological peak synchrony rule, though with acoustic data instead of perceptual trends.
McClave (1998)	-Only correlation statistics for three of the subjects. The statistical analyses were completed for each participant individually. -McClave does not report what proportion of the "over" 125 gestures were propositional versus beat types. -Only 1/3 participants demonstrated a trend of parallel movement of pitch and hand (e.g., falling hand movement with corresponding falling pitch movement). -Further analyses were biased in that they only used data from the participant who exhibited the tendency to have parallel speech and gesture movements. -Nonetheless, the trend of parallel pitch and hand movements was only upheld for propositional gestures, not beat gestures. -For this one participant, the gestural stroke was the best correlate with pitch changes. -"Stress and strokes of propositional gestures tended to converge for all three subjects the majority of the time" (p.84) -Propositional gesture's stroke coincided with the tone unit nucleus 53% of the time. -Propositional gesture's stroke coincided with another stressed syllable of the tone unit an additional 25% of the time.
Yasinnik, Renwick & Shattuck-Hufnagel (2004)	-Each speaker was analyzed individually. -Speaker 1: 90% (158 of 206) hits occurred on a pitch-accented syllable. Labeling was done while listening to the speech signal). -Speaker 2: 90% (117 of 130) hits occurred on a pitch-accented syllable for multisyllabic words. 65% (75 of 116) hits occurred on a pitch-accented syllable for monosyllabic words. Labeling was done in silence. -Speaker 3: irrelevant, only measured pause times of the intonational phrases.

Table 10 *Speech and Gesture Elicitation and Classification Procedures: Prosody and the Temporal Synchrony of Speech and Gesture in Controlled Speech Production*

Citation	Temporal synchrony as a function of prosody?	Spontaneous	Controlled nd/or	Gesture Types	Identification of Prosodic Prominence	Temporal boundaries
de Ruiter (1998) Experiment 1	No	-	+ Picture (array of 4) Gesture also :d	Deictic	Lexically stressed syllables within bi- and trisyllabic words	*Acceptable precision* Gestural apex (point when movement is at its max extension) and pointing initiation (point when hand velocity was more than 1% for the trial) Speech measurement points were beginning and end of each article, noun, and stressed syllable
de Ruiter (1998) Experiment 2	Yes	-	+ Picture Gesture :d	Deictic	Contrastive stressed syllables within mono- and trisyllabic color adjectives and trisyllabic nouns	*Acceptable precision* Gesture Speech Synchrony=beginning of gestural apex – beginning of utterance Other speech measurement points included adjective onset, adjective offset, noun onset, noun offset, and stressed syllable onsets and offsets for adjectives and nouns
Loehr (2004; 2007)	Yes	+ Six conversations	-	Iconics, deictics, metaphorics, emblems, beats, and head movements	Pitch-accented syllables using ToBI system	*Imprecise* Gestural apex (the final video frame of the gestural stroke), also preparations, holds, and retractions Synchrony: two events (i.e., pitch-accent and gestural apex) occurring within 275 msec
Rochet-Capellan et al (2008)	Yes	-	+ Nonword	Deictic	Apex of jaw-opening	*Acceptable precision* Measures between initiation of point, apex of point, jaw initiation and jaw apex for the first and second syllable of the nonword

Table 11 *Summary of Methodology and Results: Prosody and the Temporal Synchrony of Speech and Gesture in Controlled Speech Production*

Citation	Stimuli	Participants
de Ruiter (1998) Experiment 1	The participants were instructed to place their hand on a fixed point on a table. They were seated facing a plate of plexiglass which held four LEDs and four different drawings projected at a time above the LEDs. There was also a fixation LED. The participants looked first at the fixation point, heard a warning tone for 500 ms, followed by a 1000 ms pause, then one of the four LEDs below the four black and white drawing would flash for 1000 ms. The participants were required to state "the X" in reference to the picture name while also pointing to the picture. There were a total of 16 pictures. There were 4 trochees, 4 iambs, 4 trisyllabic words with first syllable stress, and 4 trisyllabic words with final syllable stress. All words consisted of single morphemes. The onset of each stressed syllable was either a plosive or sibilant. There were a total of 96 trials, with picture location and presentation carefully controlled.	*n*=12 right-handed, native Dutch speakers (9 females; 3 males), participants reported normal vision Data was only analyzed for 8 of the 12 participants due to equipment failure (*n*=1), lifting of finger instead of whole hand and finger (*n*=2), and totally separate movements of the hand and speech (*n*=1)
de Ruiter (1998) Experiment 2	The experimental procedures were generally the same as in Experiment 1. The pictures were also the same, however, they were presented either in an array of 4 identical objects of different colors, or 4 different objects all of the same color. Participants were asked to "emphasize" what distinguished a presented picture while simultaneously pointing to the picture.	*n*=11 right-handed, native Dutch speakers (9 females; 2 males), participants reported normal vision Data was analyzed for only 8 of the 11 speakers due to not lifting the entire hand (*n*=1), failing to respond within time allotted (*n*=1), and "lack of time" (*n*=1)
Loehr (2004; 2007)	The participants were asked to have a natural conversation on any topic for one hour. Four clips of conversation were chosen for analysis for a total of 164 seconds with a total of 147 gestures. The conversational sample was not randomly chosen.	N=15 English speakers, separated into 6 conversational groups. The participants in each were friends.
Rochet-Capellan et al (2008)	Labeling of nonwords *papa* and *tata* while simultaneously pointing to the target written word. Stress assignment was indicated by a 'following the syllable to be stressed.	n=20 young adult Brazilian Portuguese speakers (4 men, 16 women)

Table 11 (continued) S*ummary of Methodology and Results: Prosody and the Temporal Synchrony of Speech and Gesture in Controlled Speech Production*

Citation	Gesture and Speech Segment Classification
de Ruiter (1998) Experiment 1	-The speech was recorded using a DAT recorder while the manual movement was monitored with a Zebris ™ CMS motion analyzer. The motion analyzer consisted of an ultrasound buzzer placed to the index finger knuckle of the right hand. The ultrasound signal is then picked up by the Zebris system. The ultrasound system has a 5 ms resolution and a spatial resolution of 1mm. Post-data collection, the speech signal and motion data were synchronized. -Gesture variables were onset of pointing and apex (i.e., point of maximal forward extension) of pointing. -Speech variables were article onset, noun onset, stressed syllable onset, stressed syllable offset, and noun offset.
de Ruiter (1998) Experiment 2	-Speech variables were article utterance onset, adjective onset, adjective offset, noun onset, noun offset. Stressed syllable onsets and offset for adjectives and nouns were also measured. -Gesture variables were onset of pointing, apex onset, apex offset. The apex was divided into two measurement points because individuals demonstrated a hold of the apex in this experiment. -A total of 4% of the trials were in error and excluded from analyses.
Loehr (2004; 2007)	-Gestures were segmented frame-by-frame (each frame=33 ms) using Anvil software (Kipp, 2001) and were classified using a modified version of McNeill's classification system. -Gestural apex was marked as the stroke's final frame or when the stroke's direction changed. -The speech signal was segmented into pitch-accents, phrase accents, and boundary tones using ToBI. A corresponding pitch track was created using the Praat tool (Boersma, 2001). -Pitch-accents were classified as "the point of highest (or lowest) pitch within the vowel of the associated stressed syllable" (p. 78). Interrater reliability for pitch-accent coding was 91% for all sound files.
Rochet-Capellan et al (2008)	-Gestures and speech variables were segmented using an Optotrack system. -10% of peak velocity=onset for both gesture and speech movement -Gesture variables: onset, forward stroke (onset to apex), pointing plateau (apex hold), and return stroke -Speech variables: onset and apex of jaw movement for first and second syllables of nonwords.

Table 11 (continued) *Summary of Methodology and Results: Prosody and the Temporal Synchrony of Speech and Gesture in Controlled Speech Production*

Citation	Results summarized
de Ruiter (1998)	-Gesture apex is temporally close to noun onset on average (M=59 msec prior to noun onset). -A regression analysis was conducted to predict the apex times using speech "landmarks, article onset, noun onset, stressed syllable onset, stressed syllable offset, and noun offset. -Noun onset was the best predictor of apex time ($p<.001$), article onset was also significant ($p<.001$). -Stressed syllable onset and offset were not significant predictors of apex time. -In addition, the apex of the gesture actually occurred earlier in the stress-final conditions relative to the stress-initial conditions which is the opposite direction anticipated. The difference in apex time was not significant. -Therefore, the findings of Experiment 1 demonstrated that stress position did not affect the temporal parameters of the gesture apex. -However, because of the limited intonational contour of the article+noun stimuli construction, a second Experiment utilizing contrastive stress was completed to further examine whether syllable prominence affects the timing of gesture.
de Ruiter (1998) Experiment 2	-Stress location was a significant variable for several analyses. -There was a main effect for stress location the amount of time between onset of pointing and apex onset (termed "Launch" period). The Launch is shorter for syllable initial stress than all other stress locations for adjectives only. -There was a main effect for stress location and the beginning of the gestural apex. The correlation between apex beginning and each stress location was significant. There was a main effect for stress location and apex duration. -There was no main effect of stress location upon the timing of beginning of pointing
Loehr (2004; 2007)	-Gestural apexes were significantly more likely to align with pitch-accented than non-pitch-accented syllables using a chi-square test ($p<.001$).
Rochet-Capellan et al (2008)	-Synchronization was observed for stressed syllables. -Synchrony of the gesture apex and jaw opening apex for stressed syllables in first syllable position. -Synchrony of onset of the gesture return stroke and jaw opening apex for stressed syllables in the second syllable position. -stress in speech always occurred sometime between the onset and offset of the gesture apex.

Of the nine investigations of interest, seven different measures of syllable prominence were utilized. Perceptual correlates of prominence were used by five of the investigations. McClave (1994, 1998) selected stressed syllables and tone unit nuclei, although a clear description of the identification process and reliability was not reported. Loehr (2004; 2007) and Yasinnik et al. (2004) used the Tones and Break Indices (ToBI; Beckman & Elam, 1997) to initially identify pitch-accented syllables.

In addition to perceptually rating the speech samples according to ToBI guidelines, Loehr also utilized acoustic waveforms and corresponding spectrograms and to mark accent as the point of highest or lowest pitch within the perceived stressed syllable. Likewise, Nobe (1996) measured prominence acoustically. Nobe identified the peak fundamental frequency and peak intensity within an intonation unit (i.e., bounded by silent pause, under one intonation contour, one semantic unit) using an oscilloscope trace in a sample of continuous speech.

Only two other sets of investigators employed a controlled paradigm to attempt to control the prosodic characteristics of the spoken productions. de Ruiter (1998) carefully controlled the task to reliably identify the syllables with lexical stress for his first experiment and contrastive stress for his second experiment, though identification of stressed and unstressed syllables was based solely on perceptual judgment. Rochet-Capellan (2008) and others also completed a tightly constrained task to control for the placement of stress on two bisyllabic nonwords and the simultaneous production of deictic gestures. To date, this is the only investigation to utilize kinematic measures of jaw movement as the correlate of prominence.

The selection of the appropriate measure of stress or prominence becomes even more difficult for spontaneous speech samples such as those analyzed by the lion's share of these researchers. Determining the syllables that are stressed and where the larger unit, such as an intonational phrase begins and ends is less than straightforward in spontaneous speech. However, the ambiguity for identifying prominence in spontaneous speech lies in the syntactic construction of lexical items in phrases, sentences, and conversation. Once placed within a larger intonational phrase unit, emphasis or accent may be placed upon any given lexical item. Therefore, it is not certain that a lexically stressed syllable is also the most prominent syllable within a phrase. However, it is often difficult to deduce syllable prominence within spontaneous connected speech and subsequently difficult to select the relevant syllable for temporal synchrony measures. Despite this inherent limitation in identifying prosodic prominence in spontaneous speech, until fairly recently, the majority of investigators chose to use spontaneous samples for their analyses.

2.6.4.1 Spontaneous speech paradigms First, the experiments that employed a spontaneous speech paradigm will be reviewed. These investigations are often vague in their description of methodology and are imprecise in their measurement of gesture-speech synchrony. Moreover, the sample sizes were small ranging from three to fifteen participants, as were the number of gestures and amount of speech analyzed.

Bull and Connelly (1985) were the first to study the concurrence of emphatic stress and body movement. They found that syllables with primary stress produced in

spontaneous conversations were likely to be accompanied by movements of the arms/hands as well as the trunk, legs/feet, and head. They examined the movement and speech of 20 British students, separated in opposite sex pairs, 10 of whom were friends, the other 10 were strangers. The participants discussed three items that they disagreed upon as determined by an "attitude questionnaire" for 15 to 20 minutes. These conversations were videotaped.

These same subjects were used in the gesture identification process. Each participant was instructed to watch the videotape with no audio playback and identify when a body movement occurred as an emphasis to speech. The participants identified emphasis body movements for themselves and their conversational partners. The coding system consisted of identifying which part of the body performed the movement, what the video frame numbers were associated with the movement, and whether the movement was "very" or "quite" emphatic. Both raters needed to agree upon whether a movement was used for emphasis in order to be included. Arm/hand emphasis movements were further classified as unilateral/bilateral and whether or not they came into contact with another object or body part.

A subset of the conversations (six pairs, three familiar and three unfamiliar) were examined to investigate body emphasis movements and their relation to *tonic stress*. The definition of tonic stress (i.e., primary stress) and how it was identified was not reported. Transcripts were completed of the conversations and "scored for primary stress" (p. 179) with interrater reliability of stress coding calculated as 79% accurate.

Bull and Connelly's (1985) results indicated that emphasis body movements were most often performed by the arms/hands and the head and that most emphasis arm/hand

movements were unilateral and did not come into contact with an object or other body part (72%). Bull and Connelly (1985) state that "a mean 90.5% of the tonic stresses within the segments of tape scored were accompanied by body movement" (p. 179) of a total of 277 tonic stresses coded. A total 540 movements were coded as related to tonic stress and were distributed across movements of the arm/hand (34%), trunk (15%), head (35%), and legs/feet (16%).

McClave (1994) was the first to examine the effect of stress upon the timing of manual gestures. She remains the only individual to examine this relationship with beat gestures specifically. This is also one of the only investigations that failed to find that gestures coincided with prominent syllables. Two conversational groups, one consisting of two females and the other with two males, were instructed to talk about any topic for one hour. There was no other information provided regarding the four participants. Additionally, the selection of conversation to be analyzed was not necessarily representative of the entire sample given that an indeterminate number ("over 50") of gestures were chosen from a portion of the conversations that was "densely packed with gesticulations" (p. 47).

The gestures and speech were then classified from this sub-sample. Each gesture was observed by advancing the video recording frame-by-frame, 1/30 second at a time. The speech signal was segmented by first identifying tone units. Tone units were classified according to pause boundaries, pitch movements, presence of anacrusis (i.e. syllables that are unstressed and spoken faster at the beginning of a tone unit), final syllable lengthening, and register changes. Virtually no information is given regarding how the syllables within the tone unit were classified as stressed or as a nucleus of the

tone unit. The points within the gesture and speech signal that was measured in terms of synchrony to the other were not reported.

McClave (1994) does not provide any quantitative data within her results. There is a spattering of qualitative examples, but no clear measurement data or statistical analyses. She summarizes the findings by stating that beat gestures co-occurred with both stressed and unstressed syllables. However, it is not at all apparent that this conclusion can be reached based upon the many caveats and complete lack of quantitative data.

McClave (1998) later conducted a re-analysis of the above data and analyzed both representational gestures and beat gestures. More gestures were chosen for analysis, but again the exact number was not disclosed. "Over 125" gestures were chosen again based upon being in a "densely packed" portion of the sample or from a section that was "noteworthy" (p. 74), thus introducing selection bias in the replication study as well. In fact, sample was further biased given that one of the four speakers was excluded secondary to not producing many gestures, though the criterion for exclusion was not stated. In contrast to her earlier study, beat gestures and representational gestures were identified according to McNeill's gesture classification system (1992). However, the number of each gesture type analyzed was not reported. In addition, the preparation, stroke, hold, and retraction of each gesture were coded. The identification of tone units was the same as in the earlier study though ToBI (Beckman & Elman, 1997) and analysis of a narrow-band spectrogram were also used to help determine the pitch contour of a tone unit and subsequent pitch shifts within the tone unit. A greater than 5 to 8 Hz pitch

variation was considered a significant shift based upon an individual's own pitch fluctuations.

Some statistical analyses were included in this later study, though they were correlational. Not only did the analyses consist of only correlations, but they were conducted for each of the three subjects individually. What's more, analyses were further reduced to only the single subject that demonstrated a trend of parallel hand and pitch movement. Hence, the validity of the findings is questionable. For this single conversational speaker, the gestural stroke was found to be the best correlate to pitch shifts for representational gestures. The gestural stroke coincided with a syllable with some degree of stress 78% of the time, 53% with a tone unit nucleus and 25% with another stressed syllable. This analysis was not completed for beat gestures.

Two other more recent studies made use of ToBI to segment prominent syllables within a larger intonation unit. The first is just as vague as McClave (1994, 1998) in the description of methodology and results. Yasinnik et al. (2004) analyzed the speech and gestures produced by three lecturers. A caveat of subject selection, besides there only be three speakers, was that two of the individuals spoke American English while the third "appeared to" speak Australian English (p. C-98). The content and degree of spontaneity (i.e., spontaneous versus reading) was not reported. Video recordings of the lectures were transferred digitally to a Macintosh computer and iMovie was used to analyze the video segments frame-by-frame. Each frame was 33 ms in duration and a total of 14.33 minutes of speech was analyzed. Pitch-accented syllables were coded within the digitized sound files. As previously stated, the classification of gesture types was distinct for this experiment. The type of movement was the determinant of gesture classification

such that if a manual movement was punctuated by an abrupt stopping point or change in direction, the gesture was classified as *discrete*. If a manual movement was more repetitive and did not have any abrupt pauses or stopping points then it was classified as *continuous*. The on- and offset of each gesture were coded as well. Synchronization was defined as the co-occurrence of a *hit* (i.e., a stop or pause of manual movement) and a pitch-accented syllable.

Because slightly different methods were used for each of the three participants, their data were analyzed separately, similar to McClave (1998). Of the three speakers analyzed, only analyses and results for the first two speakers are relevant for the current discussion. For Speaker 1, 158 of 206 (90%) hits occurred in the same video frame as a pitch-accented syllable. For Speaker 2, 117 of 130 (90%) hits co-occurred with a pitch-accented syllable for multisyllabic words. Fewer hits coincided with pitch-accented syllables for monosyllabic words (65%, 75 of 116). The primary difference between the analyses for the two speakers is that the coder was able to listen to the speech signal when identifying gestures for Speaker 1 but not for Speaker 2. The high percentages for co-occurrence are especially striking due to the coding of a hit in contrast to a gestural onset, stroke, or apex. Any given gesture could consist of multiple hits but only one onset or apex. Therefore, even though there are potentially more hits than apexes, they still coincide with pitch-accented syllable the vast majority of the time. Yet again, this experiment is sorely lacking in numerous controls and has many caveats which limit the external validity of these intriguing findings. Similarly to McClave's investigations, there was no measurement of the degree of synchrony between speech and gesture, just

the perceived overlap of a prominent syllable with some point within the manual movement.

Loehr (2004; 2007) is one of only two researcher teams (also Rochet-Capellan et al., 2008) to consider what degree of temporal distance between a point in the gesture and the speech signal reaches synchronization for not just prosodic prominence but within all speech-gesture synchrony literature. Loehr's dissertation (2004) and later published manuscript (2007) consisted of a variety of questions, one of which was "do the unit boundaries of each modality align" (p.71). Loehr was interested in investigating whether a specific level or levels of prosodic assignment (i.e., pitch-accent) again coded according to ToBI guidelines, were likely to synchronize with the gestural apex. In his study, the gestural apex was defined as the video frame which held either the gestural stroke's final movement or a change of direction of the stroke.

All manual and head movements were coded for 15 participants who were separated into conversational groups of two or three speakers, yielding a total of 6 one-hour conversations. The participants, all friends, were encouraged to discuss any topic. Only 164 seconds of the six hours of conversation, from a total of four clips was chosen by author with a resulting 147 gestures. The distribution of gesture types (e.g., representational, beat, head movements) was not reported. Perceptual judgments of stressed syllables within the conversational speech segments were made and pitch accented syllables were judged perceptually and by identification of highest or lowest pitch "within the vowel of the associated stressed syllable" (p. 187).

Loehr did not measure the degree of synchrony between gestural apex and pitch-accent, but was more systematic in his classification of synchrony that most other

researchers. He pondered "how close in time must two annotations be to be considered 'near' each other" (p. 99). Based upon averages within his own data, he calculated that a pitch-accented syllable and gestural stroke that occurred within 275 ms of each other were synchronous, which was approximately equal to one standard deviation of the average time interval between the beginning or ending of a gestural apex and the nearest tone. Even though this measurement is better motivated than simply observing whether a given video frame hold a manual movement and a prominent syllable, it is not clear why both the beginning and end point of the apex were collapsed and what point within each tone was coded. Furthermore, 275 ms is at least if not more than the average syllable length (Rusiewicz, 2003). Therefore, it is not clear what gestural strokes are actually aligning with in the speech signal. After coding whether or not a gestural apex occurred within 275 ms of any particular point within the boundary tones, it was found that gestural apexes were significantly more likely to align with pitch-accented syllables that syllables that were not pitch-accented according to a chi-square test (p<001). Interrater reliability was also completed for coding of pitch-accented syllables and was acceptable (91%).

This investigation offers insight on the general co-occurrence of head and body movements in conversation, though the potentially selection process of analyzed speech units was potentially biased and certainly limited. It is interesting to note that the tempo of hand movements, head movements, and speech seemed to share a common tempo of a third of a second. However, the measures are do not yield the quantitative data to strongly support this somewhat anecdotal statement empirically. Additionally, the experiment is limited by the video frame resolution of 33 milliseconds for the precise

identification of the time of gesture apex. The caveat of video time resolution is shared by a number of relevant studies, including the study conducted by Nobe (1996).

Nobe completed a series of five experiments in his 1996 dissertation. Only Experiment 4 is relevant for the current discussion. Nobe was also the only investigator to use the ever-popular cartoon narration task as stimuli. In fact, the narratives for this investigation were initially collected by David McNeill and were based upon the Sylvester and Tweety cartoons that were often used by McNeill (1992) and others. A total of 12 episodes were shown to 6 individuals (five females, one male) who were native English speaking graduate students. After watching each episode, each individual immediately recalled the events of the cartoon to a naïve listener. A total of 48 representational gestures were chosen "randomly" from the 72 descriptions, however Nobe does not provide information regarding how these gestures were chosen, nor what proportion of the sample these 48 gestures represented. Additionally, the process of identifying and classifying the gestures and the reliability of this process was not reported.

A prominent syllable was defined as a syllable with peak f0 and/or peak intensity within an intonation unit. Therefore, it was possible for a syllable to possess both peak f0 and intensity, or just one of the two. These measures were made using SoundScope software, but there were no details of the manner in which the peak f0 and intensity were measured or how reliable these measures were. Furthermore, the methodology for identifying gestural onsets and strokes was not described.

Much like the majority of the other studies described (Loehr, 2004; McClave, 1998; Yasinnik et al., 2004) there was a tendency for gestures to coincide with prominent

syllables. Thirty-six (75%) of the gestural strokes "co-existed with" with a syllable with both peak f0 and peak intensity (p. 53). Another six gestures (12.5%) co-existed with a syllable with peak f0 and three gestures (6.3%) co-existed with a syllable with peak intensity. Thus, 95.8% of gestural strokes coincided with a prominent syllable and gesture was slightly more likely to occur with syllables with peak f0 than peak intensity. Based upon these data, Nobe posited a rule of acoustic peak synchrony, which parallels McNeill's phonological peak synchrony rule but refines it to acoustical rather than perceptual results. The investigation does offer a novel and useful approach for classifying prosodic prominence, though is lacking in a number of other areas such as sample size, gesture selection and classification, and reliability. Another drawback of the experiment was that the results are purely descriptive with no accompanying statistical analyses. Also, it is not clear if there was a particular point within the peak syllable and within the gestural stroke that was the focus for synchronization classification. Like all other studies to date, there was no measure of the amount of synchrony (i.e., temporal interval) between speech and gesture. Finally, this study like the other four outlined thus far, utilized spontaneous speech samples which are difficult to control in regards to both speech and gestures produced. Subsequently, the last set of experiments that are discussed were completed using controlled paradigms that are quite distinct from those that used spontaneous samples.

2.6.4.2 Controlled paradigms Even though the studies described above offer valuable insight onto the perceived temporal relationship of speech and gesture in

naturalistic speaking situations, it is not possible to state that there is a predictable and consistent synchronization of gestures and prosodically prominent syllables from these findings. However, there are three investigations that indeed attempted to elicit speech and gesture utilizing a controlled paradigm (de Ruiter, 1998; Krahmer & Swerts, 2007; Rochet-Capellan et al., 2008). As will be exemplified, the collective findings that prosodically prominent syllables and gesture have a tendency to temporally co-occur both in spontaneous speech and controlled speech production, at least in simplistic, constrained stimuli in non-English productions, further support the relevance of and need for the current project.

Not only did de Ruiter construct and describe the Sketch Model in his 1998 dissertation, but he also completed a series of three experiments to test a number of predictions made by the Sketch Model. The first two of these experiments examined the effect of stress upon the synchronization of a gestural apex and it lexical affiliate in a constrained task. Lexical stress within bi- and trisyllabic nouns was manipulated in Experiment 1, while lexical and contrastive stress in mono- and trisyllabic adjectives and nouns was manipulated in Experiment 2. Likewise, de Ruiter was the only researcher to control for gesture type and occurrence. de Ruiter was also the only investigator to measure a temporal interval of synchrony in contrast to a general co-occurrence of some point in the speech and gesture production. Moreover, de Ruiter completed temporal measurements for a number of dependent variables within the speech and gesture productions which allowed for greater specificity of synchronization points. Indeed, these experiments are far less naturalistic than those that employed spontaneous speech tasks. However, these experiments provide essential information on the potential effect

of prosodic prominence upon speech-gesture synchrony without any number of extraneous variables due to the unstructured nature of the task impeding the validity of the findings.

Twelve right-handed Dutch speakers with normal vision participated in Experiment 1, though 4 were excluded secondary to equipment failure and incorrect manual movements. The participants were seated at a table and instructed to place their hand on a designated neutral position on the table. Four pictures were displayed on a vertically mounted plate of plexiglass at the far edge of the table. These pictures were projected via slide projectors. A red LED was under each picture and an additional LED was in the center of these pictures and acted as a fixation point. Each picture was illustrated with white lines on a black background and was 18 x 18 cm. First, the participant heard a warning tone for 500 ms while the fixation LED flashed. After a 1 s pause, one of the LEDs under the pictures flashed for 1s. The participant was required to label the indicated picture while also pointing to the picture. Each verbal response was to be given as a [determiner] + [noun] construction. A total of 2500 ms was allotted for the participant's response after each flashing LED. A total of 16 different nouns were used as stimuli which consisted of four trochees, four iambs, four trisyllabic words with initial position stress, and four trisyllabic words with final position stress. The presentation of the pictures was carefully controlled and resulted in a total of 96 trials. Speech was recorded using a DAT recorder while the manual movement was monitored with a Zebris ™ CMS motion analyzer. The motion analyzer consisted of an ultrasound buzzer placed to the index finger knuckle of the right hand. The ultrasound signal was then picked up

by the Zebris system. Post-data collection, the speech signal and motion data were synchronized.

There were many dependent variables in de Ruiter's (1998) study, especially compared to similar investigations of prosodic stress and speech-gesture synchrony. Speech signal dependent variables included determiner onset, determiner offset, noun onset, stressed syllable onset, stressed syllable offset, and noun offset. Deictic gesture dependent variables included the onset of pointing and the apex of gesture (i.e., "maximal forward extension") (p. 33). A number of analyses were completed, though the analyses of stress position are the focus here. Specifically, de Ruiter asked, "what is it that the speech/gesture system attempts to synchronize with the apex?" (p. 35). The short answer to this question according to the findings of Experiment 1 was that the speech/gesture system did *not* attempt to synchronize with lexically stressed syllables. de Ruiter completed a regression analysis that used the variables of the speech signal to predict when the apex of gesture would occur. The onset and offset of the stressed syllable were not predictive of apex ($p<.45$ and .73, respectively). Furthermore, the gestural apex was not reached later as would be expected for final position stress relative to initial position stress. Though, it is important to note the apex never occurred following the stressed syllable. These findings indicated that despite the results of the other descriptive studies, the timing of gesture was not be affected by prosodic prominence. Yet, de Ruiter rightly acknowledges that the limited scope of the paradigm resulted in a limited prosodic contour and that lexical stress within a bi- or trisyllabic word may not be the equivalent of a "peak syllable" (p. 36) in a more naturalistic context. Experiment 2 attempted to

address this limitation by expanding the length and prosodic variability within the stimuli.

A more naturalistic task was utilized by including contrastive stress. de Ruiter (1998) chose contrastive stress because he was interested in measuring the effect of a pitch-accented syllable within a phrase with more than one stressed syllable. When contrastive stress is present, the pitch-accented syllable is made even more prominent. He states, "the new design enhances the phonetic realization of stress, it allows for a wider range of stressed syllable positions, and it introduces a more marked intonational contour in the production of the speech" (p. 37). Data was analyzed for 8 Dutch speakers in Experiment 2, though 3 other participants were excluded due to time constraint effects and incorrect manual movement. The general setup of the experiment was identical to Experiment 1. The primary difference between these two studies was the stimuli used. The pictures were the same; however, they consisted of different colors to introduce a contrastive element to the task. For example, four pictures of a butterfly were shown, each comprised of a different color. The participant would respond by placing emphasis on the contrastive lexical form, in this case the color of the illuminated picture (e.g., *the GREEN butterfly*). The adjectives were all color descriptors that were either monosyllabic or trisyllabic with final position stress. The number of each canonical shape was not reported. de Ruiter was forced to use these canonical shapes because there are no multisyllabic color descriptors with initial position stress in Dutch. The nouns used were identical to Experiment 1.

The pictures were presented in an array of four that were either four identical pictures of different colors or four pictures that were different but with the same colors.

The structure of the required response was [determiner]+[adjective]+[noun] (e.g., *the green butterfly*). The participants were instructed to place emphasis on the contrastive element of the picture of interest (i.e., picture with flashing LED underneath) while pointing to the picture. The participants were given four seconds to respond in this experiment. Following 24 practice trials, 96 experimental trials were presented in a carefully controlled manner. In addition, half of the sample was presented in reverse order of the other half of the group to further control for order effects. Four percent of the experimental trials were excluded due to participant error.

The dependent variables were modified because of the lengthened stimuli in Experiment 2. The onset of pointing was still measured for the deictic gesture, though the apex was actually measured according to two time points instead of one. This is because participants held on to the apex with a pause of movement prior to retraction instead of a more continuous apex and retraction in the previous experiment. Hence, the beginning and end of gestural apex were measured. The apex beginning was defined as the point when "the forward extension of the hand exceeds 95% of the maximum extension reached during the entire trial" and the apex end was defined as the point when "the hand is extended less than 95% of the maximum extension" (de Ruiter, 1998; p. 39). Another dependent variable is derived from subtracting the onset of pointing from the onset of the apex. This dependent variable is the *launch*. Finally, the duration of the apex was also calculated by subtracting the apex onset time from the apex offset time. Speech dependent variables included utterance onset, adjective onset, adjective offset, noun onset, noun offset, stressed syllable onset for both adjectives and nouns, and stressed syllable offset for both adjectives and nouns.

Again, this review only is focused on the analyses of stress upon the dependent variables. de Ruiter also rightly acknowledges that contrastive and lexical stress cannot be completely dissociated since the contrastive element will always occur on the lexically stressed syllable, at least in this protocol. Therefore, the factors of stress and contrast were collapsed as one factor in the analysis. In contrast to Experiment 1, prosodic prominence was a significant factor in a number of analyses. The onset of the deictic gesture occurred earlier for the first two stress locations (530 and 528 ms, respectively) than for stress locations 3 and 4 (551 and 553 ms, respectively). This results was significant using a one-tailed t-test [$t_1(7) = -2.05$, $p = .04$, $t_2(15) = -1.86$, $p = .04$]. Similarly, the launch was longer when the prominent syllable occurred later in speech compared to earlier ($F(3,21) = 9.89$, $p < .001$), though post-hoc analysis demonstrated that this main effect was due to a significant difference between the first stress position and the other three positions. The launch was also longer for syllables with final position compared to initial position for adjectives, but not for nouns. As would be expected from the launch analyses, stress position also had a significant effect on the apex onset time. A significant correlation ($r = .61$, $p < .001$) between apex onset and stress position further confirmed this relationship. Finally, apex duration also increases with stress position ($F(3,21) = 22.64$, $p < .001$).

In sum, a prominence effect was found only when a more naturalistic prosodic context was examined. Though, it can be argued that the prosodic contour of an phrase like *the GREEN crocodile* and the affiliated deictic gesture is still very constrained. It may seem that by taking away the ability to create novel speech and gesture that this would impede the ability to relate the findings to spontaneous speech. This is somewhat

true, but, by taking away the novelty of the task a confounding variable is also controlled. In spontaneous speech tasks, it is quite possible that lexical frequency may interfere with any potential prosodic stress effect since lexical frequency has also been found to affect the timing of the gesture. Therefore, it is actually beneficial to provide the participant with the required vocal response. Likewise, the constrained deictic gesture is more of a benefit than a hindrance. In order to measure strictly the effect of prosody upon gesture, it is not critical that the gesture be of any particular type or independently constructed by the participant. In other words, the aim is to measure a phonologic/phonetic effect upon the timing of gesture, not a conceptual/lexical effect. Again, by providing the required gesture response and by making the manual response a simple deictic gesture, the lexical and cognitive requirements are lessened resulting in greater assurance that the effect that is being studied is truly the one that is being manipulated. Lastly, the response requirements would not affect the timing relationship between a given syllable and the gestural apex at such a finite level. At the most, the gesture may roughly co-occur with the speech signal because it is a relatively short vocal response and because the participants were instructed to point at the picture while naming it. Only if there was truly an effect of prosodic stress would you expect the timing of the speech and gesture to vary depending upon the carefully controlled stress conditions.

It is difficult to account for the findings of Experiment 2 within the scaffold of the Sketch Model. These results are actually support for the Growth Point Model and refute de Ruiter's assertion that the Formulator does not interact with the gesture production system. de Ruiter states that "this suggests that there is some form of information exchange between phonological encoding and pointing planning or even execution

processes" (p. 46). He later continues, "the finding cannot be incorporated at the level of the conceptualizer or gesture planner, because these processes do not have access to information at the word or syllable level" (p. 61). Subsequently, de Ruiter postulates interdependent manual and articulatory motor processes that may account for these findings, in short a "phase locking mechanism at the level of lower level motor planning, instead of a higher level synchronization process" (p. 61). The Sketch Model has not yet been modified to account for the findings of Experiment 2. Additional research is required that manipulates prosodic prominence in a controlled paradigm both in constrained and naturalistic contexts. Experiments 1 and 2 of the current project aim to accomplish this goal.

Similar to de Ruiter, Rochet –Capellan and colleagues manipulated lexical stress in a controlled paradigm with non-English speaking participants to examine whether gestures synchronized with syllables with lexical stress. Twenty Brazilian Portugeuse speakers were instructed to produce two simple, nonword, phonetic forms, /*papa*/ and /*tata*/ while simultaneously pointing to the written nonword projected onto a screen in front of them. Also designated in the written word was whether the participant was to place stress on the first syllable, *pa'pa* or *ta'ta*, or on the second syllable *papa'* or *tata'*. The text stimuli were presented either in a near position (10 cm from midline) or far position (50 cm from midline). They were also shown a picture of a "smiley face" and were instructed the label they produced was to be thought of as the "smiley face's" name to make the task more natural. Though tightly constrained to only include a single nonword response, the phonetic context and manipulation of the independent variables (stress position, spatial position, and consonant) were well-controlled and the 160

experimental trials were randomized for each of the four blocks, yielding 40 experimental trials per block.

Also like de Ruiter's use of the Zebris ultrasound system, a novel methodology for measuring the gesture and speech variables was employed by Rochet-Capellan and her colleagues. Both the finger movements of the pointing gesture and jaw movements of the spoken responses were measured using Optotrak (Northern Digital, Waterloo, Ontario, Canada). The onset, offset, and apex of jaw and finger movements were tracked using three-dimensional infra-red emitting diodes (IREDs). Although this methodology is frequently employed to measure speech movements, this is the first investigation to integrate the use of IRED tracking for measuring the synchronization of gestures and speech.

A number of time points were measured for each trial. The initiation, apex, and termination of the pointing gesture were recorded. Additionally, the time from gesture onset to apex (forward stroke), time from gesture apex to the start of the return gesture stroke (pointing plateau), and the return stroke (leaving plateau to the termination of the gesture) were recorded. The initiation of jaw movement for each of the two syllables, as indicated by the point of reaching 10% of peak velocity, was recorded for each trial. Likewise, recordings were made of two jaw apices (i.e., the maximum displacement of the jaw marker) for the two syllables in each trial.

Indeed, results of this study indicated that synchronization occurred between gesture and speech movements as a function of prosodic stress. However, results were dependent upon not only presence or absence of stress, but also syllable position. As predicted, jaw apex and gesture apex occurred closely in time, though only when

syllables in the initial position were stressed. When, second syllables were stressed, the jaw apex was closely aligned temporally with the onset of the return stroke and the pointing plateau was significantly lengthened (157 ms) compared to when stress was on the first syllable (127 ms). Though it is interesting to note that despite syllable position, stress was always produced sometime during the time between the onset and offset of the gesture apex (pointing plateau). The authors discussed their findings as evidence for a "bidirectional link between the hand/finger and the jaw" (p. 1519) and support for a dynamic systems perspective of interaction and coupling of two oscillatory systems, as hypothesized in the current project.

Yet, Rochet-Capellan et al.'s findings are contradictory to de Ruiter's results from his first experiment. While de Ruiter did not find that stress position in single words did not affect the time of gesture apex, Rochet-Capellan and others found that indeed gesture apex was produced synchronously with the jaw apex of stressed syllables at least in the first position of words. An obvious difference, other than language produced by the participants, is that de Ruitier manipulated lexical stress using real lexical items, while Rochet-Capellan and her co-authors manipulated stress using phonetically simple nonword forms. Though Rochett-Capellan et al.'s experiment and both experiments conducted by de Ruiter share many features as well, namely an unnatural context consisting of single word or phrase production produced by non-English speaking participants. Additional research is required that manipulates prosodic prominence in a controlled paradigm that extends beyond the word/phrase level with English-speaking participants. The experiments of the current project aim to accomplish this objective.

2.6.4.3 Summary of prosodic prominence and temporal synchrony Though the majority of investigations found that gestures and prominent syllables have a tendency to co-occur (de Ruiter, 1998; Loehr, 2004; McClave, 1998; Nobe, 1996; Rochet-Capellan et al., 2008; Yasinnik et al. 2004), the psychological reality of this assertion is disputable for several reasons. First, McClave (1994) demonstrated that beat gestures occurred with unstressed syllables just as frequently as with stressed syllables. de Ruiter (1998) also found, with a more convincing empirical paradigm than McClave's earlier study, that lexical stress did not affect the timing of the gestural apex. Second, the abundant caveats in the affirmative investigations greatly diminish the validity of the findings. Perhaps the greatest drawback of these investigations is the manner in which synchrony is defined and subsequently measured. Gesture and speech are not absolutely coordinated. In fact, it is highly unlikely that any two events, especially two human movement events, are separated by an interval of 0 ms. As a result, measuring the degree of synchrony between a point in the manual gesture and the speech signal potentially would lend greater insight to the effect of a variable, in this case prosodic prominence, upon temporal synchronization. Such a measurement opens up greater predictions regarding what may or may not reduce the interval between these two time points, in contrast to the more typical observation of a manual movement and some ambiguous point of the associated prominent syllable co-occurring within a 33 ms video frame. Such a measurement was only included in Rochet-Capellan and others' work.

Another limitation of these investigations, with the exception of de Ruiter's (1998) experiments, is the complete lack of systematic control of potential confounds as well as the failure to directly manipulate of the independent variable of interest, namely

prosodic stress. It is nearly impossible to address these concerns with spontaneous speech and is therefore not surprising that the five experiments that utilized a spontaneous speech paradigm are the most difficult to determine what is causing any possible relationship between stress and gesture timing. de Ruiter's manipulation of lexical and contrastive stress was novel and provides the strongest data regarding this hypothesis. Yet, de Ruiter's two experiments produced conflicting findings. In Experiment 1 lexical stress was did not predict the timing of a gestural apex, while in Experiment 2 contrastive stress within a larger intonational contour did tend to affect the timing of the gestural apex.

Despite these many methodological limitations and variations, it is intriguing that many suggested a potential interaction between prosodic prominence and the timing of the associated gesture. These findings highlight the need for a more well-controlled empirical protocol to further examine this relationship. The experiments conducted by de Ruiter and Rochet-Capellan and colleagues stand out among the other investigations and the most theoretically based and carefully controlled. For this reason, the basic methodology of these experiments were expanded upon for this project's first and second experiments.

2.6.5 Gesture and speech perturbation

Perturbing the production of speech or gesture can also lend insight on the effect of processes below the level of the Conceptualizer upon the timing of gesture. If there is no interaction at the level of the Formulator or below, then perturbing one the production systems should not affect the planning and execution of the other system. However, if there is interaction between the two systems at points lower than the Conceptualizer, then perturbing one system could hypothetically alter the timing of the other. In fact, an effect of perturbation upon the corresponding movement indicates a level of interaction even lower than the Formulator. It can be hypothesized that perturbing the speech or gesture movement would only effect the timing of the affiliate initiated movement if the two motor systems were entrained as presented in an earlier section. Evidence to support this claim comes from one study that perturbed gesture production (Levelt et al., 1985) and studies of speech perturbation caused by stuttering (Mayberry & Jaques, 2000; Mayberry, Jaques, & DeDe, 1998), DAF (McNeill, 1992), and speech errors (de Ruiter, 1998) (Tables 10 and 11). However, there has been no study to date that has systematically investigated the effects of speech perturbation upon the temporal parameters of accompanying gestures in a controlled paradigm.

2.6.5.1 Perturbing gesture Levelt and colleagues (1985) conducted a series of four experiments and sought to explain whether the temporal relationship of speech and

gesture was interactive or ballistic (i.e., independent). This study is particularly interesting to relate to the Sketch Model since de Ruiter based his model on Levelt's model of speech production. Levelt and colleagues' work supported the notion that speech and gesture do in fact interact during the planning phase of production but are then independent during the execution phase. Only the fourth experiment is relevant for examining the temporal relationship of gesture and speech production and more than two decades later is still the only investigation in which gesture was impeded.

In this fourth experiment, fourteen right-handed Dutch-speaking participants were seated in view of four red light-emitting diodes (LEDs). Each participant was instructed to point and state "this light" or "that light" as an LED was illuminated. The finger movements were measured by way of an infrared system and the participant's voice onset time was also measured. In order to study the hypothesized interaction of speech and gesture during motor execution, each participant's wrist was attached to an apparatus that could alter the load imposed upon the pointing arm. The apparatus "basically consisted of a suspended mass attached by means of a cord running over a system of pulleys to the subject's wrist" (Levelt et al., p. 155).

The dependent variables measured were voice onset time and apex of the pointing gesture. The primary independent variable was the time the load was applied, beginning of the gesture or halfway through the gesture. Participants also conducted the task with no load applied. In short, Levelt et al. (1985) found that speech was halted only when the load was applied at the beginning of the gesture, not when the gesture was halfway to termination. Voice onset times were similar for the no load condition and halfway-loaded condition, but voice onset time was significantly longer, on average 40

milliseconds longer, for the beginning-loaded condition. When no load was applied in Experiment 4 as well as the preceding three experiments, speech and gesture were tightly linked in time. The investigators concluded that gesture and speech are ballistic "since motor execution will fly blind on whatever it was set out to do in the planning phase, or at least without concern for the other motor system involved" (p. 135).

A major caveat to Levelt et al.'s work is that the task was simplistic and nearly automatic. This controlled study offered possible support for the independence of speech and pointing gestures during execution for tasks with no semantic, syntactic, or prosodic demands but does not offer insight on the temporal relationship of speech and gestures in conversational speech or even in experimentally controlled utterances. Impeding gesture is inherently more difficult than impeding speech. Perhaps that is why Levelt and colleagues work is the only study of the effect of gesture perturbation on gesture-speech synchrony. Perturbing the speech signal for the purpose of measuring the effect on gesture timing is also difficult to achieve, though three types of studies have demonstrated that altering the timing of speech by some type of perturbation affects the timing of the corresponding gesture.

2.6.5.2 Perturbing speech McNeill (1992) was the only investigator that experimentally altered the timing of the speech signal in order to measure the effect on the timing of coinciding gestures. McNeill accomplished the perturbation of speech by imposing delayed auditory feedback during two experiments. The description of this work is far from scientific and was presented within his 1992 text. There is no report of

quanitative data and no report of the measure of temporal synchrony. Nevertheless, the results offer some interesting information and are often used as additional evidence of speech-gesture synchronization and integration.

McNeill's first experiment consisted of an unreported number of participants sho were subjected to a 200 ms auditory delay during Sylvester and Tweety cartoon narrations. They also narrated the cartoons with no imposed delay. As expected from myriad previous research utilizing delayed auditory feedback procedures, the 200 ms auditory delay resulted in slowed speech and dysfluencies characterized by prolongations and repetitions. In addition, gestures were produced more frequently and with greater perceived complexity when the auditory delay was imposed to when there was no delay. The criteria for judging gesture complexity was not described nor are the measurement procedures for determining the temporal relationship of gesture and speech. Yet, McNeill reports that "despite the slowing down and consequent disturbance of rhythm, the gesture is still synchronous with the coexpressive part of the utterance" (p. 275) and later state that the "the relationship of gesture to speech in time is ordinarily firmly fixed" (p. 278).

The results of second exploratory experiment presented by McNeill are actually contradictory to such statements. Two of McNeill's colleagues actually participated in this second experiment. These individuals recited predetermined utterances from memory with a series of continuous iconic gestures that were also predetermined. For example, the phrase *you take the cake out of the oven* was accompanied by the iconic gesture of hands taking a cake out of the oven. Auditory delays of 0, 0.1, 0.2, 0.3, 0.45, and 0.5 seconds were imposed during the participants' recitation of the sentences. The

participants were required to "talk though" the DAF and force "themselves to vocalize continuously, no matter how strong the impulse was to slow down" (p. 280). A critical flaw was that only the 0.2 s delay condition was able to be identified after the completion of the study because the identification key was lost for the other conditions. In contrast to the maintenance of synchrony despite the auditory delay in the first experiment, gesture had a tendency to precede the corresponded spoken affiliates in the second experiment. McNeill explains these disparate results as a function of spontaneous versus recited utterances. It is also quite possible that any effect or lack there of in this second experiment many have resulted from a participant bias given that they were McNeill's colleagues. Hence, the only two experimental paradigms to directly perturb speech yielded discordant results in regards to the effect on the temporal parameters of associated gestures.

A cartoon narration task was also employed by Mayberry and colleagues' investigations of the timing of gesture in relation to speech produced by adults and children who stutter (Mayberry & Jaques, 2000; Mayberry, Jaques, & DeDe, 1998). Observing the timing of gesture and speech during spontaneous dysfluencies enables one to study a naturally occurring perturbation of speech. Six adults who stuttered and six typical speakers participated in their first investigation (Mayberry, Jaques, & DeDe, 1998). In addition, two eleven-year-old boys who stuttered and two typically speaking boys completed the same task in a second component of the study. Later, Mayberry and Jaques (2000) presented the same data from the adult participants in another manuscript, though the reported results are the same with no more elaboration of quantitative measures.

The findings were similar for both the adults and children with the exception that the young participants produced fewer overall gestures than the adult participants. Dysfluent utterances were accompanied by fewer gestures. Normal dysfluencies produced by both the control and experimental group were accompanied by gestures while atypical dysfluencies produced by the experimental group were accompanied by few gestures. Interestingly, in the rare instances that a gesture was produced with an atypical dysfluency, the manual movement would either cease movement midair or fall to rest and then begin again within milliseconds of the resolved lexical affiliate.

Gesture onset always synchronized with fluent speech and never on a repetition or prolongation of atypical dysfluencies in contrast to typical dysfluencies in which the gesture onset did coincide with the onset of the dysfluent lexical affiliate. Mayberry et al. (1998) concluded "gesture execution is so tightly linked to speech production during spontaneous expression that gesture almost never uncouples from speech, even in the face of frequent and often massive disruptions to speech expression caused by stuttering" (p. 85). These results are interesting though are flawed by the recurring issue of including only observational accounts, no explicit criteria the classification of synchrony, and a lack of quantitative data. Not only is it not clear what the degree of synchrony is between speech and the accompanying gesture is in typical and atypical dysfluencies, by Mayberry and others do not make predictions about what stage of speech and gesture production the synchrony between the two is achieved and subsequently maintained in atypical dysfluent utterances. In fact, basic descriptive statistics regarding the number of utterances, percentage with dysfluencies, type of dysfluencies, percentage with accompanying gestures, and so on were not reported.

The last piece of perturbation evidence for shared processing of gesture and speech comes from speech error data. In addition to de Ruiter's (1998) investigation of prosody's role in speech/gesture temporal synchronization and the construction of the Sketch Model, he also presented ad hoc data from speech errors that occurred during one of his prosodic stress tasks. A total of 28 speech errors were produced by six of the eight participants in de Ruiter's second experiment. The details of this task were outlined in the previous section. To review, the participants were required to point to one of 4 pictures while labeling the item with an emphasis on the contrastive element of the picture. For instance, both a green and violet crocodile were presented and if the violet one was indicated the participant would point and state *the VIOLET crocodile*. Eleven errors were consisted of overt repairs and 17 were hesitations. The individuals tended to hesitate or repair their color adjectives because some were the less frequent of the options for the color such as *violet* and *purple*. Additionally, de Ruiter acknowledged that the stimulus pictures for the lizard and crocodile looked similar.

These data can be considered a reflection of a perturbation of speech since on average the onset of speech occurred 166 ms later than in non-error responses. What is more important though is that not only did the onset of speech occur later but the timing of the deictic gesture was also delayed to stay in sync with the speech signal. The onset of the gesture was delayed by 84 ms and the duration of the gesture launch (time from onset to apex) took an additional 117 ms on average, yielding a delay of gesture apex of 184 ms. de Ruiter completed an additional correlational analysis which indicated that the alteration of gesture timing did in fact occur on individual trials and was not simply a false impression created by means across the trials.

Table 12 *Speech and Gesture Elicitation and Classification Procedures: Perturbation and the Temporal Synchrony of Speech and Gesture*

Citation	Speech and gesture remain synchronized with perturbation	Spontaneous	Controlled Speech and/or gestures	Gesture Types	Type of Perturbation	Temporal boundaries
Levelt et al. (1985)	Yes-if perturbed early No-if perturbed late	No	Yes for both	Deictic	Load randomly applied to wrist	Voice onset time Gesture apex
McNeill (1992) Experiment 1	Yes	Yes	No	Not specified	DAF	Not specified
McNeill (1992) Experiment 2	No	No	Yes	Iconic	DAF	Not specified
Mayberry & Jaques, (2000); Mayberry, Jaques, & DeDe (1998)	Yes	Yes	No	Not specified	Fluency disorder	Not specified
de Ruiter (1998) Experiment 2	Yes	No	Yes	Deictic	Speech errors	Gesture onset Gesture launch (onset to apex) Speech onset

Table 13 *Summary of Methodology and Results: Perturbation and Temporal Synchrony of Speech and Gesture*

Citation	Stimuli	Participants
Levelt et al. (1985)	One of four LEDs would illuminate with the response of pointing to the light while stating *this light* or *that light*	n=14 adults
McNeill (1992) Experiment 1	Cartoon narration of Sylvester and Tweety cartoon	Not reported
McNeill (1992) Experiment 2	Recitation of memorized utterances	n=2 adults
Mayberry & Jaques, (2000); Mayberry, Jaques, & DeDe (1998)	Cartoon narration task	n=6 typical adults; n=6adults with fluency disorder; n=2 eleven year-old boys with fluency disorders
de Ruiter (1998) Experiment 2	Four contrastive pictures presented, one of which was designated as the target picture. The participants were required to point to the picture while stating a three word phrase including the contrastive element such as *the VIOLET crocodile*. The stimuli that were included in this analysis were 28 items that were erroneously produced with either overt repairs or hesitations.	n=6 adults

Table 13 (continued) *Summary of Methodology and Results: Perturbation and the Temporal Synchrony of Speech and Gesture*

Citation	Gesture and Speech Segment Classification
Levelt et al. (1985)	Gesture and speech were stringently controlled, therefore no classification procedure was necessary
McNeill (1992) Experiment 1	Not specified
McNeill (1992) Experiment 2	Not specified
Mayberry & Jaques, (2000); Mayberry, Jaques, & DeDe (1998)	Not specified
de Ruiter (1998) Experiment 2	Gesture and speech were stringently controlled, therefore no classification procedure was necessary

Table 13 (continued) *Summary of Methodology and Results: Perturbation and the Temporal Synchrony of Speech and Gesture*

Citation	Results summarized
Levelt et al. (1985)	Participants' speech and gesture productions remained synchronized when the wrist perturbation occurred at the beginning of the gesture, not when applied halfway through the gesture
McNeill (1992) Experiment 1	Increased frequency and "complexity" of gestures with DAF Perceived synchronization of speech and gesture with DAF
McNeill (1992) Experiment 2	Gestures preceded speech
Mayberry & Jaques, (2000); Mayberry, Jaques, & DeDe (1998)	Fewer gesture produced with atypical dysfluent utterances Gestures produced with atypical dysfluencies characterized by a cessation of movement with a re-initiation of movement occurring with the lexical affiliate was fluently produced
de Ruiter (1998) Experiment 2	The onset of the gesture was delayed to remain synchronized with the lexical affiliate

2.6.5.3 Summary of perturbation and temporal synchrony Like the manipulation of prosodic stress, perturbing speech or gesture tests the hypothesis that speech and gesture form an interactive system and that their interaction occurs below the level of the Formulator. Thus far, the studies conducted on this topic have pointed to a tendency for speech and gesture to remain synchronized even when one of the movements is halted. On the other hand this work is preliminary and exploratory and there is data to suggest that synchrony may not be maintained in all circumstances (McNeill, 1992, Experiment 2).

Unlike the manipulation of prosodic stress, perturbing the production of speech tests an interaction of speech and gesture at a lower level than the phonological encoder (i.e., the motor programming level). While finding an effect of prosodic stress upon the degree of speech/gesture synchronization may provide the first systematic evidence for some relationship of speech and gesture, the underlying mechanism of the relationship is ambiguous. A prosodic stress effect may be indicative of the phonetic plan from the Formulator communicating with

the Gesture Planner. On the other hand, a prosodic stress effect could actually reflect self-entrainment of gesture and prominent syllables. Even de Ruiter (1998) postulated that a "phase locking mechanism at the level of lower level motor planning, instead of a higher level synchronization process" (p. 61) may be responsible for gesture synchronizing with prosodically strong syllables.

2.7 SIGNIFICANCE

Thus, by measuring the temporal parameters of gesture following manipulation of not only prosodic stress, but also speech perturbation, one can begin an exploration of the mechanism of speech/gesture synchronization. This project acts as a fundamental contribution to the gesture literature by offering a controlled research paradigm with a reduction of confounding variables prevalent in similar investigations of the temporal synchronization of speech and gesture. Secondly, the use of acoustic analyses of the speech signal and novel use of capacitance analyses of deictic gestures is not only a contribution to the literature because of its innovativeness, but also because of the improved temporal precision of measurement. Also, it is notable that this measure is the only interval-based measure of speech-gesture synchronization to date. Subsequently, this new methodology and measure of the co-production of gesture and speech can be utilized for testing many other research questions regarding the temporal relationship of speech and gesture in both controlled and more naturalistic tasks. Furthermore, the experiments lend insight for both models of gesture and speech production and potential mechanisms of interaction between the gesture planner and the speech production system. This

is the first investigation to specifically examine the role of not only the phonological encoder, but also temporal entrainment upon the degree of synchronization between a manual gesture and speech. Additionally, the experiments will add to the growing literature not only on prosody but also the interaction of spoken language processes with motor behaviors. Lastly, this investigation provides a foundation for a future plan of research on the interaction of speech, language, and manual motor processes during development as well as in typical and disordered populations.

3.0 EXPERIMENT 1

3.1 RESEARCH METHODS

3.1.1 Purpose

The purpose of Experiment 1 was to assess the influence of (i) contrastive pitch accent, (ii) syllable position, and (iii) their interaction on the degree of synchrony between the apices of deictic gestures directed toward a visual display and vowel midpoints of target syllables produced within corresponding carrier phrases.

3.1.2 Experimental Design

Experiment 1 consisted of a two-way, within group repeated measures design. The primary variables are contrastive pitch accent (present or absent) and syllable position (first or second). These variables were manipulated via presentation of the bisyllabic compound noun stimuli. The gesture apex-vowel midpoint (GA-VM) interval for each syllable was measured as the dependent variable.

3.1.3 Participants

Participant enrollment began following approval of the research protocol by the University of Pittsburgh Institutional Review Board. Right-handed, monolingual English speakers without a history of speech, language, hearing, developmental, or neurological disorders were enrolled in the study. Participants were between the ages of 18 and 40. Because there were no anticipated differences in vowel duration or gesture time based upon gender, both male and female participants were enrolled.

Power was set at 0.80, α=.05. The estimated effect size was derived from de Ruiter's Experiment 2 from his dissertation work (1998). As the reader recalls from the literature review, this experiment is the most similar existing investigation to the current project. It is the only experiment that directly manipulated contrastive stress. Furthermore, the experiment utilized deictic gestures and measured the affect of contrastive stress on the temporal parameters of the participants' deictic gestures. The effect size from the statistical results from this study were calculated and used to estimate an appropriate sample size for Experiment 1 and 2 of the current investigation. The effect of contrastive accent upon the duration of the launch (gesture onset to apex) and duration of the apex (onset to offset of the apex) are not only relevant to the GA-VM interval measured in both of the present experiments, but are in fact dependent variables of the second experiment. de Ruiter's results indicate an effect size of contrastive stress upon the gesture launch time was d =0.76 and the effect size of contrastive stress upon the duration of the apex was d=2.37. These two effect sizes correspond to a large and huge effect, respectively, according to conventional classifications. Based upon this information, Cohen's definition of a large effect size (d=0.8) was chosen with an across-condition correlation of 0.5 resulting in a sample size of 15 individuals. This effect size and subsequent sample size are conservative

given de Ruiter's findings. However, the conservative estimate helped to ensure adequate power if there was greater variability in the current experiments due to the differences of between de Ruiter's work and this experiment (e.g., temporal measure of only gesture vs. interval measure between gesture and speech segments, Dutch vs. English speakers, de Ruiter's inability to separate metrical stress and contrastive accent).

Participants were recruited via flyer, and Psychology 101 courses at the University of Pittsburgh, the University of Pittsburgh clinical research website (http://www.clinicalresearch.pitt.edu/), and word of mouth (Appendix A). Fliers were posted within the University of Pittsburgh community and distributed in the mailboxes of students within the School of Health and Rehabilitation Sciences. Additionally, some participants were provided extra credit for their participation within a given course when approved by the instructor. Both a verbal announcement and written information regarding the investigation were given directly to the students of these courses. Additionally, each individual was reimbursed fifteen dollars for their participation time. They were told that they would receive ten dollars with the possibility of receiving a five-dollar "bonus" if they performed the task accurately. This "bonus" served as an incentive to remain attentive throughout the task. In fact, each participant received a total of fifteen dollars, regardless of accuracy.

Participants responded to the recruitment notice via Email or phone call to the principal investigator. A standardized phone script was used to review the purpose, criteria, and procedures of the study. The principal investigator described the study to the potential participant in greater detail including the time required for participation. If the individual expressed continued interest in participating in the experiment, they were then scheduled for completion of both the screening and experimental procedures. No private, identifying

information regarding the participants was recorded until after the participants were consented, with the exception of the individual's name, phone number, and Email address. This information was necessary in order to contact the individual. If an individual chose not to participate, missed their appointment or could not be reached to reschedule, or did not agree to the consent process, their identifying contact information was destroyed. Likewise, no family history information was collected at any time.

The primary investigator reviewed each page of the consent form with participants upon their arrival for their scheduled experimental session. After the consent was reviewed and signed, the participants were screened via interview questions (Appendix B) to ensure that according to their report they were right-handed, monolingual English speakers who did not have a history of speech, language, hearing, developmental, neurological or motoric disorders. Furthermore, each individual was required to pass an audiological screening to assure that they could hear at a level of at least 25 dB at 500, 1000, 2000, and 4000 Hz. These characteristics were exclusionary due to the possibility of resultant effects upon speech and/or gesture response.

The individuals also were questioned regarding past places of residence so that differences in regional dialects would be identifiable. Even though consideration of dialect is included in the study's design, each stimulus has only one permissible form of metrical stress assignment regardless of dialect according to the Merriam-Webster online dictionary (www.m-w.com) and no differences of dialect were anticipated for the relative timing of speech and gesture. Demographic information including age, gender, ethnic/racial background, and highest level of completed education was also collected.

Lastly, a vision screening was completed prior to the initiation of the experiment to assure that a visual deficit did not interfere with the participants' ability to complete the trials

accurately. It was permissible for a participant to have normal or corrected-to-normal vision. Two lines from the Snellen eye chart (Snellen, 1862) were presented to each participant. The Snellen eye chart has been traditionally used and consists of only ten letters, C, D, E, F, L, N, O, P, T, Z. A total of 15 letters were shown to each participant in the same font style and size (44 point Times New Roman) projected on the Plexiglas, identical to the font and size of the text prompts used in the experimental trials. Likewise, the distance to the screen was the same distance that was calibrated for the experimental trials (see section 3.1.5). Participants were to be excluded from the study if they failed to name any of the fifteen letters correctly.

3.1.4 Stimuli

Stimuli consisted of 60 frequent bisyllabic compound words represented by color illustrations. See Appendices C, D, and E for a written list and examples of the illustrations. Each stimulus was a noun with a concrete visual representation (e.g., *surfboard*). Furthermore, each stimulus was composed of a single morpheme (e.g., no plurals) and had only one semantic meaning. Also, the stimuli were composed of word pairs. Fifteen of the stimulus pairs shared the first syllable of the compound word (e.g., *lifeguard* and *lifeboat*). The other fifteen shared the second syllable of the compound word (e.g., *rowboat* and *lifeboat*). Because the first syllables of the word pairs were the same, contrastive stress could be manipulated on the second syllable. Likewise, contrastive pitch accent was manipulated on the first syllable for the word pairs that shared the second syllable of the compound word. The presence and absence of pitch accent was manipulated in response to prompt questions that were visually presented with 44 point Times New Roman text prior to each stimulus presentation. *Is the football above the square*? is an

example of a prompt question. The prompt question was presented visually, rather than auditorily, to reduce the potential confounds of varying prosodic and temporal characteristics of an acoustic signal. Though it is an interesting line of inquiry, there is no research to date to suggest that prosodic contours differ in response to a visual stimulus compared to an auditory stimulus.

Stimuli were presented via Stimulate software onto an 18-inch by 24 -inch piece of Plexiglas that is positioned horizontally using a back projection system. Stimulate is a custom software package created by Necessity Consulting. It is a script driven program utilized for the presentation of video and auditory stimuli as well as the collection of various types of data. Stimulate was also used for the synchronized collection of acoustic and capacitance data in the current series of experiments. Velum was placed on the Plexiglas to reduce glare from the projection system. Likewise, the stimuli were back projected to not only reduce glare from the light source of the projector, but to eliminate shadows created by the gesturing hand during the task when front projection is used. The screen resolution was set at 1024 x 768 pixels.

Illustrations of the stimuli were set at a standardized height of 2.25 inches, though the actual size of the stimuli was considerably greater when projected onto the 18 x 24 inch screen (i.e., approximately 4 .5 inches in height). The stimuli were presented two at a time. One of the illustrations acted as the target and the other as a distractor. A distractor was included to decrease automaticity of responses. The illustrations were presented in one of four possible locations (i.e., above, below, to the right, to the left) relative to an illustration of one of eight possible shapes, (1) square, (2) circle, (3) cross, (4) triangle, (5) diamond, (6) star, (7) rectangle, and (8) oval. The shapes were set at a standardized .75 inch height. A variety of shapes and locations were chosen to increase attention to the task and reduce the risk of rote phrasing

patterns. The shape was always in the central position of the screen and the four locations of the stimulus drawings were the same distance from the center shape to yield similar gesture spaces for all four positions (See Figure 11). Each drawing was approximately one inch from the centrally located shape. Each target word was presented in two of the four possible locations resulting in 120 trials per participant. Estimating that each trial took approximately 15 seconds, the total time required for the experimental trials of Experiment 1 was 1800 seconds or approximately 30 minutes. Because a repeated measures design was utilized, presentation of target and distractor illustrations, position relative to the shape, and type of shape was randomized in order to control for sequencing effects. Furthermore, the location of each stimulus and the type of reference shape was randomized to decrease anticipation of response and sequencing effects.

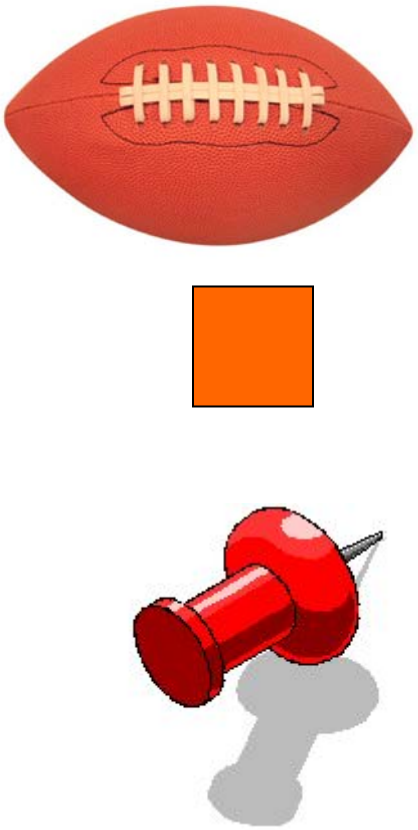


Figure 11 Example of a stimulus presentation. The preceding text prompt for a contrastive trial with this display would be, *Is the baseball above the square?*. The required spoke response then would be, *No, the FOOT'ball is above the square.*

3.1.5 Equipment and Data Collection Procedure

Individuals who passed the screening questions sat in a stationary chair for recording of their speech and manual gestures. They extended their arms on the two arms of the chair and their right hand further extended and resting on the starting position as described below. The participants were gently restrained from moving their torso forward by fastening a Velcro fabric belt around their lower torso and chair back. Also, each participant was instructed to keep their legs and torso still during the trials to further control the distance from the gesture acquisition device and reduce extraneous body, arm, and hand movements.

The individuals directly faced the Plexiglas screen and the gesture movement acquisition device. Their right hand rested on an optical reflective sensor that also acts as the before-mentioned starting position. The sensor is light-sensitive and was used to measure the time of hand lift and return during the gesture movement. See Figure 12 for a visualization of the equipment and participant setup. The hand rested on the sensor in front and slightly to the right of the participant.

Figure 12 Equipment Setup.

A head-worn hypercardioid condenser microphone (Crown CM-312AHS) mounted on to a Sony MDR-V6 headphone was placed on the participant's head with a microphone to mouth distance of approximately two inches. A cardioid microphone was chosen because its heart-shaped sensitivity pattern is optimal for reducing noise from other directions and is most commonly used for vocal recordings. The frequency range of the microphone is appropriate at 50 to 17,000 Hz with sensitivity up to 3.8 mV/Pa low impedance of 75 ohms. The gain of the audio signal was amplified using a TAPCO Mix 120 Mixer.

The spoken responses were digitally audio recorded at a sampling rate of 48 kHz for acoustic analyses. A high sampling rate was chosen to assure adequate recording of high frequency consonant sounds should phonemes other than vowels be incorporated into post hoc

analyses. Thus, this sampling rate was more than sufficient to capture the formant frequencies of English vowels. A sampling rate of 48 kHz was also chosen because it is the sampling rate used by digital media including PC sounds cards. Vowel midpoint was calculated from the acoustic signal collected for each trial. External filters are not necessary for the acoustic or capacitance recordings. The acoustic output was captured with the sound card (C-Media AC97; C-Media, Inc., Version 5.12.1.25) situated on the motherboard of the custom-built PC.

The Facilitator (KayPENTAX™, Model 3500) was utilized to amplify the spoken productions of the participants via Sony MDR-V6 Monitor Series headphones. Each participant spoke into the lapel microphone provided by the Facilitator and the output from the microphone was routed to the device. The device was set at a constant level and amplified each participant's vocal output to approximately 70 dB SPL. The Facilitator is capable of speech-voice amplification with a pass band of 70 to 7800 Hz. Amplification and playback of responses were conducted in Experiment 1 only to be parallel to the procedures of Experiment 2.

The movement of the deictic (i.e., pointing) gestures was captured using capacitance sensors within a modified theremin device (Theramax PAiA 9505KFPA) (Figure 13) and recorded using a Dataq acquisition system (WinDaq/Lite, DI-400; Dataq Instruments WinDaq Acquisition, Version 3.07). A traditional theremin is an electronic musical instrument that requires no direct physical contact to produce and modulate sound. The theremin produces a very unique sound that has been used in soundtracks for sci-fi films, other film soundtracks, and even popular music. The theremin was first inadvertently developed by Russian physicist, Leon Theremin, in 1921 while working with capacitive sensors and short-wave radio equipment for the Russian government. Capacitance by definition is the ability of a circuit element to store an electrical charge (http://www.qprox.com/background/capacitance.php). The amount of

electrical charge stored by an object is dependent upon the distance between it and another object. Capacitance is the charge associated with two adjacent objects that are separated by a non-conducting agent, most often air. The amount of capacitance increases as the size of the objects increases and the distance between them decreases. The amount of capacitance is also dependent upon the material composition of objects. For instance, metal conducts much better than plastic.

Figure 13 Theremax theremin modified for equipment setup as shown in Figure 12.

A theremin works according to these properties. Much like in the mechanics of touch screen devices, it is the ability of human flesh to hold an electrical charge that interacts with the conductors of the theremin. The human body is an excellent conductor and can accumulate charge well. Thus, we are able to accumulate static charge by creating friction between our shoes and carpet while walking across a room. We also will accumulate more charge when walking next to a wall than in the middle of a room since the human body, like any object, has greater capacitance with decreased distance from another object. In fact, the human body can be detected up to a .5 meter away from capacitance sensors (http://www.qprox.com/background/capacitance.php). This distance is much greater than the two (i.e., at closest approximation to the antenna) to eighteen inch space (i.e., at furthest distance from antenna) applied in the current investigation.

It is important to note that there was no physical contact with the Theramax theremin. The device takes advantage of the electrical charge that is inherent within the body/hand and the antennas of the theremin. The device and the participants were electrically isolated. That is, there was no charge distributed by the device to the participant. Properties of capacitance, similar to the device used in these experiments, are utilized in an abundance of commercial devices such as iPods, cellular phones, microwaves, and other touch screen machines.

A theremin consists of two antennas: the horizontal antenna modulates frequency and the vertical antenna modulates intensity. Specifically, for the Theramax unit used in these experiments, a Hartley variable oscillator operated at a frequency of 750 kHz. For each antenna, "the signal from this variable oscillator is mixed with a constant reference frequency in a ring modulator and the result passed through a single pole of low pass filter to leave only the

difference between the variable and reference oscillator frequencies" (http://www.paia.com/theremax.asp). The reader is referred to the PAiA website for technical details and schematics of the design and mechanics of the device. Thus, the signals of the two oscillators are mixed and yield output signals, notably the sum and difference of the two frequencies. The difference of the two frequencies is audible. For example if the frequency of one oscillator is 750 kHZ and the other is 752 kHZ, then 2 kHZ is audible. An oscillator is de-tuned as one's hand is moved towards an antenna, resulting in a change in the capacitance level and a change in the frequency or intensity output signal.

As mentioned previously, these output signals are typically used to create music. The Theremax theremin used in the present investigation is also capable of creating sound as the result of modulating frequency and amplitude of a tone. In addition to a signal that can be perceived auditorily there is also a visual output signal of each antenna. Importantly, the visual output of the theremin is linear. There are two output channels, one for the horizontal frequency antenna and one for the vertical intensity antenna. Thus, in addition to hearing the resulting tone, one can visualize the linear signal associated with the frequency and intensity channels.

Consequently, the methodology of this project takes advantage of the linear output of the two antennas for the measurement of the time of gesture apex. As previously discussed, capacitance and the associated electrical charge increase as an object nears another. Therefore, as the hand and finger draw closer to the frequency antenna while executing a deictic gesture, the charge increases. The maximum charge corresponds to the point of maximum extension, i.e., gesture apex, of the finger and hand movement. The peak voltage of the associated linear output is displayed and recorded on a Dataq oscilloscope as the time point of gesture apex.

A single antenna of the theremin was housed horizontally behind the Plexiglas screen. The voltage trace and peak were recorded as the participant brings their finger in approximation to the screen during the pointing movement. The distance of the participant to the theremin and screen was calibrated prior to the initiation of the experiment. To calibrate the participant-to-screen distance, the participant was first asked to maximally extend their arm and fingers in relation to the screen. Next, a two-inch foam block was placed between their index finger and the screen to assure that the participant-to-screen distance is calibrated consistently across participants. Thus the location of each participant's seated location differed dependent upon the length of their extended arm, hand, and index finger. However, each participant's index finger was two inches from the screen at the point of maximum extension. Lastly, the starting position was slightly cupped in a convex fashion to designate neutral position. The synchronized acoustic and capacitance recordings were archived from the PC to an external hard drive.

3.1.5 Task

3.1.5.1 Stimulus Presentation: Familiarization First, the participants were shown the stimulus pictures to ensure their familiarity with and correct verbal production of the stimulus item. Each of the 60 stimuli were displayed individually in the center of the Plexiglas screen and cycled through twice. The first rotation presented the stimuli along with the associated verbal label printed in black 44 point Times New Roman font. Prior to the second rotation, the participants read instructions in 44 point Times New Roman font to label each picture as it was

3.1.5.2 Instructions The instructions for the practice and experimental trials were the same and were presented via written text (44 point Times New Roman) on the screen. The instructions emphasized the requirement to answer the prompting question while pointing to the correct response. The participants were also given a 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 response were imperative. However, there were in fact no confederates that were identifying their responses. The participants were given this information only to increase the validity of their responses since they may have viewed the dual-response as redundant.

Lastly, a single audiovisual model was shown within the instructions. The audiovisual model consisted of a female completing four practice trials, two with pitch accent and two with neutral accent. The purpose of the audiovisual model was to encourage accurate pointing (e.g., completely lifting and extending the arm and hand as far as possible with a straight arm) and the placement of pitch accent when appropriate.

Instruction Screen 1:

Please begin with your hand on the outline of the hand. This is the starting point.

Instruction Screen 2:
You will first read a question. This question will ask about the picture you will see on the next screen.

Instruction Screen 3:
The pictures you will see are the same pictures you just labeled. When you see the picture on the next screen, you will answer the question while also pointing to the picture.

Instruction Screen 4:
It is important that you both label and point to the picture. After you're done, your responses will be shown to two different individuals. One of these individuals will be only listening to your voice and the other individual will be viewing in silence and will only be able to see your pointing gestures.

Instruction Screen 5:
Let's walk through an example. For instance, you read a question similar to the following, "Is the tugboat above the square?".

Instruction Screen 6:
Sometimes the answer to the question will be a "no" answer. In this case, please emphasize what is different about the picture that makes the answer "no".

Instruction Screen 7:
For example, if a picture of a steamboat appeared above a square a "no" response is required. In order for the response to be completely understood, please emphasize what is different about the picture as follows. "No, the STEAMboat is above the square."

Instruction Screen 8:
Let's see an example. (Audiovisual model will be shown)

Instruction Screen 9:
Other answers will be "yes" answers. For example, if a picture of a tugboat appeared above a square, then the correct answer would be "Yes, the tugboat is above a square". When a "yes" answer is required, there is no need to emphasize the picture label in any way.

Instruction Screen 10.
Let's see an example. (Audiovisual model will be shown.)

Instruction Screen11:

Remember…Always point to the picture while answering the questions. Please lift your hand from the starting position when pointing.

Instruction Screen 12:
After each response, you will be given time to place your hand on the starting position.

Instruction Screen 13:
Would you like these instructions repeated?

Instruction Screen 14:
Okay….Let's try a few examples.

Instruction Screen 15:
Remember to:
**Keep your back against the chair*
**Extend your arm when pointing*
**Both point and speak when responding*

3.1.5.3 Stimulus Presentation: Practice Trials The participants were seated looking at the screen with their hand in a neutral starting position. Each participant completed eight practice trials which were identical to the presentation format of the experimental trials. These eight trials consisted of four compound nouns not present in the experimental set (*steamboat, tugboat, handshake, handcuffs*). Each picture was shown in two different positions. Furthermore, each word was produced with and without contrastive accent. If the participant misunderstood the directions or did not respond with the appropriate verbal or manual response during any of these practice trials, the examiner provided verbal feedback regarding the appropriate response. For instance, if the participant did not lift the hand from the neutral position during the practice trials, they were instructed to lift and extend their hand toward the screen. Additionally, the instructions were repeated if the participant incorrectly produced the carrier phrase or did not mark contrastive accent. Correct or incorrect accent placement during the practice trials was perceptually judged online by the examiner. It was important that the participants provided an

appropriate response during the practice trials since there was no feedback or repetition during the experimental trials. The experimental trials commenced after the eight practice trials were completed.

3.1.5.4 Stimulus Presentation: Experimental Trials Each individual completed 120 trials in Experiment 1. The format for stimulus presentation was identical for practice and experimental trials. Each trial consisted of a shape in the center of the screen and two stimulus pictures. The presentation order of the target words, the distractor words, and their location were randomized. Likewise, the presentation order of the eight shapes (*square, circle, cross, triangle, diamond, star, rectangle, oval*) were randomized and each trial was unique in wording. The objective of decreasing predictability of response was to decrease the likelihood of an individual initiating the deictic gesture or spoken response prior to the stimulus display.

The beginning of each trial was marked by a 6 second presentation of the question prompt (e.g., *Is the bathtub above the square?*) followed by the 7 second 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 seconds. The rationale for randomizing the ISI was to further reduce anticipation of response initiation (e.g., Franks, Nagelkerke, Ketelaars, & van Donkelaar, 1998).

Of the 120 trials, 60 required a contrastive accent response (e.g., *No, the HOTtub is above the square*" and the other 60 required a neutral response (e.g., *Yes, the bathtub is above the*

square". Additionally, 60 of the responses manipulated accent on the first syllable while the other 60 manipulated accent on the second syllable of compound words. Thus, 30 responses elicited contrastive accent on the first syllable, 30 responses elicited neutral accent on the first syllable, 30 responses elicited contrastive accent on the second syllable, and 30 elicited neutral accent on the second syllable. It is important to recall that each target word was presented twice, resulting in the number of trials noted above.

To reiterate, the participant was provided no information or feedback during the experimental trials. Thus, there was also no repetition of trials if a participant provided an incorrect response. Production of incorrect contrastive stress, an incorrect label of the stimulus picture, failure to respond prior to the presentation of the next stimulus item, and failure to point were considered as incorrect responses. A list of all required responses was followed during the experimental trials. If an item was produced in error, the examiner noted the error and the accuracy of the production was later reviewed during the data reduction process. Data points that were produced incorrectly were removed from the final data set. A discussion of excluded trials is presented in the Results section.

3.1.5.5 Data Reduction GA-VM interval was measured for each target syllable. The acoustic signal was analyzed using Adobe Audition 3.0 software to obtain the time of vowel midpoint from a wide-band spectrographic display with a 300 Hz bandwidth filter. Each vowel of the target syllable was isolated according to a modification of the criteria outlined by Higgins and Hodge (2002). They describe the procedure to isolate a vowel as follows "cursors were placed on the first and last glottal pulses that excited the first two formants of the vowel segment….the length of the vowel was measured as the distance between the cursors" (p. C-45).

However, Higgins and Hodge's procedures needed to be altered for the present investigation due to differences in phonetic contexts between their stimuli, *hVd, hV*, and the current bisyllabic word stimuli, that were utilized in the current set of experiments. For instance, many of the stimuli were composed of vowels that were influenced by coarticulatory effects of adjacent nasal (e.g., *thumbprint*) and liquid (e.g., *wheelchair*) phonemes. Therefore, the guidelines employed by Shriberg, Campbell, Karlsson, Brown, McSweeney, & Nadler, 2003) were combined with those originally proposed for this investigation. Their guidelines were to identify vowels "by strong glottal pulsing in the presence of formant structure…nasalized portions of the vowel were in the vowel nucleus duration measures….vowels were segmented from the offset of the preceding consonant to the onset of the following consonant" (p. 561). The vowel onset, offset and subsequent duration was measured for each of the target syllables. Vowel midpoint was calculated by dividing each of the vowel durations by two.

The time of gesture apex of each deictic gesture was extracted from the output voltage signal of the theremin recorded by the Dataq system. As an individual pointed to the stimulus picture, they raised their arm and hand, formed a pointing shape with their index finger extended, and moved the hand and finger toward the screen. Moving the hand and finger toward the screen also brings the hand and finger closer to the antennas of the theremin. As noted previously, the maximum voltage charge of the horizontal frequency antenna occurs at the point of maximum finger extension (i.e., at the point that the finger is closest to the horizontal antenna). This point of maximum extension is the gesture apex as illustrated in Figure 14. This time point was automatically extracted using the Stimulate software package. Also, the examiner verified the measure of gesture apex to assess the presence of outliers. If the automatic extraction of gesture apex was unsuccessful for a trial, the examiner utilized both

the voltage trace from the frequency antenna, intensity antenna, as well as the acoustic signal generated from the antennas to discern the time of gesture apex. If the movement produced an invalid response for the capacitance sensors of the theremin, the video recording was used for final backup analysis. The trial will be discarded if the movement is produced erroneously (e.g., failure to lift hand completely).

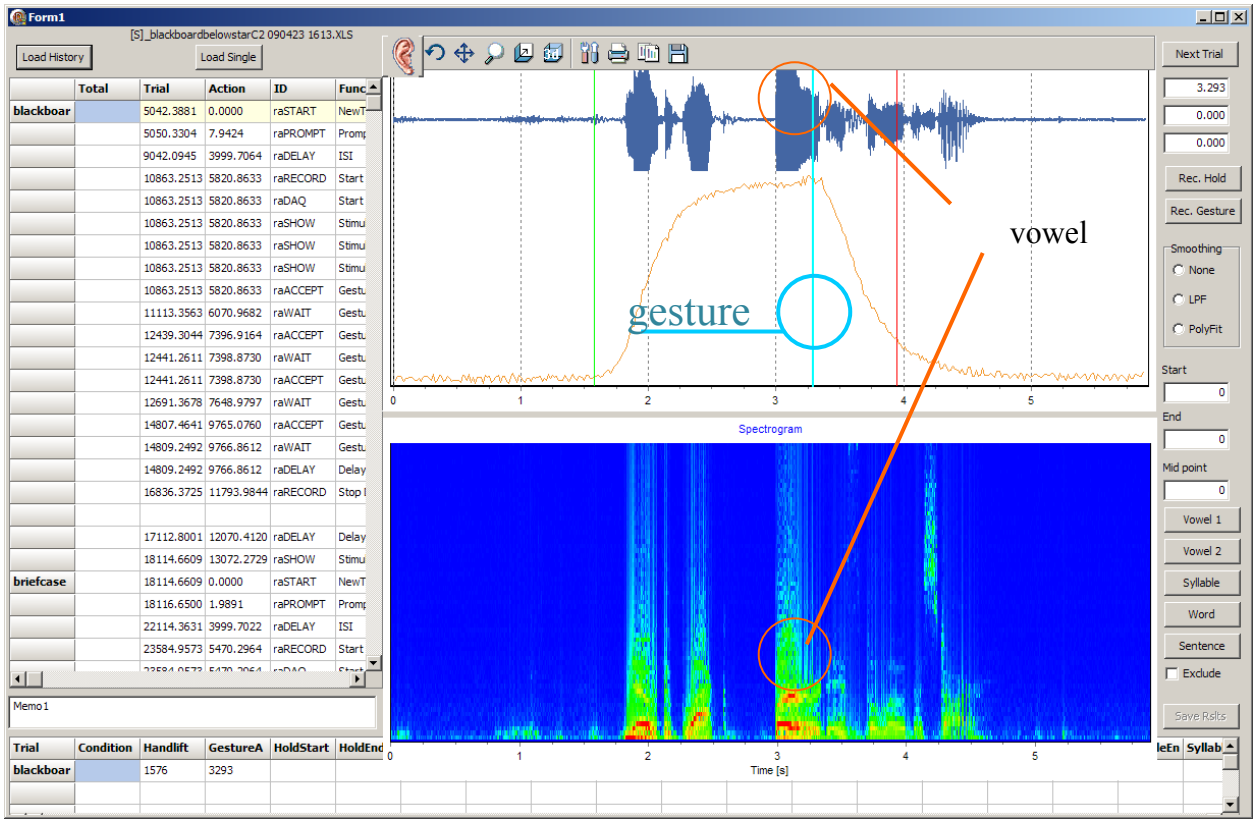


Figure 14 Capacitance voltage trace with gesture apex location and corresponding acoustic signal with approximate vowel midpoint highlighted.

The interval between gesture apex and vowel midpoint was then calculated for each trial. Please refer to Figures 15 and 16 for examples of the predicted differences for GA-VM intervals. The interval is equal to the amount of time in milliseconds between the gesture apex

and vowel midpoint. The interval was reported as an absolute number given that the measurement could be negative if the vowel midpoint occurs prior to the gesture apex. If the interval was not recorded as an absolute number then the negative values would offset the positive values when summed. The information regarding which time point occurred first was also recorded in order to analyze whether the gesture apex occurred prior to the vowel midpoint or vice versa.

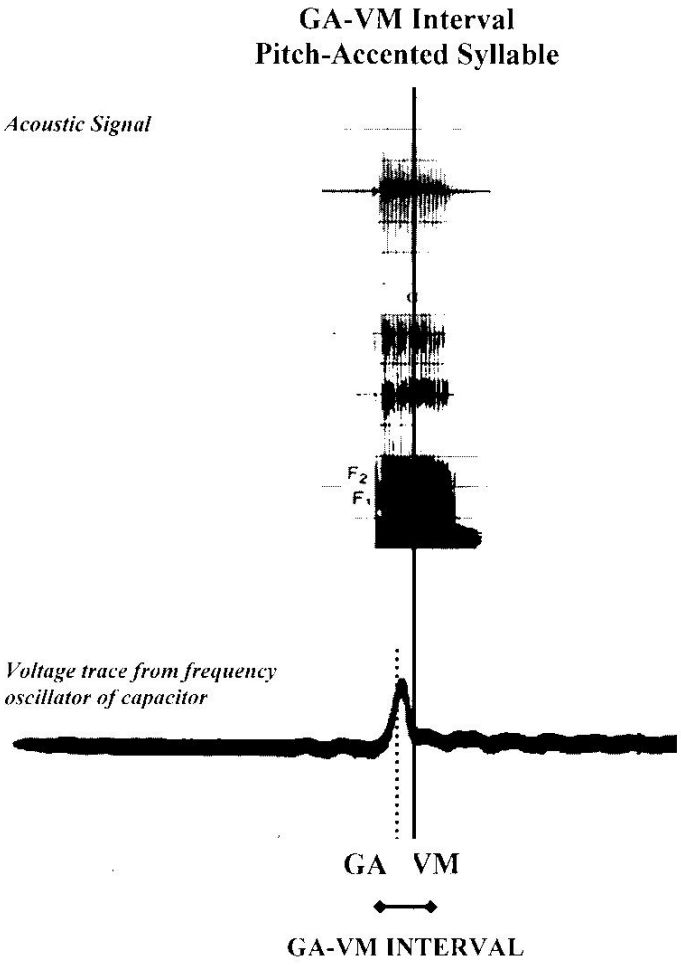


Figure 15 Predicted GA-VM interval of pitch accented syllable and deictic gesture.

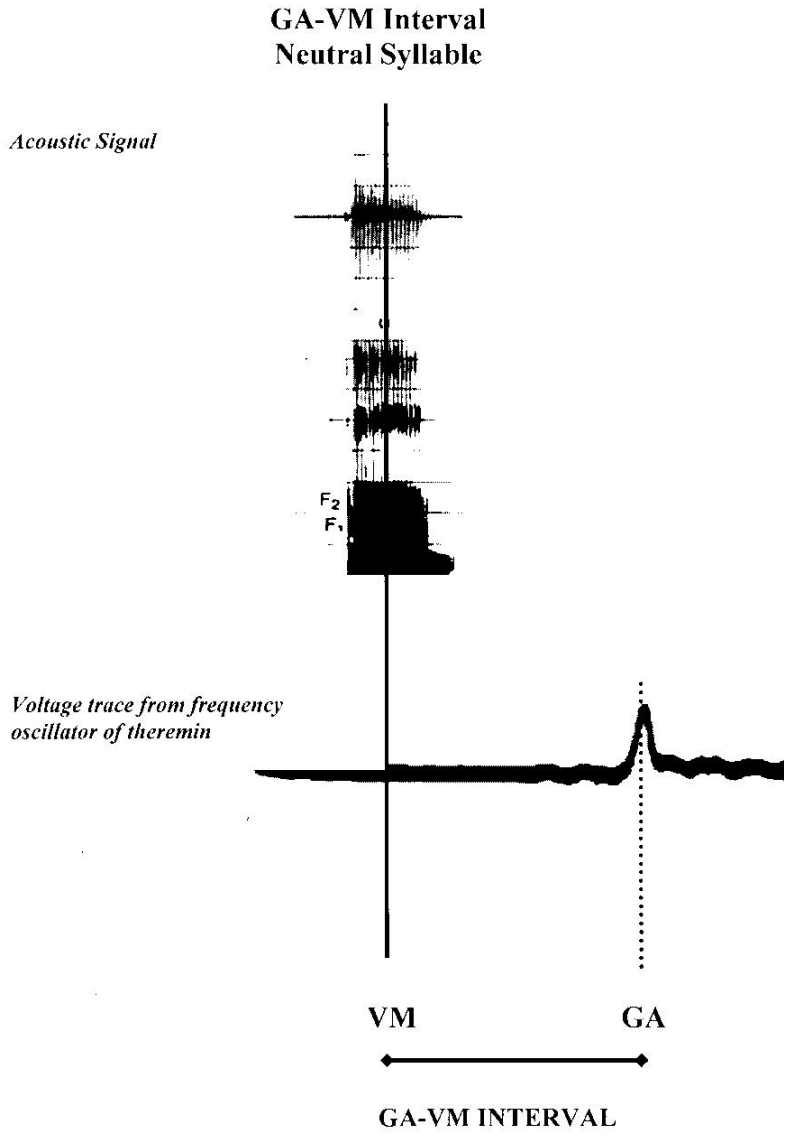


Figure 16 Predicted GA-VM interval of neutral syllable and deictic gesture.

The GA-VM dependent variable was summed for each of the four conditions (i.e.., first syllable-contrastive accent; first syllable-neutral accent, second syllable-contrastive accent, second syllable-neutral accent) for descriptive analyses (e.g., mean, standard deviation, etc.) and statistical comparisons. As described above, there were 120 total data points for the dependent variable for each participant, 30 from each of the four conditions (i.e., presence of pitch accent on first syllable, presence of pitch accent on second syllable, absence of pitch accent on first syllable, absence of pitch accent on second syllable). Hence, 1800 total data points were projected across the 15 participants, with 450 measures of the GA-VM interval in each of the four conditions.

Two (i.e., 13.3%) of the participants were randomly chosen for interrater reliability. A second, independent judge was trained to identify vowel onsets and offsets for all 120 data points for each participant. Thus, there were a total of 240 points for reliability. Correlation coefficients for GA-VM measurements were calculated. Additionally, the means and standard deviations of the two judges' vowel measurements are also reported. Lastly, the number and percentage of excluded trials due to inaccurate stress production, incorrect spoken label production, or incorrect gesture movement are reported.

3.1.6 Statistical Analyses

In addition to descriptive statistics, a two-way repeated measures analysis of variance (ANOVA) was completed for the dependent variable with the significance level set at $\alpha = .05$.

The independent variables were the presence/absence of contrastive pitch accent and first/second syllable position. The dependent variable was the GA-VM interval.

The results were assessed for deviation from normality. Stem and leaf plots generated by SPSS 16.0 indicated some concerns regarding normality of the data for some conditions. As a result, the values for GA-VM were transformed by computing the base 10 logarithm of (x + 1). The results of the ANOVA run on the log-transformed data were consistent with the results of the ANOVA's run on the original data. Therefore, the results presented are from the original, non-transformed dataset.

3.2 RESULTS OF EXPERIMENT 1

3.2.1 Included and excluded participants

A total of 18 individuals were recruited from the University of Pittsburgh community for Experiment 1. None was subsequently excluded on the basis of the exclusionary criteria (e.g, failed hearing or vision screening). The data collected from three of the participants were later excluded due to equipment failure. Thus, the results presented are for fifteen participants as projected by the initial power analysis.

The four male and 11 female participants ranged in age between 21-33 years (*M*=24 years, *SD*=3.3 years) and completed between 15 and 19 years of education (*M*=16.8 years,

SD=1.3 years). All 15 participants were Caucasian and did not speak any languages fluently other than English. Each participant was paid 15 dollars and several also received extra course credit for taking part in the study. The experimental trials required approximately 30 minutes to complete, though breaks were also permitted after each block of trials.

3.2.2 Dependent Variable: GA-VM interval

A total of 120 trials were presented to each participant, yielding a total of 1800 possible responses for this experiment. Seventy-five (4.2%) of these responses were produced in error and excluded from the analyses. Two independent raters listened to each trial to rate accuracy of response. If there was a disagreement between the two raters, a consensus rating was reached by a third independent rater who made the final judgment regarding the accuracy of the response.

Consensus ratings were required between zero and three times per participant. Responses were excluded if the participant produced the spoken response with incorrect stress placement (59), produced the wrong target (13), or failed to respond (3). The number of excluded trials is less than the 6.5% in a similar study using far more simplistic stimuli (Rochet-Capellan et al., 2009). There were no exclusions based upon inaccurate or incomplete gesture movements. In addition to the response produced in error, the paired response of the target was also excluded. For example, if *briefcase* was produced with stress on the second syllable instead of the first syllable in the contrastive condition, then the neutral response of briefcase was also excluded. Thus, a total of 150 (8.4%) responses were excluded from the analyses, yielding 1650 responses.

These 1650 data points were aggregated for each participant and then compared using a two-way repeated measures ANOVA.

Interrater reliability was completed using all trials for two of the participants, or 13% of the data set. A Pearson product-moment correlation coefficient was computed to assess the relationship between the two individuals completing acoustic measures of vowel on- and offsets. Adequate reliability was attained (r=.890) and similar means and standard deviations were noted for the author and reliability volunteer, (*M*=1856.82 ms, *SD*=467.49 ms) and (*M*=1858.70 ms, *SD*=418.25 ms), respectively. No reliability was calculated for the gesture movements due to the automatic extraction of these times points by way of Theramax theremin and recorded by the Dataq system.

Tables 14 and 15 and Figure 17 show the descriptive data for GA-VM intervals across conditions (pitch accent on syllable 1, neutral accent on syllable 1, pitch accent on syllable 2, and neutral accent on syllable 2). The two-way repeated measures ANOVA revealed a significant main effect of *syllable position* [$F(1,14) = 20.268$, $p<.0000$] and *pitch accent* [$F(1,14) = 7.499$, $p<.016$]. There was no significant interaction between these two factors [$F(1,14) = 2.220$, $p<.158$]. The main effect of syllable position was in the predicted direction; however, the main effect for pitch accent was in the opposite direction as predicted. Thus, the mean GA-VM interval was shortest for first position syllables with no pitch accent (*M*=65.16 ms, *SD*=6.94 ms) and longest for second position syllables with contrastive pitch accent (*M*=250.44 ms, *SD*=146.58 ms). This finding is opposite to stated prediction that syllables with contrastive pitch accent would exhibit the greatest synchronization.

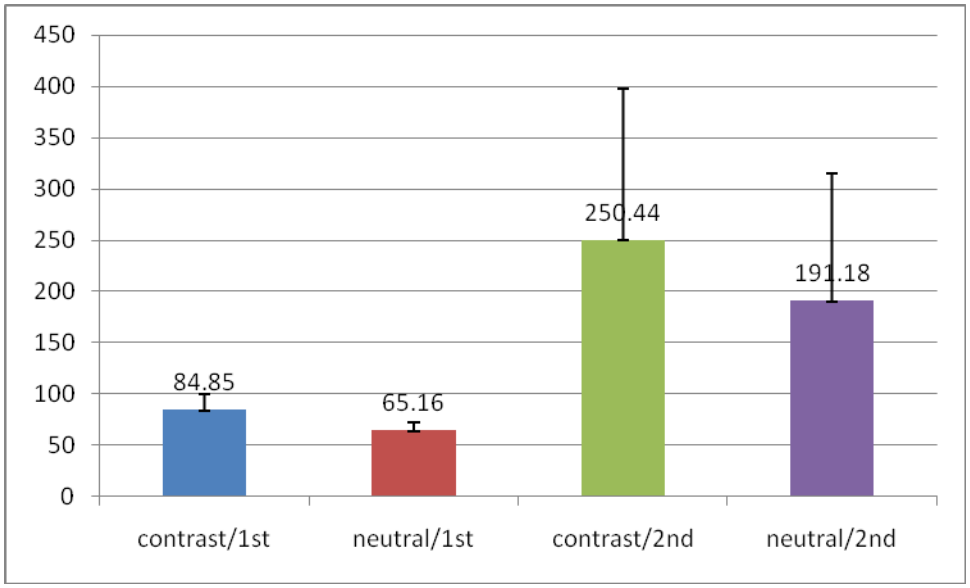


Figure 17 GA-VM intervals (ms) for each condition. Error bars represent one standard deviation.

There was one extreme outlier of the 15 participants sampled as indicated by stem and leaf plots generated by SPSS 16.0. This participant was a 25 year old male. Though he was able to complete the task according to the stated instructions, several notes were made by the examiner during the experimental trials. These notes included, "did not always fully extend arm", "little synchrony noted visually-held apex longer for some of the trials". These perceptual observations translated to exceptionally long GA-VM intervals for this participant in the second syllable with neutral stress conditions. The mean GA-VM interval for this condition was 191.18 ms (*SD*=123.62 ms) and this participant averaged 528.50 ms(*SD*=192.70 ms). Nonetheless, when this individual was removed from the dataset, the results of the analyses did not change. A main effect for *syllable position* [$F(1,13) = 20.720.$, $p<.001$] and *pitch accent* [$F(1,13) = 10.254$, $p<.007$] was found. There was not significant interaction for syllable position and pitch accent [$F(1,13) = 3.583$, $p<.081$]. Thus, the data resulted in failing to reject H0[3].

Table 14 *Descriptive Results for GA-VM Intervals (ms).*

Condition	Mean (ms)	Standard Deviation (ms)	Range (ms)
+PA/1st	84.85	16.54	64.43-132.10
-PA/1st	65.16	6.94	53.88-77.71
+PA/2nd	250.44	146.58	69.02-537.98
-PA/2nd	191.18	123.62	65.43-531.50

Table 15 *Analysis of Variance Summary Table for GA-VM Intervals. * Indicates Statistical Significance.*

Variable	*df*	*SS*	*MS*	*F*	*p-value*	η^2	*power*
Contrast	14	23377.24	23377.24	7.499	.016*	.49	.721
Position	14	318876.05	318876.05	20.268	.000*	.591	.987
Con. x Posit.	14	37029.61	2644.97	2.20	.158	.857	.284

As expected, vowel duration (ms) was increased for pitch accented syllables. Syllables in the first position were longer in duration when produced in the contrastive accent condition (*M*=154.66 ms, *SD*=51.74 ms) relative to when the same syllables were produced in the neutral accent condition (*M*=129.99 ms, *SD*=36.10 ms). This difference was significant according to a dependent samples *t*-test [$t(410)=-11.82$, $p<.000$]. Likewise, syllables in the second position were longer in duration when produced in the contrastive accent condition (*M*=234.19 ms, *SD*=81.45 ms) relative to when the same syllables were produced in the neutral accent condition (*M*=154.66 ms, *SD*=51.74 ms). This difference was also significant according

to a dependent samples *t*-test [$t(428)=-18.15$, $p<.000$]. Thus, speakers modulated contrastive pitch accent as a function of segment duration as found consistently in prior literature. Unexpectedly, vowel duration was longer for both second position syllables than their first position syllable counterparts, (see Figure 18). One would predict that vowel durations would be longer for first position neutral syllables compared to second position neutral syllables because of the trochaic metrical pattern bias of English. On the contrary, second position neutral vowels averaged 154.66 ms (*SD*=51.74 ms) in duration compared to the average duration of 129.99 ms (*SD*=36.10 ms) for first position neutral vowels. Interestingly, the contrastive pitch accented vowels in the second position were also greater in duration (*M*=168.88 ms, *SD*=60.76 ms) than vowels in the first position with pitch accent (*M*=129.99 ms, *SD*=36.10 ms).

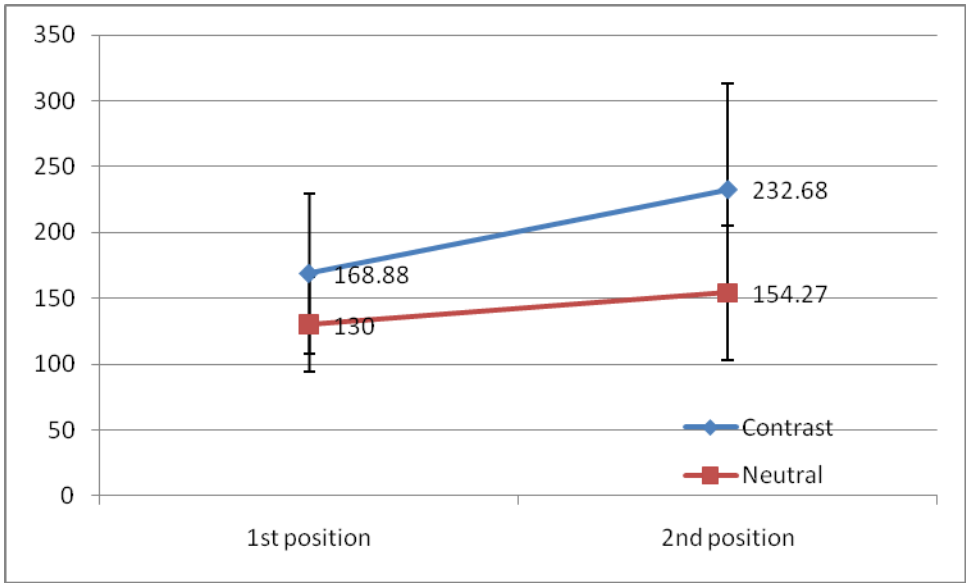


Figure 18 Vowel durations (ms) for each condition. Error bars represent one standard deviation.

In summary, both $H0^1$ and $H0^2$ were rejected based upon the significant GA-VM intervals as a function of contrastive pitch accent and syllable position and the fact that a nondirectional alternative hypothesis and subsequent two-tailed statistical procedure was applied. HO^3 was also rejected because there was no interaction of syllable position and presence/absence of pitch accent. The apex of deictic gestures was more likely to synchronize with first position syllables compared to second position syllables as hypothesized. Yet, the main effect of contrastive accent is not consistent with the motivated direction of the predicted results. In other words, a relationship between the timing of a deictic gesture and the corresponding lexical item does exist based upon this data set, though the effect of syllable position and prosodic prominence result in an explanatory quagmire.

3.3 DISCUSSION OF EXPERIMENT 1

The purpose of Experiment 1 was to investigate the influences of prosodic stress and syllable position on the timing of the point of maximum extension of deictic gestures. Though it is often observed that gestures and their lexical affiliates roughly occur in time during spoken language production, this study is only one of three that has directly manipulated prosodic stress and controlled the gesture elicitation procedure. The present experiment stands alone in experimentally studying the effect of prosodic stress and syllable position on the timing of deictic gestures in a complete multiword utterance and for English-speaking participants.

Overall, it was clear that there was a coordination of speech and gesture and that the time of gesture apex was differentially influenced by (1) first versus second syllable position and (2) presence versus absence of contrastive pitch accent. The greatest synchronization on average (i.e., smallest mean GA-VM interval) was noted for first position syllables with neutral stress while the least synchronization (i.e., largest mean GA-VM interval) was measured for second position syllables with contrastive stress. This finding was inconsistent with the hypotheses initially presented.

The lack of synchronization of gesture apices to prosodically stressed syllables in Experiment 1 is consistent theoretically with de Ruiter's (1998; 2000) Sketch Model and is evidence against Tuite's Rhythmic Pulse Model. The Sketch Model posits that there is no interaction between the speech and gesture production systems at the level of the Formulator, the mechanism responsible for lexical retrieval and prosodic stress assignment among other planning processes. Thus, the manipulation of contrastive pitch accent on a given syllable should not have affected the timing of the corresponding deictic gesture. According to de Ruiter, speech and gesture are initiated at the same time, though the onset of gesture most often occurs before the onset of the lexical affiliate because "gesture production is less complex, and therefore less time consuming, than for speech" (de Ruiter, 1998, p. 61). The current results do not indicate the gesture planner interacts with the phonological encoder due to the lack of synchronization between syllables with contrastive pitch accent and deictic gesture. If there were no interaction between processing levels lower than the Conceptualizer, then one would expect both the gesture and lexical affiliate to be initiated relatively simultaneously. The Sketch Model postulates that the processing involved in produced a gesture, especially a deictic gesture, is less complex, takes less time and therefore causes the gesture to precede the lexical affiliate. If this is accurate, then

one would expect the greatest synchronization (i.e., smallest GA-VM intervals) to be measured for the first syllable position, regardless of pitch accent assignment. Indeed, this was the case in the current experiment such that the GA-VM intervals were significantly lower for the first syllable position, regardless of pitch accent assignment. In other words, the apex of gestures coincided with the initial onset of a word, not with prosodically prominent syllables.

What is perhaps most striking is the extremely small GA-VM intervals averaged across participants for the first position syllables. The intervals between a gesture apex and the first syllable midpoint were no greater than 65 and 85 *milliseconds* for neutral and contrastive conditions, respectively. To put this in perspective, 100 ms is the benchmark for software developers so that processing will seem instantaneous and is also the human reaction time for advanced athletes. In fact, one of the fast human reflexes, the patellar (i.e., knee jerk) reflex requires approximately 50 ms. This extreme tightness between the gesture apex and vowel midpoint was unanticipated for neutral syllables. However, there is previous research that demonstrated a tight interval for first position stressed syllables. Rochet-Capellan and others (2008) found that jaw apex-pointing apex intervals averaged only 11 ms for stressed, initial syllables in the nonsense word task. In contrast, the jaw apex-pointing apex interval for second position stressed syllables was nearly 15 times longer, 151 ms in duration. These authors did not make any comparisons between stressed and unstressed syllables like in the current experiment (e.g., *PA*'pa vs. *pa*'PA) but rather the differences in the timing of deictic gestures dependent upon first or second position stress (e.g., *PA*'pa vs. pa*PA*'). However, the general findings of Rochet-Capellan et al. and Experiment 1 certainly share parallels. Future analyses of Experiment 1 data could include measuring the vowel midpoint of the nontarget syllables and the

non-absolute values of the GA-VM intervals to further relate these findings to those of Rochet-Cappellan et al. and support the selection of dependent variables for later investigations.

There are several theoretical postulations to account for these seemingly puzzling findings. First, it is possible the gesture apex was not the appropriate time point to use as the synchronization point to the spoken response. Gesture apex was chosen for valid reasons and because of the desire to record a single time point within the gesture for the dependent measure. Yet, when looking at entire movement trajectory, perhaps the maximum point of extension is not the ideal time point for such a measure. 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. A visual inspection of random gesture and speech displays across participants further supports the need to record additional points in the gesture, in particular the gesture launch duration and midpoint. Figure 19 is a single, but seemingly representative example of a speech/gesture response. The light blue line is the gesture apex and occurs after the vowel midpoint of the target syllable, which is in the second position and produced with contrastive pitch accent. In contrast, the gesture onset occurs prior to the initiation of the vowel,

with the gesture launch midpoint occurring at a more synchronous moment with the target vowel midpoint.

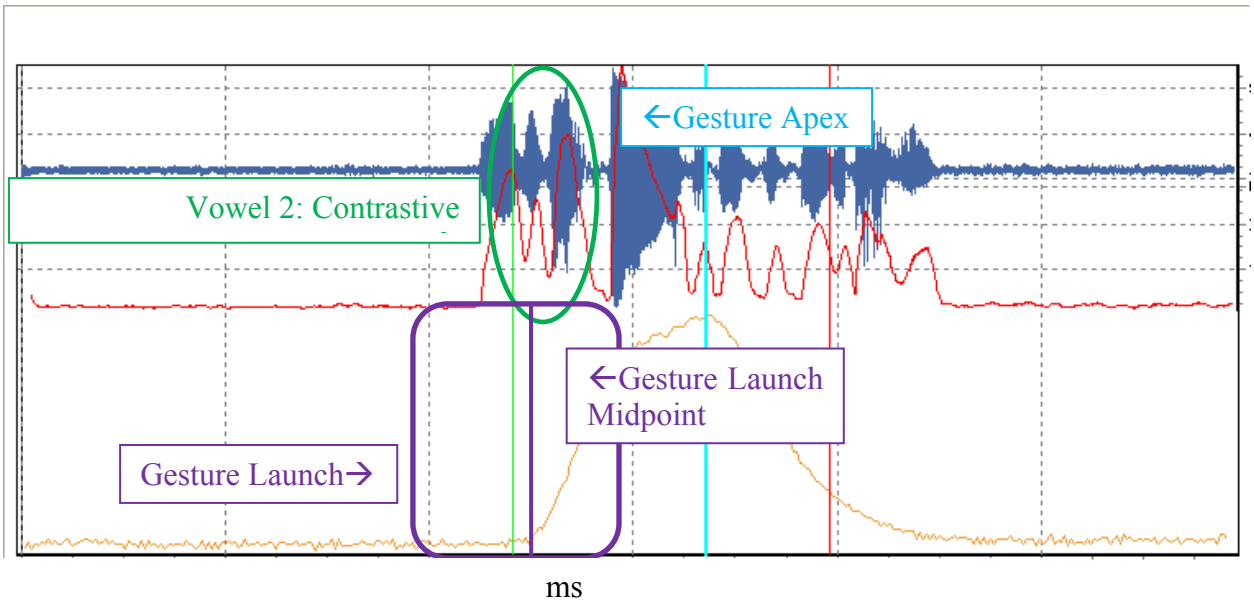


Figure 19 Comparison of gesture launch and gesture apex in relation to the stressed second position syllable. The response is: "No, the lightBULB' is below the square". The top trace is the acoustic signal and the bottom signal is the voltage trace obtained from the theremin.

It also is possible that other points in the gesture are altered to align with prosodically prominent syllables, even in this experiment. For instance once an apex is reached, it can be held before return to a rest position. An individual may hold a gesture apex to wait for the stressed syllable, especially for stressed syllables in the second position. This hypothesis also was put forth by de Ruiter (1998) who found that the duration of gesture apices were indeed longer for later stressed syllables in a contrastive stress task. He states, "if the operation of the strict phonological synchrony rule is by itself not sufficient to obtain full synchronization, the

conceptualizer might compensate by lengthening the gesture upon receiving feedback from the comprehension system" (p. 45).

Hence, measuring additional variables other than the gesture apex may clarify why greater synchronization was noted for neutral syllables than pitch accented syllables. Certainly, the synchrony of gesture apices and first position syllables is not unexpected. On the other hand, the finding that gesture apices synchronized with *unstressed* syllables was an unexpected finding. One would predict that gestures would either synchronize with prosodically prominent syllables or not synchronize as a function of this variable at all. Upon first glance, it appears that the data do not support a lower-level entrainment of the speech and gesture, such as that proposed by Tuite (1993) and Iverson and Thelen (1999). In fact, this study was not explicitly designed to examine the dynamic entrainment of speech and gesture. Yet the observation that gestures synchronized with the contrastive stress independent variable, albeit the control condition, leads one to inquire further about the role of lower-level processes in the timing of speech and gesture. Therefore, it is proposed that additional work be completed to systematically study the temporal relationship of gesture and prosodically stressed syllables employing additional manipulation of the spoken response and additional dependent measures. The objectives of Experiment 2 were in line with such a proposal.

4.0 EXPERIMENT 2

4.1 RESEARCH METHODS

4.1.1 Purpose

The objective of Experiment 2 was to assess the influence of (i) contrastive pitch accent, (ii) syllable position, (iii) speech perturbation, and (iv) their interaction during the production of deictic gestures directed toward a visual display and vowel midpoints of target syllables within corresponding carrier phrases produced by typical adults on the degree of synchrony between deictic gestures and speech. The temporal synchrony of speech and gesture was examined by measuring the dependent variables of (a) GA-VM interval and two temporal parameters of deictic gestures, (b) total gesture time and (d) gesture launch time.

4.1.2 Experimental Design

Experiment 2 consisted of a three-way (2 x 2 x 2), within group repeated measures design. The three variables of Experiment 2 were contrastive pitch accent (i.e., present or absent), speech perturbation (i.e., presence or absence of 200 ms auditory delay), and syllable position (first or second). These variables were manipulated via presentation of the bisyllabic

compound noun stimuli. The dependent variables measured were the GA-VM interval for each target syllable as well total gesture time and gesture launch time.

4.1.3 Participants

Though the participants for Experiment 2 were independent from the sample of Experiment 1, the inclusionary and exclusionary criteria for the participants were the same as were the recruitment, consent, and screening procedures. As stated in Chapter 3, effect sizes from de Ruiter's (1998) study of the effect of contrastive stress upon the timing of deictic gestures were calculated and guided the estimate of effect size and resultant sample size for this investigation. Power was set at 0.80, α=.05. An estimated effect size of d=0.8 with an estimated across-condition correlation of r=0.5 yielded a sample size of 12 participants. Again, however, 15 participants were enrolled to be conservative.

4.1.4 Stimuli

The stimuli were identical to those used in Experiment 1. Likewise, the presentation order of each stimulus picture, location, and centrally located shape were randomized as in Experiment 1. In contrast to Experiment 1, the stimuli were presented a total of four times, resulting in a total of 240 trials per participant. Each target word was presented twice under the influence of DAF and twice without DAF. The same comparisons were made for contrastive

pitch accent and syllable position both with and without DAF imposed by doubling the amount of trials from Experiment 1. Similar to Experiment 1, there were 120 unique responses; though in this experiment they were produced twice, once with DAF and once without DAF. Estimating that each trial took approximately 15 seconds, the total time required for the experimental trials of Experiment 2 was 3600 seconds or approximately 60 minutes for each participant.

4.1.5 Equipment and Data Collection Procedure

The equipment and data collection procedures were also identical to Experiment 1 with two notable exceptions. First, an optical sensor was placed at the starting location directly in front of each the participant's right arm, which was bent in a forward direction. This sensor is light-sensitive and automatically recorded the time of gesture onset and offset. The synchronized acoustic, capacitance, and optical sensor recordings were archived from the PC to an external hard drive.

Second, the participants' speech was perturbed secondary to the manipulation of the presence and absence of delayed auditory feedback (DAF). DAF is described by Pfordresher & Benitez (2007, p. 743) as "a constant time lag (that) is inserted between produced actions (e.g., piano keypress) and the onsets of auditory feedback events (e.g., the onset of a pitch)". DAF causes a breakdown of fluency in typical speakers characterized by decreased speech rate, prolonged voicing, increased speech errors (e.g., phoneme exchanges), increased vocal intensity, and increased dysfluencies (e.g., prolongations and part-word repetitions) (e.g., Burke, 1975; Howell & Archer, 1984; Stuart, Kalinowski, Rastatter, & Lynch, 2002).

A 200 ms auditory delay was chosen. The rationale for selecting a 200 ms auditory delay is based upon the ubiquitous finding that this duration yields the most consistent breakdowns in the temporal execution of speech produced by typical adults secondary to the asynchrony the individual's spoken production and auditory feedback of their speech (Finney & Warren, 2002; Marslen-Wilson & Tyler, 1981; Stuart, et al., 2002). Additionally, a 200 ms delay is typically chosen with delayed auditory feedback (DAF) because it is approximately the length of an average syllable (Smith, 1992).

Consequently, 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 different for DAF conditions compared to normal auditory feedback (NAF) conditions. Although the speech of many typical speakers is perturbed by DAF as previously described, it cannot be assumed that all speakers produced delayed spoken productions. Therefore, the sentence durations for each speaker were compared as a function of NAF and DAF conditions to assure that utterances spoken under DAF were longer than the same sentences produced without DAF. Mean time error (MTE) for each individual was measured 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 word production under the influence of DAF compared to NAF conditions. Elman 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, 1983). The greater the MTE measurement (recorded in milliseconds), the

greater the time difference between DAF and NAF conditions. Sentence duration was also analyzed across participants using a three-way ANOVA.

The Facilitator (KayPENTAX^{TM}, Model 3500) was utilized to amplify and delay the spoken productions of the participants. The Facilitator is an external device that is capable of real time amplification, delayed auditory feedback, immediate loop playback, speech range masking, and metronomic pacing. Each participant spoke into the microphone provided by the Facilitator and the output from the microphone was routed to the device. The Facilitator is capable of presenting an auditory delay between 10 and 500 ms in 10 ms increments, though only a 200 ms delay was used in this experiment as motivated earlier. The feedback, real-time and delayed, was presented via Sony MDR-V6 Monitor Series Headphones.

The device was set at a constant level and amplified each participant's vocal output to approximately 70 dB SPL, regardless of the presence or absence of an auditory delay. The Facilitator is capable of speech-voice amplification with a pass band of 70 to 7800 Hz. Higher loudness levels have been found to elicit greater speech disruptions (e.g., Elman, 1983). Furthermore, the majority of experiments have manipulated the intensity of the auditory feedback to be at a comfortable listening level and it is commonly accepted that 70 db SPL approximates typically conversational loudness levels. In fact, Howell and Sackin (2002) found that this loudness level approximated 70 dB SPL for all eight of the participants in their study of the effects of DAF on syllable repetition. The nonaltered auditory feedback (NAF) conditions also were amplified to 70 dB SPL to remain consistent with the DAF conditions.

The issue of order effects is addressed in various ways in the speech perturbation literature. While researchers who have imposed a mechanical load upon an articulatory structure like the lower lip presented the experimental and control trials in a randomized fashion

(e.g.., Munhall, Löfqvist, & Kelso, 1994), those that have imposed an auditory delay typically used a block design to counterbalance the DAF and NAF conditions to control for practice effects (e.g., Pfordresher & Benitez, 2007; Finney & Warren, 2002; Howell & Dworzynski, 2001; Howell & Sackin, 2002; Jones & Striemer, 2007; van Wijngaarden & van Balken, 2007). In line with previous research, DAF and NAF conditions were counterbalanced. Blocks of 20 trials of either the DAF or NAF conditions were presented in succession. For example, 20 trials of DAF were followed by 20 trials of NAF, then 20 trials of DAF, 20 of NAF and so on. The presentation order of the two conditions was randomized across participants such that some began with NAF trials and others with DAF trials. Additionally, the trials within each block were randomized to control for practice effects of the stimuli that may have affected the temporal parameters of the other two independent variables (i.e., syllable position and contrastive pitch accent).

Another notable difference between investigations who utilized a mechanical load to perturb speech and those who utilized DAF is the control for habituation and anticipation of the perturbation. Typically, a mechanical load is imposed on a small number of trials to control for any effects of habituation and anticipation. For example Munhall et al. (1994) imposed a load to the lower lip on only 12% of the trials. There is no such control for habituation in the DAF literature. The same trials that are presented in the NAF conditions are also presented in the DAF conditions. Likewise, the counterbalanced NAF and DAF blocks prohibit the effectiveness of reducing the number of experimental trials since they still will be presented in succession and, thus, anticipated. There is concern about individuals habituating to an auditory delay when used for prolonged lengths of time as a fluency-enhancing technique but this is not addressed in studies that utilize DAF in a similar manner to the present experiment. Thus,

identical stimuli were presented in the NAF and DAF counterbalanced conditions. Each target word was presented twice, once in the presence of DAF and once without. Thus, a total of 240 trials were completed by each participant, 120 in the NAF condition and 120 in the DAF condition.

4.1.6 Task

4.1.6.1 Instructions The instructions were similar to those in Experiment 1 though information regarding the DAF was provided to reduce possible disruptions and subsequent errors that could occur from the unanticipated perturbation. Although alerting the participants to DAF introduced knowledge of the perturbation, anxiety about possible equipment failure and/or incorrect response on the part of the participants was reduced by providing this information. The instructions conveyed that their speech would sound louder to them through the earphones and that on some trials their speech could be heard later than they when they actually speak.

4.1.6.2 Stimulus Presentation: Familiarization The familiarization procedure was similar to Experiment 1. However, the participants also were familiarized with the auditory delay. After they were familiarized with the stimulus pictures, they were told that they would read, but hear their voice later than when they actually spoke. They were then visually presented with the first four sentences from the Rainbow Passage (http://web.ku.edu/idea/readings/rainbow.htm). Each sentence was back-projected in isolation via text on the Plexiglas screen. The participants were instructed to read each sentence at comfortable loudness level. A 200 ms delay was imposed by the Facilitator while they read the four sentences. They were not given any further

instruction regarding whether to "ignore" or to "pay attention" to the DAF. The familiarization with the DAF was simply to demonstrate to the participants that the delay was a known and acceptable component of the experiment.

4.1.6.3 Stimulus Presentation: Practice Trials Participants completed the same eight practice trials as in Experiment 1. They were not exposed to delayed auditory feedback during the practice trials.

4.1.6.4 Stimulus Presentation: Experimental Trials Each individual completed 240 trials in Experiment 2. The stimulus presentation for the experimental trials was exactly the same as in Experiment 1. Remembering that double the number of trials was presented in Experiment 2, 60 trials required a contrastive accent response (e.g., *No, the HOTtub is above the square*) and the other 60 required a neutral response (e.g., *Yes, the bathtub is above the square*). Additionally, 60 of the responses placed accent on the first syllable while the other 60 placed accent on the second syllable of compound words. Thus, 60 responses had contrastive accent on the first syllable (e.g., *HOTtub*), 60 had neutral accent on the first syllable (e.g., *hottub*), 60 had contrastive accent on the second syllable (e.g., *lifeGUARD*), and 60 had neutral accent on the second syllable (e.g., *lifeguard*).

Again, unlike Experiment 1, a 200 ms delay was imposed during 50% of the trials. Therefore, 30 trials in each condition were produced with DAF and the other 30 without DAF resulting in 120 total trials produced under the influence of DAF and 120 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. The examiner provided no information or feedback during the experimental trials.

4.1.7 Data Reduction

Measurement of the GA-VM interval was completed for each trial as described for Experiment 1. A number of additional dependent measures of gesture were obtained in this experiment. These three variables were total gesture time, gesture launch time, and sentence duration. Total gesture time equals the time measured in milliseconds between gesture onset and offset as recorded by the optical sensor. Gesture launch time equals the time in milliseconds between gesture onset and gesture apex. Acoustic analyses were completed using Adobe Audition 3.0 software to calculate sentence duration for each trial. Total sentence duration was equal to the duration from the onset of the utterance to the offset of the utterance. Specifically, the acoustic waveform was analyzed and the onset of the initial word of the carrier phrase and the offset of the shape label were isolated. The time between these two points in milliseconds equals the total sentence duration. Lastly, MTE was measured for each individual by comparing the mean difference between the total sentence duration of the trials produced without an auditory delay and the trials produced under the influence of a 200 ms delay.

Interrater reliability was calculated for 13.3% of the participants (n=2). Just as in Experiment 1, a second, independent judge was trained to identify the onset and offset of the target vowels. For this experiment the second judge also measured the onset and offset of the sentence. A Pearson product-moment correlation coefficient was computed to assess the relationship between the two individuals completing acoustic measures. Adequate reliability was attained for both vowel onsets and offsets (r=.981) and sentence onsets and offsets (r=.997). Similar means and standard deviations of the on and offset points were noted for the author and reliability volunteer for both vowels and sentences. The mean vowel time points for the author were 2800.12 ms, *SD*=490.43 ms) and 2792.43 ms (*SD*=489.79 ms) for the second judge. Likewise, the author's mean sentence on and offset measures were 2174.39 (*SD*=1307.38 ms) compared to 2167.26 ms (*SD*=1293.89 ms) for the second judge. No reliability was calculated for the gesture movements due to the automatic extraction of these times points by way of Theramax theremin and recorded by the Dataq system.

The gesture dependent variables and GA-VM interval listed above were summed for each of the eight conditions (i.e., (1) first syllable-contrastive accent with NAF; (2) first syllable-contrastive accent with a DAF; (3) first syllable-neutral accent with NAF; (4) first syllable-neutral accent with DAF; (5) second syllable-contrastive accent with NAF; (6) second syllable-contrastive accent with DAF; (7) second syllable-neutral accent with NAF; (8) second syllable-neutral accent with DAF) for comparisons for descriptive analyses (e.g., mean, standard deviation, etc.) and statistical comparisons.

4.1.8 Statistical Analyses

As noted earlier in this chapter, a three-way ANOVA was completed with sentence duration as the dependent variable to analyze differences in utterance length between DAF and NAF trials. A dependent samples t-test also was performed to test the prediction that pitch accented vowels would be longer than vowels with no contrastive pitch accent. In addition to descriptive statistics, a separate three-way repeated measures analysis of variance (ANOVA) was completed for each of the dependent variables. The independent variables were the presence/absence of contrastive pitch accent, the presence/absence of speech perturbation via 200 ms auditory delay, and first/second syllable position. The dependent variables were GA-VM interval, total gesture time, and gesture launch time. Post-hoc simple main effects were analyzed using Bonferroni corrected pairwise comparisons.

4.2 RESULTS OF EXPERIMENT 2

4.2.1 Included and Excluded Participants

A total of 28 individuals were recruited from the University of Pittsburgh community for this study and were independent of the participants enrolled in the first experiment. There were no participants excluded on the basis of the exclusionary criteria (e.g., failed hearing or vision screening). The data collected from 13 of the participants were later excluded due to equipment failure, specifically a software problem that caused the gesture onset and offset time points to be omitted from data recording. Thus, the results presented are for fifteen participants, three more than projected by the initial power analysis.

The four male and eleven female participants ranged in age between 22-31 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). All 15 participants were Caucasian and did not speak any languages fluently other than English. Each participant was paid 15 dollars and several also received extra course credit for taking part in the study. The experimental trials required approximately 60 minutes to complete, though often the participants took breaks during the procedure.

4.2.2 Data Reduction

A total of 240 trials were presented to each participant, yielding 3600 possible responses for this experiment. Of these, 169 were excluded (4.7%) of these responses were produced in error and excluded from the analyses. Two independent raters listened to each trial to rate accuracy of stress placement. If there was a disagreement between the two raters, a consensus rating was reached by a third independent rater who made the final judgment regarding the accuracy of the response. Consensus ratings were required for zero to five responses per participant. Responses were excluded if the participant produced the spoken response with incorrect stress placement (77), produced an error on the target (e.g., speech error, hesitations, coughed; 78), failure of DAF equipment (7), or gesture error (7; e.g., scratched head before pointing). As in Experiment 1, the number of excluded trials was less than the 6.5% in a similar study using bisyllabic nonword stimuli and corresponding pointing gestures (Rochet-Capellan et al., 2009). Errors for each participant ranged from 0 to 44 excluded responses (M=11.3, SD=11.0).

A conservative approach was taken for including trials that were important for later paired comparisons that may be of interest. Therefore, in addition to the response produced in error, all three paired responses of the target also were excluded. To be clear, these additional three responses were excluded even though they actually were produced accurately. For example, if *briefcase* was produced with stress on the second syllable instead of the first syllable in the contrastive condition and with DAF, (1) the same response produced without DAF was excluded as were (2-3) the neutral responses of *briefcase* produced in the DAF and NAF conditions. Thus, a total of 676 (18.8%) responses were excluded from the analyses, yielding 2924 responses.

The results were assessed for extreme outliers and deviation from normality. Stem and leaf plots generated by SPSS 16.0 did not indicate extreme outliers for any of the ANOVA's performed. Based upon the same stem and leaf plots and assessment of the means and standard deviations, there were concerns regarding normality of the data for some conditions of some of the dependent measures. As a result, the values for GA-VM, total gesture time, gesture launch time, and sentence duration were transformed by computing the base 10 logarithm of (x + 1). The results of the ANOVA's run on the log-transformed data were consistent with the results of the ANOVA's run on the original data. Therefore, the results presented are from the original, non-transformed dataset.

4.2.3 Sentence Duration and Effects of Speech Perturbation

The data first were analyzed to evaluate the effects 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. Mean time error (MTE) was calculated for each participant (Elman, 1983). Indeed, 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. The mean MTE was 867.53 ms (*SD*=681.52 ms) and ranged from 120 to 2346 ms across participants. In other words, on average utterances were 867 ms longer in the DAF condition than when spoken in the NAF condition. The average MTE for the participants was greater than the average MTE of 381 ms (*SD*=343 ms) calculated by Elman (1983). The increase in average MTE is likely accounted for by the discrepancy in task requirements since the participants in Elman's investigation were

only asked to open and close their lips in a tapping fashion. As expected, there was increased variability of sentence duration for DAF trials as compared to the same responses produced without an auditory delay.

A three-way ANOVA was performed to assess the impact of DAF on the duration of the spoken responses, as well as the other manipulated variables (i.e., contrastive pitch accent and syllable position). As expected, only two main effects were significant. Indeed, sentences were significantly longer in duration when produced with DAF [$F(1,14) = 24.049$, $p<.0000$]. 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) as illustrated in Figure 20. Taken along with the MTE calculations, it was confirmed that DAF perturbed the spoken productions of each participant, allowing further investigation of the effects of this perturbation on the temporal characteristics of the corresponding deictic gesture, thus allowing the investigation of the effects of all 15 participants. 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); [$F(1,14) = 39.420$, $p<.0000$]. This 369 ms increase was confirmation that participants increased the duration of vowels when produced with contrastive pitch accent. Additional analyses of vowel durations will be presented later. There were no significant interactions observed for sentence duration.

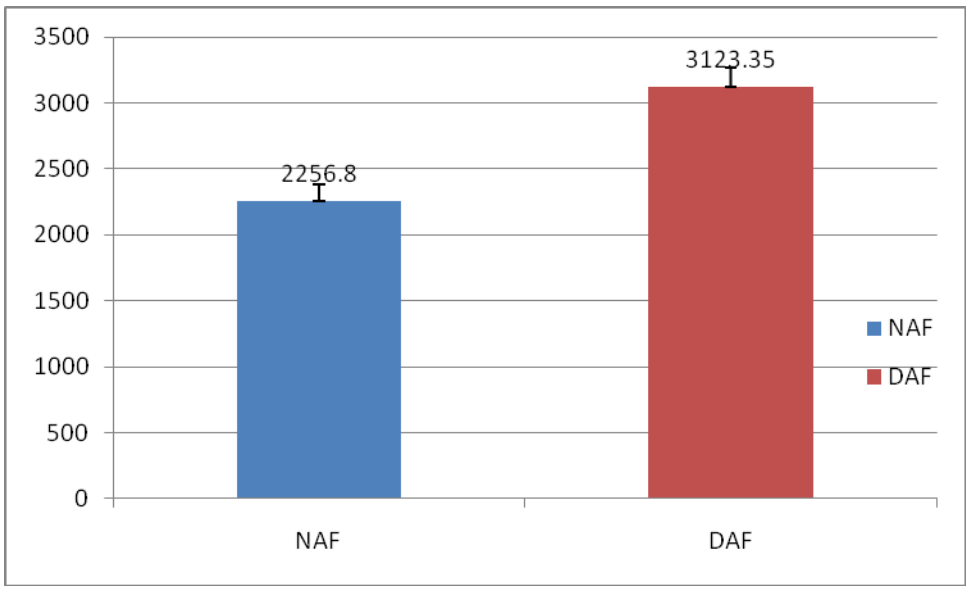


Figure 20 Sentence duration times (ms) for DAF and NAF conditions.

4.2.4 Vowel Durations

Vowel As expected, vowel duration (ms) was 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), just as in the first experiment of this investigation. This difference was significant according to a dependent samples *t*-test [*t*(1461)=-19.314, *p*<.000]. Vowels were also longer in DAF conditions (*M*=265.84 ms, *SD*=122.73 ms) than vowels produced in without auditory feedback (*M*=194.69 ms, *SD*=85.42 ms). This difference was also significant [*t*(1461)=-19.618, *p*<.000].

Once again, though still unexpectedly, vowel duration was longer for both second position syllables than their first position syllable counterparts (see Figure 21). Second position syllables averaged 235.07 ms (*SD*=109.55 ms) compared to 190.38 ms (*SD*=99.80 ms) for first

position syllables [$t(1461)=-11.17$, $p<.011$]. Moreover, second position syllables significantly differed in duration as a function of contrast [$t(730)=17.781$, $p<.000$]. Vowel durations were almost 100 ms longer on average for second positions syllables with contrastive pitch accent (M=275.59 ms, SD=113.36 ms) relative to when the second position syllables were produced with neutral stress (M=195.22 ms, SD=87.19 ms). Thus, individuals significantly lengthened vowel durations in the presence of DAF, when assigned contrastive pitch accent, and when in the second position of the target word (see Figure 21).

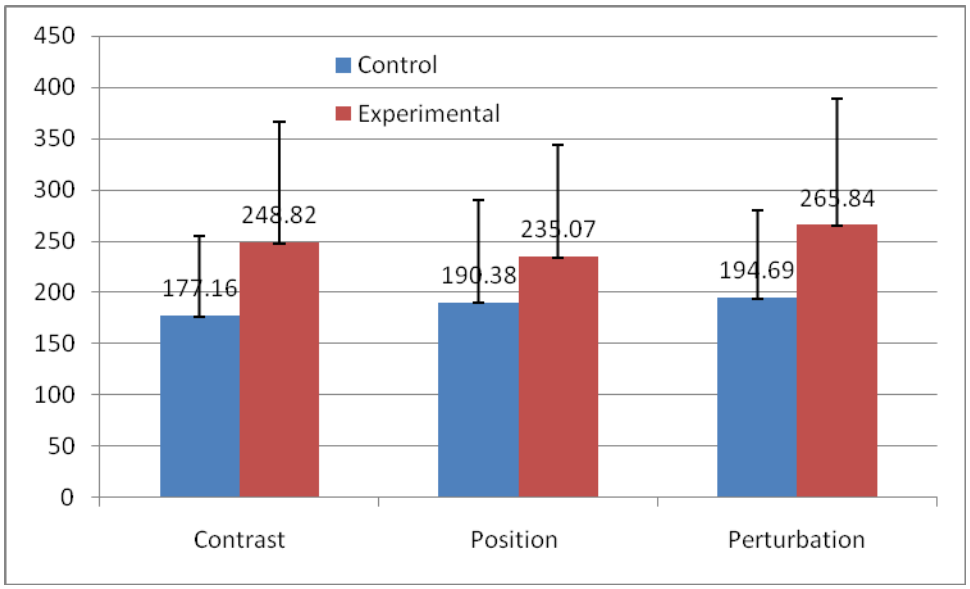


Figure 21 Vowel durations (ms) for control and experimental conditions: contrast (neutral and contrastive accent), syllable position (first and second position), and perturbation (NAF and DAF). Error bars correspond to one standard deviation.

4.2.5 Dependent Variable: GA-VM Interval

Table 16 and Figure 22 show the descriptive data for GA-VM intervals across for the three independent variables, contrast, syllable position, and speech perturbation. The three-way repeated measures ANOVA revealed a significant main effect only of *perturbation* [$F(1,14) = 6.593$, $p<.022$]. There was a significant interaction between *perturbation* and *syllable position* [$F(1,14) = 17.063$, $p<.001$]. Refer to Table 17 and Figure 23 for presentation of summary data.

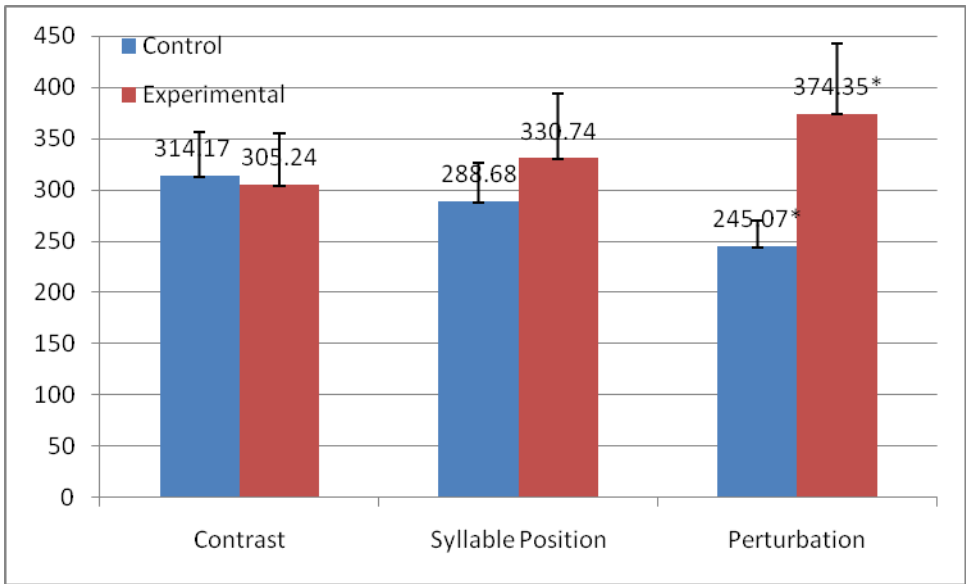


Figure 22 GA-VM intervals (ms) for control and experimental conditions: contrast (neutral and contrastive accent), syllable position (first and second position), and perturbation (NAF and DAF). Error bars correspond to the standard error. * Indicates significant main effect.

Table 16 *GA-VM Intervals (ms) for Each Condition (-PA: Neutral Accent; +PA: Contrastive Accent).*

Condition	Mean (ms)	Standard Deviation	Range
-PA/1st/NAF	250.26	110.30	69.41-428.66
-PA/1st/DAF	322.28	232.92	83.66-895.07
-PA/2nd/NAF	238.52	166.82	59.90-535.40
-PA/2nd/DAF	445.63	327.42	63.48-1043.08
+PA/1st/NAF	274.77	131.37	109.26-510.29
+PA/1st/DAF	307.41	217.06	57.30-801.50
+PA/2nd/NAF	216.73	155.53	69.73-691.48
+PA/2nd/DAF	422.06	399.43	63.43-1452.70

GA-VM intervals were significantly longer for target syllables produced with an auditory delay (*M*=374.35 ms, *SD*=68.19 ms) as compared target syllables produced without an auditory delay (*M*=245.07 ms, *SD*=25.53 ms). In regards to the interaction effect, the GA-VM intervals for NAF trials were shorter for second position syllables (*M*=227.63 ms, *SD*=37.72 ms) than for first position syllables (*M*=262.51 ms, *SD*=28.42 ms). The results for GA-VM intervals for the DAF trials were in the opposite direction. GA-VM intervals for DAF trials were shorter for first position syllables (*M*=314.85 ms, *SD*=56.08 ms) than for second position syllables (*M*=433.85 ms, *SD*=91.55 ms). Counter to the predicted outcomes, GA-VM intervals were not significantly shorter for first position syllables or for syllables with contrastive pitch accent.

Table 17 *Analysis of Variance Summary Table for GA-VM Intervals; * Indicates Statistical Significance.*

Variable	*df*	*SS*	*MS*	*F*	*p-value*	η^2	*power*
Contrast	14	2392.178	2392.178	.148	.707	.010	.065
Position	14	53059.231	53059.231	.630	.441	.043	.115
Perturbation	14	501379.999	501379.999	6.593	.022*	.320	.666
Con. x Posit.	14	5668.783	5668.783	.627	.442	.043	.115
Con. x Pert.	14	3177.110	3177.110	.895	.360	.060	.143
Posit. x Pert.	14	177609.448	177609.448	17.063	.001*	.549	.970
Cn x Pos x Pert	14	2648.814	2648.814	.343	.567	.024	.085

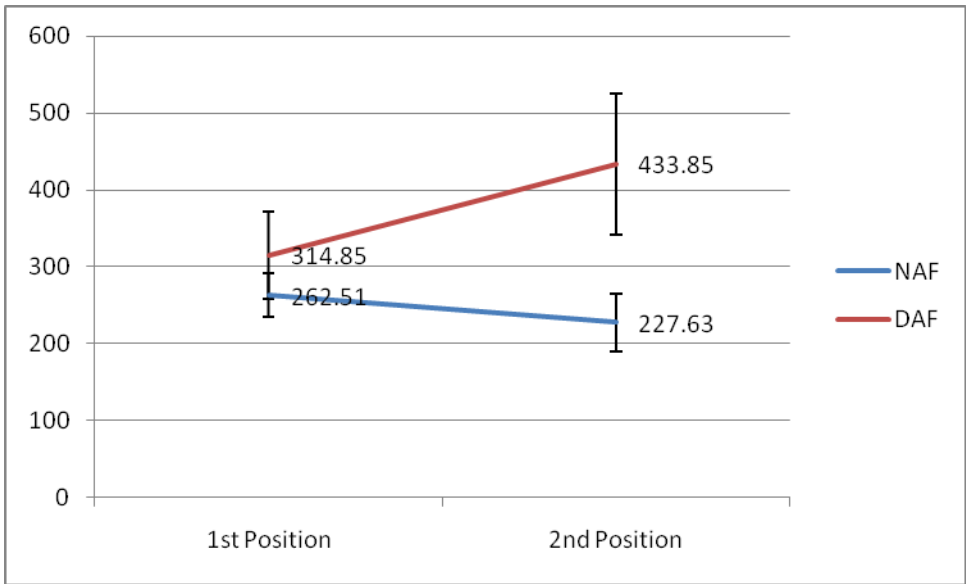


Figure 23 GA-VM intervals (ms) for perturbation by syllable position. Error bars correspond to standard error.

The significant two-way interaction of *perturbation* x *position* was further analyzed utilizing a post hoc analysis of simple main effects using the Bonferroni correction. The analysis of simple main effects allows one to examine the effects of one independent variable while the other independent factor is held constant. Significance was adjusted from .05 to .025 because two post hoc tests were performed to interpret the interaction between DAF and syllable position. As presented in Table 18, the effect of DAF was only significant for the second syllable position [t(30)=3.84, p<.0003] and not for the first syllable position [t(30)=.97, p<.170]. Hence, DAF perturbed the synchronization of the gesture apex and vowel midpoint only for the target syllables in the second position.

Table 18 *Results of Post Hoc Simple Main Effects of Perturbation x Position Interaction. df=Degrees of Freedom, t=Test Statistic, and p=Significance Level. * Indicates Statistical Significance.*

Pairwise Comparisons for Syllable Position	***df***	***t***	***p***
DAF versus NAF: first syllable	30	.97	.170
DAF versus NAF: second syllable	30	3.84	.0003*

These results are counter to the predicted outcomes. H_0^1 and H_0^2 were not rejected because there were no significant differences between the GA-VM intervals as a function of presence/absence of pitch accent and first/second syllable position, respectively. Failure to reject H_0^1 and H_0^2 is also contradictory to the findings of Experiment 1 where a significant main effect was found for both contrast and position for GA-VM interval. The findings do result in rejection of H_0^3 and H_0^4. GA-VM intervals were significantly longer for DAF trials compared to NAF trials. Likewise, an interaction of syllable position and DAF was demonstrated such that GA-VM intervals were longest for trials produced with DAF and second position target syllables.

The contradictory findings to Experiment 1 support the earlier proposal that additional measures of the gesture movement were required to elucidate whether speech and gesture are temporally entrained. The conclusion that there was less synchrony for DAF trials than NAF trials does not provide support for tight entrainment of the two systems when faced with perturbation. However, additional dependent variables are necessary to explore the temporal relationship of speech and deictic gesture.

4.2.6 Dependent Variable: Total Gesture Time

Total gesture time was predicted to increase as a function of increased spoken response time, whether due to DAF or contrastive pitch accent. Though there were significant effects found for the dependent measure of total gesture time, gestures required approximately 1 ½ seconds to complete regardless of condition (see Figure 24). On average, gestures were 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). Total gesture time was longer for second position syllables (*M*=1575.09 ms, *SD*=123.09 ms) compared to first syllable position (*M*=1534.34 ms, *SD*=130.19 ms). Lastly, it was predicted that the time to complete a gesture would be longer for sentences produced with DAF relative to NAF and this was indeed the case. Total gesture time for DAF conditions averaged 1609.40 ms (*SD*=150.55 ms) compared to 1500.03 ms (*SD*=103.39 ms) for NAF conditions.

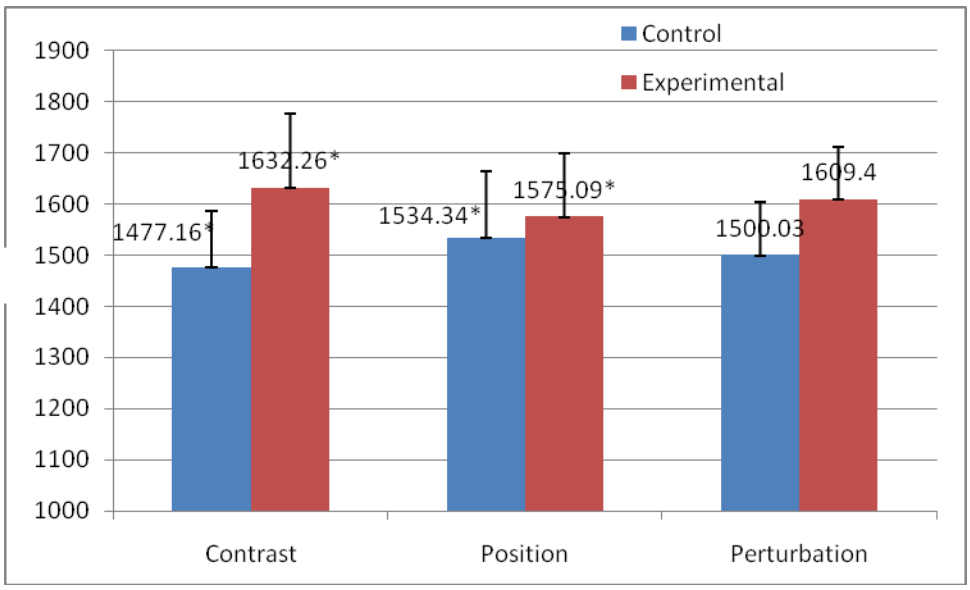


Figure 24 Total gesture time (ms) for control and experimental conditions: contrast (neutral and contrastive accent), syllable position (first and second position), and perturbation (NAF and DAF). Error bars correspond to standard error. * Indicates significant main effect.

Total gesture time was analyzed for syllable position, presence/absence of contrastive pitch accent, and presence/absence of speech perturbation with a 2 x 2 x 2 ANOVA (see Tables 19 and 20). A significant main effect for *contrast* [$F(1,14) = 10.087$, $p<.007$] and *syllable position* [$F(1,14) = 6.344$, $p<.025$] emerged. There was a significant two-way interaction between *contrast* and *syllable position* [$F(1,14) = 23.004$, $p<.000$] (see Figure 25). Total gesture time was longest for trials that held contrastive pitch accent on the second syllable (M=1671.87 ms, SD=140.28 ms). Furthermore, there was an average 79 ms difference between the total gesture time for first and second position condition when produced with contrastive stress and only a 2 ms difference for first and second position condition when produced with neutral stress.

Table 19 *Total Gesture Time (ms) for Each Condition.*

Condition	Mean (ms)	Standard Deviation	Range
-PA/1st/NAF	1408.40	327.73	1060.17-2033.14
-PA/1st/DAF	1543.65	546.94	1018.37-2759.00
-PA/2nd/NAF	1445.90	368.92	1045.97-2250.89
-PA/2nd/DAF	1510.71	476.20	1008.33-2420.85
+PA/1st/NAF	1529.50	465.60	1049.90-2508.00
+PA/1st/DAF	1655.79	711.63	982.41-3229.29
+PA/2nd/NAF	1616.30	462.24	1109.03-2570.89
+PA/2nd/DAF	1727.44	636.54	1069.80-3089.11

Table 20 *Analysis of Variance Summary Table for Total Gesture Time; * Indicates Statistical Significance.*

Variable	*df*	*SS*	*MS*	*F*	*p-value*	η^2	*power*
Contrast	14	721625.745	721625.745	10.087	.007*	.419	.839
Position	14	49824.507	49824.507	6.344	.025*	.312	.649
Perturbation	14	358870.516	358870.516	4.434	.054	.241	.500
Con. x Posit.	14	44409.168	44409.168	23.004	.000*	.622	.994
Con. x Pert.	14	2619.955	2619.955	.153	.701	.011	.065
Posit. x Pert.	14	13735.943	13735.943	2.142	.165	.133	.276
Cn x Pos x Pert	14	5732.915	5732.915	1.046	.324	.070	.159

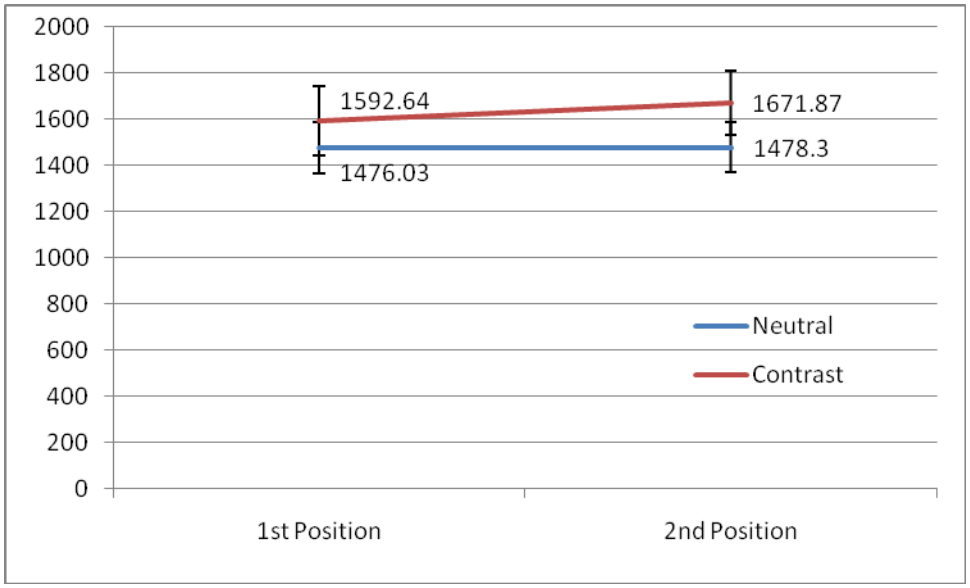


Figure 25 Total gesture time (ms) indicating a significant two-way interaction for *position x contrast.* Error bars correspond to standard error.

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. Significance was adjusted from .05 to .025 because two post hoc tests were performed to interpret the interaction between contrast and syllable position. As presented in Table 21, the effect of syllable position was only significant for the second syllable position [t(30)=17.50, p<.0000] and not for the first syllable position [t(30)=1.57, p<.063]. These results signify individuals lengthened the time of their gesture when trials were produced with contrastive pitch accent, but reached significance only for the condition with accent on the later second syllables.

Table 21 *Results of Post Hoc Simple Main Effects of Contrast x Position Interaction. df=Degrees of Freedom, t=Test Statistic, and p=Significance Level. * Indicates Statistical Significance.*

Pairwise Comparisons for Syllable Position	***df***	***t***	***p***
Contrast versus neutral: first syllable	30	1.57	.063
Contrast versus neutral: second syllable	30	17.50	.0000*

The increased total gesture time for the DAF condition relative to the NAF condition was marginally significant [$F(1,14) = 4.434$, $p<.054$]. A subsequent calculation of effect size resulted in a small-medium effect size ($f=.19$). That is, even though there were increased gesture times on average for DAF trials, this difference did not reach significance or a substantial effect size.

On average, individuals did not produce gestures, from the time of onset to offset, with any vast variability. However, the significant effects of position, contrast, and marginal effect of perturbation indicates that individuals altered the timing of a gesture in response to changes within the spoken response. The significant interaction of *position x contrast* demonstrated that

individuals lengthened their gesture time the most when speakers produced prosodic prominence on the second syllable of the target word. Yet, in regards to the stated hypotheses, H_0^5 could not be rejected because there was no significant difference of total gesture time for DAF compared to NAF trials. The next dependent measure presented, gesture launch time, enabled a more focused exploration of this potential temporal relationship, specifically from the onset to the apex of the deictic gesture.

4.2.7 Dependent Variable: Gesture Launch Time

It was hypothesized that gesture launch time would increase as the spoken response increased in duration. Spoken response time could increase either secondary to speech perturbation via DAF or secondary to the production of contrastive pitch accent. Results for gesture launch time are shown in Tables 22 and 23 and Figure 26. The time of gesture onset to gesture apex was influenced by several factors. As expected, the longest gesture launch times were observed for the trials spoken with DAF (*M*=834.52 ms, *SD*=98.64 ms) compared to the shortest gesture launch times noted for trials spoken without an auditory delay (*M*=789.61 ms, *SD*=63.82 ms). Yet, only one main effect was significant. Like total gesture time but unlike GA-VM, 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.

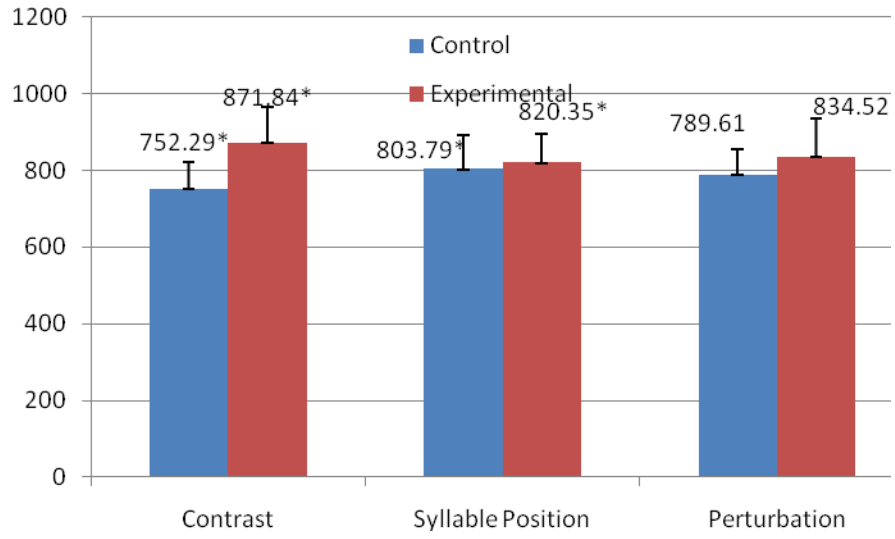


Figure 26 Gesture launch time (ms) for control and experimental conditions: contrast (neutral and contrastive accent), syllable position (first and second position), and perturbation (NAF and DAF). Error bars correspond to standard error. * Indicates significant main effect.

Even though there was a significant main effect of syllable position on total gesture time, this effect was not found for the initial segment of the total gesture (i.e., gesture launch). In fact, gesture launch times for first and second position syllables differed only by 17 ms on average. However, syllable position did interact with contrast to effect the duration of gesture launch time analogous to the dependent measure, total gesture time. There was a significant two-way interaction between *contrast* and *syllable position* [$F(1,14) = 5.910$, $p<.029$] (see Figure 26). The average time from gesture onset to gesture apex 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 Figure 27, 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.

Table 22 *Descriptive Data for Gesture Launch Time (ms) for Each Condition.*

Condition	Mean (ms)	Standard Deviation	Range
-PA/1st/NAF	734.11	218.79	469.67-1261.57
-PA/1st/DAF	792.66	423.86	424.39-2085.00
-PA/2nd/NAF	734.88	210.97	475.69-1135.56
-PA/2nd/DAF	747.52	261.81	448.79-1257.11
+PA/1st/NAF	815.55	287.04	507.86-1524.29
+PA/1st/DAF	872.82	443.53	507.00-2183.86
+PA/2nd/NAF	873.90	286.95	530.32-1577.78
+PA/2nd/DAF	925.09	427.55	535.67-2145.67

Table 23 *Analysis of Variance Summary Table for Gesture Launch Time; * Indicates Statistical Significance.*

Variable	*df*	*SS*	*MS*	*F*	*p-value*	η^2	*power*
Contrast	14	428759.617	428759.617	17.880	.001*	.561	.975
Position	14	8226.406	8226.406	.728	.408	.049	.125
Perturbation	14	60521.668	60521.668	1.202	.291	.079	.176
Con. x Posit.	14	45048.364	45048.364	5.910	.029*	.297	.619
Con. x Pert.	14	2605.854	2605.854	.939	.349	.063	.148
Posit. x Pert.	14	5069.546	5069.546	.785	.391	.053	.131
Cn x Pos x Pert	14	2974.495	2974.495	.663	.429	.045	.118

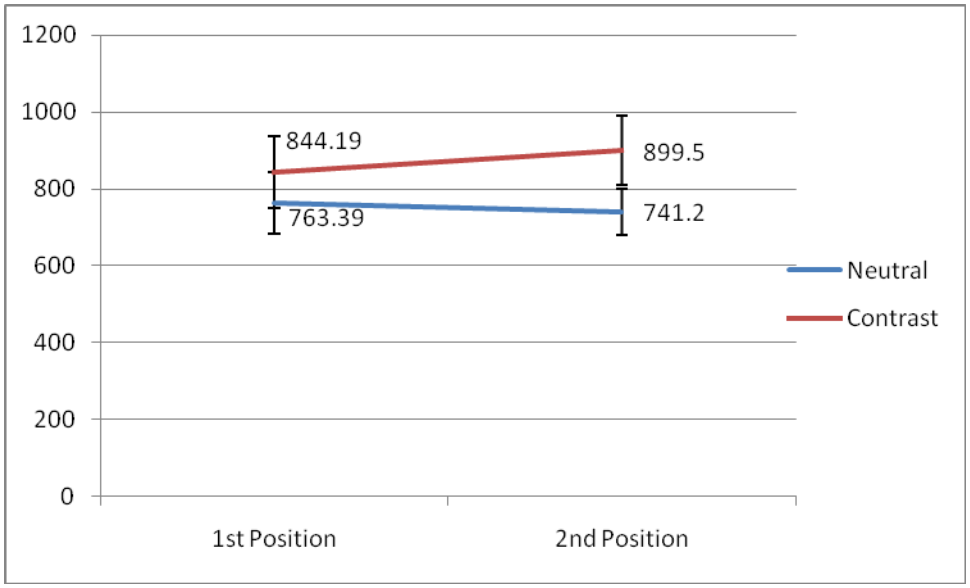


Figure 27 Gesture launch time (ms) indicating a significant two-way interaction of *position x contrast.* Error bars correspond to one standard deviation.

The significant two-way interaction of *contrast* x *position* was further analyzed utilizing a post hoc analysis of simple main effects using a the Bonferroni correction. Significance was adjusted to .025. As presented in Table 24, the effect of contrast was significant for both first [$t(30)=2.49$, $p<.009$] and second syllable position [$t(30)=4.88$, $p<.0000$]. 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.

Table 24 *Results of Post Hoc Simple Main Effects of Contrast x Position Interaction. df=Degrees of Freedom, t=Test Statistic, and p=Significance Level. * Indicates Statistical Significance*

Pairwise Comparisons for Syllable Position	***df***	***t***	***p***
Contrast versus neutral: first syllable	30	2.49	.009*
Contrast versus neutral: second syllable	30	4.88	.0000*

There were no other significant interactions though the gesture launch times across the eight possible conditions are noteworthy in several ways. Mean gesture launch time was longest for trials which were produced not only with contrastive pitch accent on the second syllable, but also with an auditory delay (*M*=925.09 ms, *SD*=427.55 ms). The shortest gesture launch times were noted for the trials produced with no contrastive accent and no auditory delay. There were two scenarios with no accent and no auditory delay, one with a first syllable comparison and the other with a second syllable comparison. The mean gesture launch times were 734.11 ms (*SD*=218.79 ms) and 734.88 ms (*SD*=210.97 ms), respectively. Thus without an influence of either DAF or contrastive pitch accent or both, there was notable consistency of the mean and variability of gesture launch time.

The primary motivation for measuring gesture launch time was to evaluate the potential effect of speech perturbation on the timing of the meaningful portion of the gesture. For this objective, H_0[6] failed to be rejected, though marginally significant and qualitatively longer gesture durations were noted for DAF trials. Alternatively, the analysis of gesture launch time offered insight on a significant relationship between contrastive pitch accent and the meaningful portion of the gesture such that gesture launch times were longer for pitch accented trials. This finding was most salient for trials produced with pitch accent on the second syllable of the target word.

4.2.8 Further Examination of Deictic Gesture Timing

A number of significant findings of Experiment 2 indicated changes in the gesture movements that co-varied with changes in the speech movements. Nonetheless, a number of inconsistencies and perplexing aspects of the data spurred the analysis of two additional variables, gesture launch midpoint to vowel midpoint (GLM-VM) interval and gesture return time. GLM-VM interval shares similarities to the measure GA-VM interval, though captures the midpoint of the launch period leading to the apex of the movement. 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. The prediction was that GLM-VM intervals would be smaller for first position target syllables and for syllables with contrastive pitch accent compared to their counterparts, second position syllables and neutral syllables, respectively. If speech and gesture were tightly entrained even when faced with perturbation, then no difference was expected to be found for the GLM-VM intervals for DAF versus NAF conditions.

Gesture return time was also calculated to further examine the data of Experiment 2. Gesture return is equal to the time from gesture apex to gesture return. Gesture return time was analyzed to provide a complete assessment of two components of the total gesture, gesture launch and gesture return. In particular, gesture return time was included to examine the discrepancy of a significant main effect of contrast and position for total gesture time, but only contrast for gesture launch. However, it is predicted that there will be no significant differences

for any of the three variables for gesture return because both the meaning and the spatio-temporal parameters of the gesture are specified within the launch portion of the gesture.

4.2.8.1 Gesture Launch Midpoint to Vowel Midpoint Interval The midpoint of the gesture launch (GLM) was calculated by dividing the gesture launch duration by two and adding that value to the time of gesture onset. Vowel midpoint was previously measured and calculated for the GA-VM interval measures. Therefore, no additional reliability measurements were necessary. A 2 (presence/absence of contrastive pitch accent) x 2 (1^{st} and 2^{nd} syllable position) x 2 (presence/absence of DAF) ANOVA was conducted on GLM-VM. The skewness for GLM-VM ranged between .316 and 1.115 (*M*=.819, *SD*=.33). Like the other three dependent measures of gesture-speech synchrony, the data were transformed by computing the base 10 logarithm of (x + 1) to increase normality. Unlike the other dependent measures, this transformation did change the results of the ANOVA. After log transformation, Shapiro-Wilk tests of normality were nonsignificant for all eight conditions, indicating an adequate normal distribution for subsequent analysis. As a consequence, the results presented for GLM-VM are from the transformed dataset.

Results were consistent with the predictions for the factors of syllable position and contrast. Like, GA-VM for Experiments 1 and 2, the interval between gesture launch midpoint and vowel midpoint was shorter on average for first position syllables (*M*=2.40, *SE*=.08) compared to second position syllables (*M*=2.67, *SE*=.08) as displayed in Tables 25 and 26 and displayed in Figure 28. Results indicated a significant main effect of syllable position [$F(1,14) = 5.301$, $p<.037$] (see Table 27). GLM-VM also was 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) as also shown in Figure 28. This main effect was also significant [*F*(1,14) = 27.848, *p*<.000] (see Table 27). A main effect of GA-VM was not noted in the results of this experiment but a main effect was established in the first experiment. However, the mean GA-VM intervals were actually longer for contrastive stressed syllables in Experiment 1.

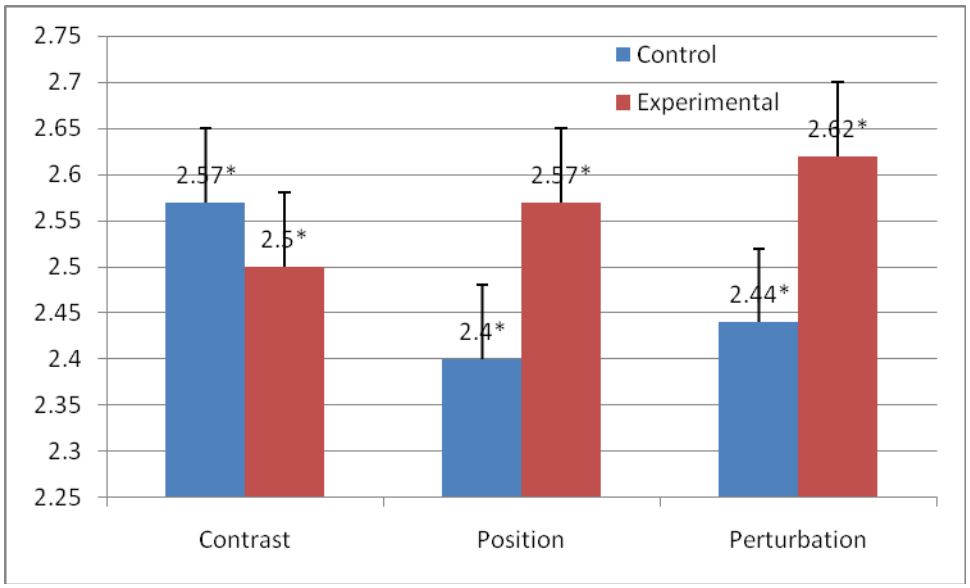


Figure 28 GLM-VM Intervals (ms) for control and experimental conditions: contrast (neutral and contrastive accent), syllable position (first and second position), and perturbation (NAF and DAF) from transformed dataset. Error bars correspond to standard error. * Indicates significant main effect.

An auditory delay resulted in longer GLM-VM intervals (*M*=2.62, *SE*=08) compared to when there was no auditory delay (*M*=2.44, *SE*=.07). This main effect was also significant [*F*(1,14) = 32.932, *p*<.000] . These results are also presented via Tables 26 and 27and Figure 28. This finding is consistent with the GA-VM interval results from both Experiment 1 and 2. There were no significant interactions.

Table 25 *Descriptive Data for GLM-VM Intervals (ms) for Each Condition: Log Transformed Data.*

Condition	Mean (ms)	Standard Deviation	Range
-PA/1st/NAF	2.35	0.30	1.93-2.83
-PA/1st/DAF	2.48	0.34	1.83-2.97
-PA/2nd/NAF	2.63	0.28	1.89-2.98
-PA/2nd/DAF	2.80	0.29	2.17-3.18
+PA/1st/NAF	2.27	0.33	1.74-2.78
+PA/1st/DAF	2.49	0.35	1.94-3.06
+PA/2nd/NAF	2.52	0.32	1.84-3.02
+PA/2nd/DAF	2.73	0.36	2.02-3.25

Table 26 *Analysis of Variance Summary Table for GLM-VM Intervals: Log Transformed Data. * Indicates Statistical Significance*

Variable	*df*	*SS*	*MS*	*F*	*p-value*	η^2	*power*
Contrast	14	.126	.126	5.301	.037*	.275	.573
Position	14	2.21	2.21	27.848	.000*	.665	.998
Perturbation	14	.966	.966	32.932	.000*	.702	1.000
Con. x Posit.	14	.023	.023	1.522	.238	.098	.210
Con. x Pert.	14	.028	.028	4.463	.053	.242	.503
Posit. x Pert.	14	.003	.003	.604	.450	.041	.112
Cn x Pos x Pert	14	.007	.007	1.873	.193	.118	.248

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 provides replication that the time between the gesture and target syllable becomes longer when speech is produced under the influence of an auditory delay.

4.2.8.2 Gesture Return Time Gesture return time is equal to the time between gesture apex and gesture offset. Hence, gesture return time plus gesture launch time is equal to the total gesture time for each individual trial. A base 10 logarithm of (x + 1) transformation was performed on the data with no change in the results. Also, three extreme outliers were identified

via stem and leaf plots generated by SPSS 16.0 software. Analyses were performed a second time with the extreme outliers removed. Again, there was no change in the results of the three-way ANOVA. Therefore, data are presented for all fifteen participants and on the original dataset prior to transformation.

As predicted, there were no significant differences between control and experimental conditions for contrast, syllable position, or perturbation (see Figure 29 and Tables 27 and 28). There were also no significant interaction effects. Gesture return times were longer for trials produced with DAF (M=774.87 ms, SE=66.05 ms) than for trials without an auditory delay (M=710.87 ms, SD=44.21 ms). In fact, gesture return times were the longest for trials produced with DAF and contrastive stress on the second syllable (M=802.35 ms, SD=275.33 ms). Although this finding was not significant, it is the same as the descriptive results for gesture launch time.

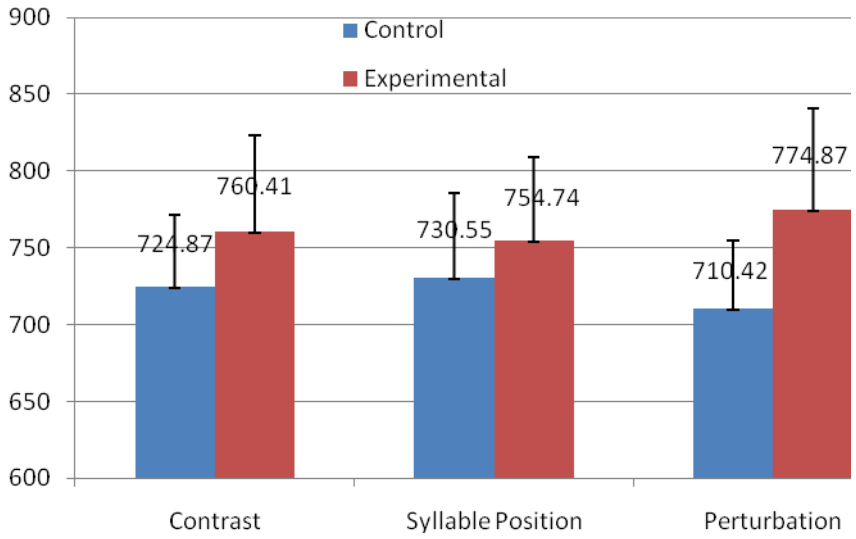


Figure 29 Gesture return time (ms) for control and experimental conditions: contrast (neutral and contrastive accent), syllable position (first and second position), and perturbation (NAF and DAF). Error bars correspond to standard error. There were no significant main effects.

Table 27 *Gesture Return Time (ms) for Each Condition.*

Condition	Mean (ms)	Standard Deviation	Range
-PA/1st/NAF	674.29	135.24	449.67-954.13
-PA/1st/DAF	750.98	251.85	494.60-1409.80
-PA/2nd/NAF	711.02	169.80	535.53-1115.33
-PA/2nd/DAF	763.19	239.53	501.27-1338.54
+PA/1st/NAF	713.95	207.43	329.86-1172.10
+PA/1st/DAF	782.96	337.36	475.41-1785.44
+PA/2nd/NAF	742.40	202.19	489.00-1185.58
+PA/2nd/DAF	802.35	275.33	534.13-1589.37

Table 28 *Analysis of Variance Summary Table for Gesture Return Time. No Significant Differences Noted.*

Variable	*df*	*SS*	*MS*	*F*	*p-value*	η^2	*power*
Contrast	14	37903.492	37903.492	1.521	.238	.098	.210
Position	14	17560.094	17560.094	1.942	.185	.122	.55
Perturbation	14	124641.833	124641.833	3.919	.068	.219	.454
Con. x Posit.	14	2.284	2.284	.000	.988	.000	.050
Con. x Pert.	14	.019	.019	.000	.999	.000	.050
Posit. x Pert.	14	2115.972	2115.972	1.319	.270	.086	.188
Cn x Pos x Pert	14	448.469	448.469	.092	.766	.007	.059

4.2.9 Summary

This experiment tested a number of hypotheses. As predicted, both sentence durations and vowel durations were significantly longer for DAF trials compared to NAF trials. Vowel durations were also significantly greater for syllables produced with contrastive pitch accent than when they were produced with no pitch accent, as well as for second position syllables compared to first position syllables.

Contrary to expectations and the results of Experiment 1, GA-VM was not affected by the presence of contrastive stress or first syllable position. However, the GA-VM interval was affected by DAF given that there was significantly greater asynchrony as measured by GA-VM intervals for responses produced with an auditory delay. An interaction of syllable position and speech perturbation was found as a result of a greater difference in GA-VM interval for NAF and DAF conditions for a second position syllable compared to the first position syllable. In summary, the time between a gesture apex and vowel midpoint did not decrease for stressed syllables or across syllable positions, though it did increase for trials produced with an auditory delay.

The remaining hypotheses corresponded to changes in the duration of the gesture that may result from elongation of the spoken response secondary to DAF. Total gesture time and gesture launch time were longer for DAF trials than NAF trials, though this finding did not reach significance for either dependent measure. The average times to execute a complete gesture and a gesture launch were significantly longer for utterances produced with contrastive pitch accent. Total gesture time was also significantly longer for second position syllables than first position syllables, though gesture launch time was not. A significant interaction effect for both total gesture time and gesture launch time was detected for *contrast* x *position*. For both measures,

the time to complete the gesture segment of interest was longest for trials produced with contrastive pitch accent on the second syllable.

Two additional analyses were conducted to further explore the temporal relationship of speech and gesture and explicate the incongruent findings, particularly for GA-VM interval. The findings for GLM-VM interval for the factor of syllable position were similar to the GA-VM interval results from Experiment 1, even though the GA-VM interval results from Experiment 2 were not. Like the findings for GA-VM interval in the first experiment, gesture launch midpoint was more synchronized to vowel midpoint for first position syllables than second position syllables. Also like the results for GA-VM interval for the first experiment, there was a main effect of contrast for GLM-VM interval in this experiment, though in the opposite and in this case, expected direction. On average, gesture launch midpoints were more synchronized with vowel midpoints of target syllables produced with contrastive pitch accent than with neutral stress. GLM-VM intervals were shortest in duration for trials produced with contrastive pitch accent on the first syllable without the influence of an auditory delay.

Taken together, the results demonstrate that the temporal parameters of speech are altered by the presence of contrastive pitch accent and an auditory delay of 200 ms. Likewise, the presence of pitch accent, speech perturbation, and position of the target syllable affected the temporal parameters of the affiliated deictic gesture. Total gesture times and gesture launch times were longer for trials produced with pitch accent than for those same responses without pitch accent. The gestures were longest when the accent is placed on the second syllables, which were the longest in duration and also produced later than first syllable targets. Subsequently, greater synchrony was noted for contrastive stressed syllables as measured by GLM-VM interval. Perturbing speech by way of DAF also elongated the interval between the gesture and

speech stream as measured by both GA-VM interval and GLM-VM interval. Perturbing and lengthening speech also resulted in the increase of total gesture time, gesture launch time, and gesture return time, albeit not significantly. These findings are enumerated further in the Discussion section that follows.

4.3 DISCUSSION OF EXPERIMENT 2

The aim of Experiment 2 was not only to test whether speech and gesture co-occur in time, but also to explore the underlying mechanism of the supp osed speech-gesture synchronization. The objectives of Experiment 1 also were addressed in this experiment. The effects of contrastive pitch accent (present versus absent) and syllable position (first versus second) on the relative timing of deictic gestures were investigated once again, although additional dependent measures in Experiment 2 provided an opportunity to expand the examination of the role of these variables on speech-gesture synchrony. The other primary distinction between the first and second experiment of this investigation was the manipulation of auditory feedback during the participants' responses. The participants heard their responses amplified via headphones for all trials, but half of the trials were delayed by 200 ms while the other trials were played in real-time.

4.3.1 Vowel and Sentence Durations

Results indicated that individuals make changes to the duration of both vowels and sentences as a result of DAF and contrastive pitch accent. These findings not only validated the procedures of the current experiment but lend support to previous studies of the effects of DAF and pitch accent on acoustic duration measures. Though DAF is a long-studied and widely employed methodology (e.g., Elman, 1983), there are very few studies that have looked at the effects of DAF at the segmental, rather than the utterance level for typical fluent speakers. To the author's knowledge, this study stands alone in identifying that individuals made adjustments to individual vowels within utterances.

The participants also consistently lengthened pitch accented syllables, especially in the second syllable position (*M*=275.59 ms), and even to a greater degree when isolating second position accented syllables produced with DAF (*M*=307.15 ms), compared to first position syllables with pitch accent which averaged 190.38 ms in duration. It has been well-established that though pitch accent is primarily thought to affect the fundamental frequency contour of a syllable and vowel, the duration of the syllable, rime, and vowel is likely to transform as well (e.g., van Kuijk & Boves, 1999). This research adds to that literature and highlights the complex nature of a variety of prosodic prominence units. The data also demonstrate that individuals may over-compensate when asked to place contrastive stress on a normally unstressed vowel, as with the bisyllabic compound word pairs which typically would be produced with lexical stress on the first syllable. In other words, participants were familiar with and automatically assigned lexical stress and even contrastive pitch accent on the first syllable of these trochaic lexical items. When asked to place emphasis in an unusual and non-automatic manner on the second position syllable, participants produced relatively greater prosodic prominence as demonstrated by greatly

increased vowel durations within these second position syllables with contrastive pitch accent. The possible implications of this unbalanced prominence assignment across first and second syllables will be discussed further in a later section.

4.3.2 Effects of Speech Perturbation

The major prediction of this second experiment that was *not* upheld was that speech and gesture would remain synchronized even when speech was perturbed via DAF. In fact, the opposite finding was observed. Speech and gestures were more *asynchronous* when produced under the influence of DAF. However, there are indications that the gesture did temporally unfold differently during DAF trials compared to their NAF counterparts. Qualitative changes in total gesture time, gesture launch time, and gesture return time suggest that there is potentially some interaction and feedback between the two motor systems.

If one only considers the finding that the intervals between gesture apices or gesture launch midpoints with vowel midpoints increases with the addition of DAF, then it would seem that de Ruiter's Sketch Model (1998; 2000) is supported (see Figure 30). As stated in the Discussion section for Experiment 1, it is possible that the formulation, planning, and programming of a gesture and its lexical affiliate are initiated simultaneously in the conceptualizer, though increased time is necessary preceding the production of speech because of increased complexities of movement planning and execution. A deictic gesture requires even less planning and programming compared to iconic gestures according to the Sketch Model. Specifically a deictic gesture is initiated in the conceptualizer along with the target word. Then the preverbal

message, coded for the "representation of the content of speech" (de Ruiter, 2000, p. 289), is sent to the formulator and the sketch, coded for the "spatio-temporal representations" of the gesture is sent to the gesture planner. There is where the interaction between the two systems ceases. The sketch holds information regarding the direction of the referent. Once sent to the gesture planner, a motor program for the deictic gesture is created by way of accessing the motor program template for the fairly consistent, conventional hand configuration of a deictic gesture. This is a much more simplistic process not only compared to the formulation of speech within the formulator, but also compared to the planning of iconic gestures which do not have a one-to-one template in the gestuary. That is to say, there are unique spatio-temporal representations that need to be incorporated for each individual iconic gesture.

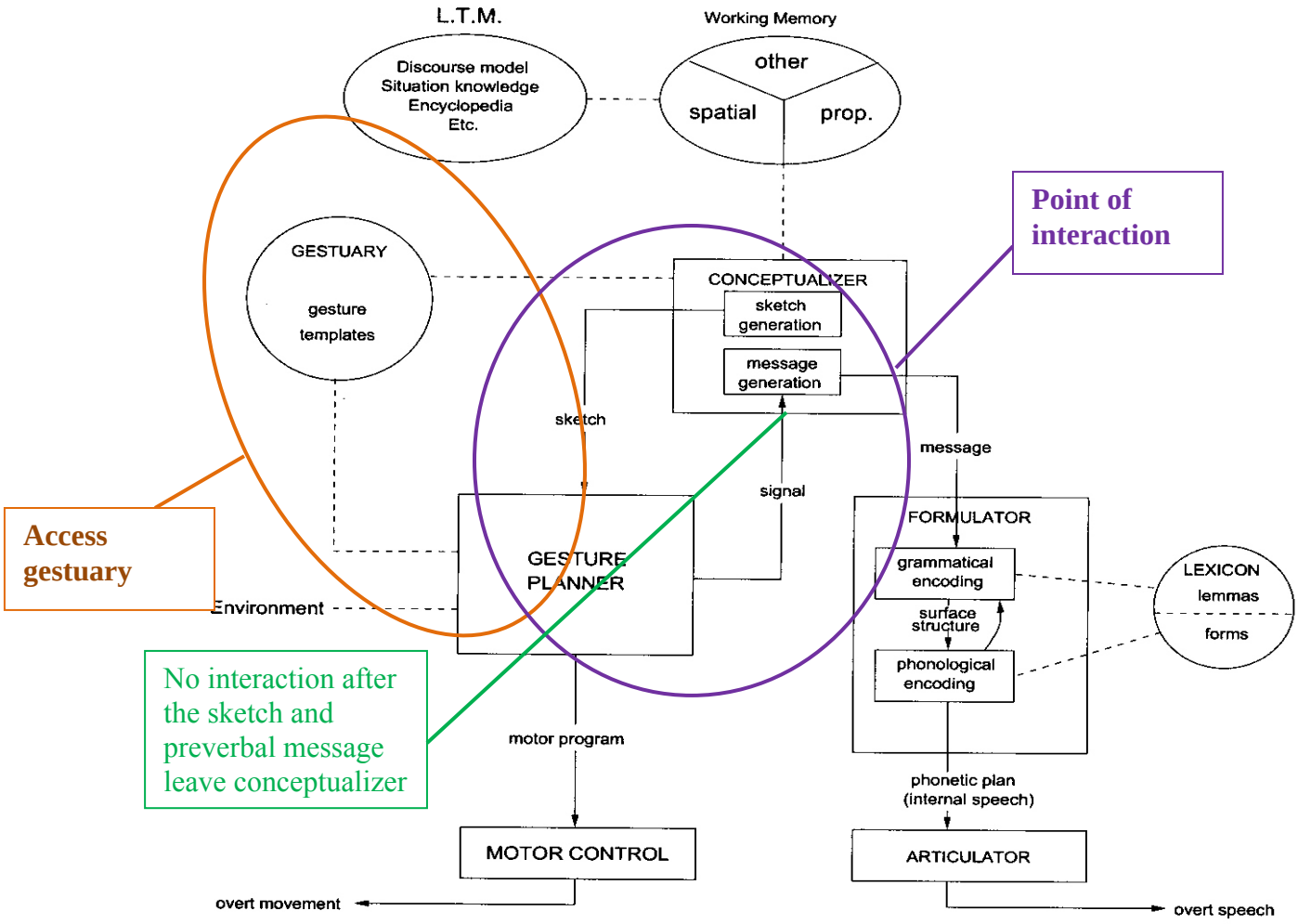


Figure 30 Specifications of speech and gesture interactions proposed by the Sketch Model. Adapated from "Gesture and Speech Production," by J.P. de Ruiter, 1998, Unpublished doctoral dissertation, Katholieke Universiteit, Nijmegen, Germany, p. 16.

If gesture and speech cease interaction before the level of the formulator, then neither prosodic stress nor speech perturbation should 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 than thought when initiated together in the conceptualizer. As a result, the

interval between a gesture apex and/or 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 three points to be made that may challenge a strict adherence to the Sketch Model that are explained below.

First, it is possible that there is indeed interaction between the two systems below a point of conceptualization but that the planning of a deictic gesture is unique in its level of convention and automatic planning, particularly in the recurring task requirement of this experiment. Gesture onsets are consistently found to occur up to a second before the onsets of the associated spoken lexical item (e.g., Morrel-Samuels & Krauss, 1992). A deictic gesture motor program is arguably simpler and more rote than an arbitrary and idiosyncratic iconic gesture. Therefore, a deictic gesture may be less susceptible to changes in its execution, in this case temporal changes. An iconic gesture may be more likely to exhibit changes in the gesture launch period, as well as pre-stroke, post-stroke, or apex holds to maintain synchrony.

This account could also reconcile the findings between this experiment and McNeill's (1992) exploratory study of the effect of DAF on gesture production. In contrast to the controlled paradigm and use of deictic gestures of Experiment 2, McNeill first had individuals spontaneously narrate a cartoon. McNeill's qualitative observations included an increase in the amount of gesture and that "gesture is still synchronous with the coexpressive part of the utterance" (p. 275). Interestingly, McNeill did not observe the same synchrony, though vaguely described in the first of his experiments, in a second protocol. In the second experiment, participants were trained to recite utterances from memory while executing a series of fluid, continuous, iconic gestures. Even though the gestures were iconic, they were perhaps accessed differently from the gestuary 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 more similar to the current study, so too were the results. In contrast to McNeill's first study in which the spontaneous gestures remained perceptually synchronized with the spoken targets, there was noted asynchrony in the second experiment. In fact, gestures were noted to precede the spoken targets just as in the current experiment. Again, the dissociation between spontaneous and prescribed gestures may explain the differences between the findings of the present experiment and Mayberry and colleagues' (Mayberry & Jaques, 2000; Mayberry, et al., 1998) work that described a cessation and subsequent synchronization for children and adults with fluency disorders while spontaneously describing a cartoon. In brief, the availability of feedback to and from the speech system may be different for automatic, conventional gestures compared to individually specified, spontaneous gestures. Future investigations that employ a more natural, spontaneous production paradigm and different gestures types could further examine these disparate findings.

Another possibility is that DAF was not the appropriate method of perturbation. 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 relatively wide range of mean durational difference between NAF and DAF trials, 120 to 2346 ms. Therefore, it is possible that individuals with greater speech disruptions when faced with DAF may also be differentially susceptible to gesture disruptions.

Moreover, DAF is a relatively slow and also continuous perturbation method. Frequency shifted feedback is another auditory feedback tool to used to alter the speech individuals with fluency disorders (e.g., Howell and Archer, 1984) and individuals with Parkinson's disease (Brendel, Lowit, & Howell, 2009). Frequency shifted feedback results in reduced speech rates

for speakers with communication disorders and typical speakers (Tourville, Reilly, & Guenther, 2008). However, it is not clear if the rate of susceptibility of frequency shifted feedback and delayed auditory feedback is the same across individuals. It is also not known if the temporal changes are similar across the two feedback methods. Thus, the effect of frequency shifted feedback upon the timing of gestures remains an empirical question.

Additionally, the continuous feedback provided by DAF and frequency shifted feedback may not be appropriate for studying the effect of speech perturbation on gesture timing since the feedback mechanism shared between the gesture and speech systems may not be continuous as suggested by Levelt et al. (1985). Levelt and colleagues imposed a load to the wrists of their Dutch-speaking participants during a deictic gesture and simple spoken response task. In short, the participants pointed to a light while stating, *this light* or *that light*. Speech and gesture were synchronized when no load was applied and also when the load was applied halfway through the gesture. However, voice onset times were longer when a load was applied to the wrist at the onset of the gesture. A controlled, automatic response was employed in this study conducted by Levelt et al. and once more, there was limited feedback and ensuing changes in the execution of one motor behavior based upon perturbation of the other. 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 & Gracco, 1988).

Accordingly, there are similar ways to perturb speech in a precise, fleeting moment during speech production rather than a continuous, relative slow auditory closed-loop of feedback. It is

possible to impose transient mechanical perturbations of the mandible (e.g., Shaiman, 1989) and lower lip (e.g., Abbs and Gracco, 1984). Usually, this literature is interpreted as evidence for the remarkable adaptability of the motor system, in this case the speech motor system as well as the dynamic coordination of structures. Such methodology could be adapted to not only study the coupling of two articulators like the mandible and tongue, but two internally coordinated structures like the jaw and hands during spoken language production. Rochet-Capellan et al.'s (2008) findings that jaw aperture was synchronized with pointing gestures further support the use of kinematic methodologies to study the entrainment of speech and gesture. Future investigations also could modify the existing rhythmic speech task constructed by Cummins and Port (1998) to systematically study the effects of perturbation of speech and/or gesture upon the other system. During Cummins and Port's investigation, individuals were required to recite the phrases *big for a duck* or *geese for a duke* along with metronome-like beeps. One could easily create iconic gestures to mirror the lexical response and ask the individuals either to produce the gestures and speech at a habitual rate and/or along with the metronome. Transient perturbations to the mandible could be employed and the resultant disruptions upon the gesture and also speech system measured.

Perturbation of gestures also can lend insight on the interactive nature of speech and gesture. Levelt and others (1985) are the only researchers to date to perturb a gesture, in their case by imposing a load on the wrist while pointing. A replication of this work would certainly be worthwhile, though there are other designs and methodologies that could address this question as well. One methodology for studying deictic gestures specifically that stands out as both valid and feasible is the double-step perturbation paradigm (e.g., Prablanc & Martin, 1992; Bard, Turrell, Fleury, Teasdale, Lamarre, & Martin, 1999). This paradigm is used to examine goal-

directed movement, such as reaching and ocular gaze. 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 (e.g., Prablanc & Martin, 1992). 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 of varying conditions. For example, the time of target change could differ between no shift of target, early shift, and later shift, similar to the perturbation times used by Levelt and colleagues.

In short, it is certainly feasible to construct paradigms that employ a constrained task with other gesture types and other speech and/or gesture perturbation methods. Future investigations may also aim to examine the effects of speech perturbation and/or gesture perturbation in a less-constrained task such as a story retell, passage reading, cartoon narration, or even spontaneous conversation.

Finally, a third argument for against strict adherence to the Sketch Model is the durational changes of deictic gestures when produced in DAF conditions. There were increases in overall duration of the gesture, gesture launch time, and gesture return time for DAF trials. Indeed the difference in duration for NAF and DAF conditions were not significant for these three dependent measures, but they were consistent. It is also worth noting that that a larger sample size may have been necessary to detect the effect of DAF upon the duration of gesture. The initial sample size was based upon data that corresponded to the effect of contrastive pitch accent, not speech perturbation via an auditory delay. The effect size of DAF for total gesture time was d=0.38, a small to medium effect size. The estimated effect size for this series of experiments was d=0.76 based upon de Ruiter's data for the effect of contrastive stress on

gesture launch time. If instead, 0.38 is used as the estimate of effect size, power set at 0.80 and alpha set at 0.05 the required sample size increases to 57 participants. Retrospectively, this difference is logical because the increase in vowel duration associated with contrastive pitch accent is likely a steadier occurrence than the anticipated variable temporal effects of auditory delays across and within participants. Hence, future investigations employing DAF should consider utilizing a small to medium effect size when calculating sample size rather than medium to large effect size.

4.3.3 Effects of Syllable Position and Contrastive Pitch Accent

The effects of syllable position and contrastive pitch accent will be discussed concurrently given that they not only were studied in Experiment 1, but also because many of the effects of these two factors interacted in the second experiment. Syllable position affected the timing of gesture in several ways. As in Experiment 1 and as predicted, greater synchrony was noted for first position syllables compared to second position syllables. However, this effect was not found for the measure of GA-VM interval as in Experiment 1, but was found for the measure of GLM-VM interval. Only one other main effect of syllable position was observed; gestures were longer when produced with second syllable position trials than first syllable position trials.

A number of significant interactions emerged for syllable position, with second position syllables seeming to affect the timing of gesture to a greater extent than first syllable positions. The execution of a complete gesture movement and the movement from gesture onset to apex required increased time for trials produced with contrastive pitch accent on the second syllable of

the target word. Even though gestures seemed to extend their temporal envelope to match the later production of the prosodically prominent second syllable, greater asynchrony was noted for second position syllables produced with DAF according to the intervals between gesture apices and vowel midpoints.

The timing of gestures changed as a function of contrastive pitch accent, though there was no effect of contrastive pitch accent for GA-VM interval in this experiment as there was in the preceding study. The data from Experiment 1 demonstrated that there were actually significantly shorter GA-VM intervals for neutral trials compared to contrastive trials. This finding was perplexing in the first experiment and led to the suggestion that the gesture apex measure may not be capturing the pulse of movement, but rather the end point of the pulse. Thus, the measurement of gesture launch time and gesture launch midpoint to vowel midpoint intervals was motivated in the Discussion section of the first experiment.

Interestingly, there was indeed 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. In fact, review of the descriptive data for the transformed dataset showed that greatest synchrony was noted for first position syllables produced with pitch accent without auditory delay (M=2.27, SE=.084) while the least synchrony was noted for second position syllables produced with neutral stress and DAF (M=2.71, SE=.072). These main effects for GLM-VM interval are in line with the alternative hypotheses stated for the originally presented GA-VM dependent variable.

Not only were gesture launch midpoints more synchronous with stressed vowel midpoints than neutral vowel midpoints, but also gesture launch times were significantly longer on average

for trials with contrastive pitch accent. Gesture launch times were longest for accented second position syllables. Total gesture time was also significantly longer for second position syllables with contrastive accent, but not for accented first position syllables. Increased gesture time demonstrates 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 longest and latest 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.

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 & Thelen (1999). If there were no entrainment and the motor processes proceeded according to the Sketch Model (de Ruiter, 1998, 2000), 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, Harris, and Kelso (1982) 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.

To conclude, 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 eyeblinks, head movements, and even torso movements are as well (Birdwhistell, 1952; 1970; Bull & Connelly, 1985; Loehr, 2004; 2007). If there was a unique and absolute cognitive-linguistic relationship responsible for gesture-speech synchrony as posited by theorists like McNeill (1992), de Ruiter (1998; 2000) and Krauss, Chen, & Gottesman (2000), then such observations would be difficult to explain.

Though Levelt (1989) is a well-specified and testable model of spoken language production, an integrative model of speech motor control such as proposed by Ballard, Robin, and Folkins (2003) which supports the coordination of motor behaviors across systems may frame studies of oral/manual interactions. Ballard et al. state that, "neurological and evolutionary evidence strongly suggest that neural networks are large, flexible, multifaceted, multifunctional, and overlapping in function" (p. 46). Though this model is largely unspecified, the fundamental tenets of the authors' postulations are certainly consistent with the rationale and findings of Experiment 2. Fusing the foundation of Ballard et al.'s integrative model of speech motor control and the specifications of a dynamic systems approach to understanding the coordination and entrainment of speech and gesture (e.g., Kelso & Tuller, 1984; Iverson & Thelen, 1999; Tuite, 1993) will be a source for future theoretical consideration and subsequent empirical testing.

5.0 CONCLUSIONS

5.1 LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH

This investigation offered insight on the temporal synchronization of speech and gestures. The three variables manipulated, syllable position, contrastive pitch accent, and speech perturbation distinctively affected the temporal characteristics of the spoken response and corresponding deictic gesture. Evidence for entrainment of the two motor systems was demonstrated, though synchronization of the two systems actually decreased when speech was perturbed. The methodology employed was unique in the level of experimental control, stimuli, and data collection procedures relative to other studies of the role of prosodic stress and speech perturbation on the synchronization of speech and gesture. Also, the timing of speech and gesture was studied using longer utterances and more natural responses than similar controlled research paradigms that examined the effect of prosodic stress on speech-gesture synchronization, and the only experimental study on the topic to enroll English-speaking individuals. This was the first systematic investigation that explored the effect of speech perturbation upon the synchrony of manual and speech movements.

The results of the study are intriguing, though also reveal several limitations of the current work and motivate extensions of this research to future empirical study of the interaction of speech and gesture systems. The limitations of the study primarily center on the measurement

of the speech and gesture signal and the required response. Some of the limitations were discussed in previous sections as well (e.g., type of gesture and perturbation method).

As was stated in the rationale for the first study, there is a multitude of ways to measure prosodic stress. Increased vowel duration was chosen as the acoustic correlate of contrastive pitch accent because of the relatively consistent and robust changes as a function of prosodic stress. However, there are other acoustic and physiologic correlates that may better expose the temporal relationship between speech and gesture. The current data support the measurement of the stroke (i.e., launch) portion of the gesture as the pulse of the oscillator, not the arguable end point of the gestural stroke. Gesture launch midpoint was analyzed for the second study with distinctly different results compared to time of gesture apex. However, the voltage traces generated by the capacitance sensor of the theremin can also be analyzed to determine the peak velocity of manual movement. Reaching studies have demonstrated that velocity decreases as an arm/hand approaches the target (e.g., Wu, Trombly, Lin, & Tickle-Degnen, 2000). It is plausible that the velocity of arm/hand movement also slows as one gets closer to the apex of a gesture. Therefore, the time of peak velocity may correspond to the point of maximum motoric effort that overlays the pulse of movement.

The optimal measure of the oscillator pulse within the speech stream is also 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 these results and those of Rochet-Cappellan et al. (2008), the speech and gesture system appear to share dynamic motor linkages. Rochet-Cappellan and colleagues kinematically studied the synchronization of finger movements and mandibular movements in during the repetition of *puhpuh* or *muhmuh* with prominence on one syllable while pointing to a smiley face. The

investigators were able to make precise measurements of timing of the jaw and finger by capturing the movement infrared-emitting diodes via an Optotrak system and found predictable relationships between jaw opening and pointing. Future investigations could utilize an optical tracking system or ideally a multiple camera high-speech optical motion capture system to better understand the entrainment of speech with other motor systems, including the hands. For instance, one could approach the logical next step in this line of research is to examine the phase transitions of oscillating speech and manual movements. Using an optical motion capture system would also allow tracking of the movements of not just hand and finger movements that accompany speech, but also facial movements to test the predictions of Tuite's Rhythmical Pulse model and experimentally replicate the descriptive and perceptual findings from previous studies (Birdwhistell, 1952; 1970; Bull & Connelly, 1985; Loehr, 2004; 2007). Sophisticated equipment like an optical motion capture system would also allow data to be collected on multiple gesture types within a three-dimensional space. To date, there has been no experimental investigation of the synchronization of speech and gestures that did not elicit deictic gestures. As discussed earlier, there is reason to conjecture that the interaction of speech and manual movements may differ based upon the type of gesture and automaticity of the task.

Indeed the present experiments employed a controlled protocol, though the responses were longer and more natural than prior research of this type. Yet, there are several limitations of the responses. First is an issue of time resolution. Significant effects of syllable position were found in the dataset, though manipulation of prosodic stress across different points in an utterance may lead to a better understanding of speech and gesture synchronization. An example of a paradigm that could be modified for American English speakers is Krahmer and Swerts

(2007) repetitive production of *Amanda went to Malta* with stress placed on an earlier or later word within the utterance.

A second limitation of the responses of Experiments 1 and 2 is the production of *yes* or *no*. Prominence could be placed upon this first word of the sentence and a pause inserted between *yes/no* and the next word of the response, *the*. Thus, the words *yes* or *no* could act as a separate intonational phrase and the prominence of the spoken item could act as a periodic attractor, rather than the target accented syllable within the compound word.

The dynamic coordination of speech and manual movements have been studied, but few studies have gone beyond examining the temporal parameters of tightly constrained vocal productions and co-occurring finger tapping. Past research has revealed a tight coupling of finger tapping with taps simultaneously produced with single, repeated syllables (e.g., *stak stak stak…)* (Chang & Hammond, 1987; Smith, McFarland, & Weber, 1986). This research also showed that when the amplitude and frequency of one movement was increased, the same parameters of the associated movement were increased as well. An extension of this paradigm that is relevant to the current study was completed by Hiscock and Chipuer (1986). Individuals were required to tap and iambic or trochaic rhythm with their right or left hand while reciting one of four, ten-syllable utterances. Two of the utterances exhibited an iambic rhythm (e.g., *the cause of crime eludes the brightest minds*) and two exhibited an irregular rhythm (e.g., *the Vancouver summer is delightful*). The rhythms were the same in one condition and mismatched in others. Although both utterances similarly decreased the rate of tapping, only the mismatch rhythm condition resulted in significant disruption of the tapping rhythm. Implementations of these investigations of in reference to speech and gesture movements, rather than speech and

finger tapping movements would help to elucidate the coordination of these two internal coupled oscillators.

Concepts such as periodic attractors, coupled oscillators, and temporal entrainment are embodied within dynamic systems theory. 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 systems, namely oral and manual movements, is blossoming, but still only its provenance.

5.2 THEORETICAL IMPLICATIONS

Iverson and Thelen (1999) and Tuite (1993) are the only investigators to formulate theories of speech and gesture production from a dynamic systems perspective. Both theories, the Entrained Systems theory (Iverson & Thelen) and Rhythmical Pulse Model (Tuite) assert that rhythm is the underlying mechanism of interaction and overlay of the two systems. Iverson and Thelen's work reflects early developmental processes within the first 18 months of life and posits that the degree of coupling of speech and gesture is dependent the level of effort and automaticity of the respective motor behaviors (also see Iverson & Fagan, 2004 and Iverson, 2010). Even though they do not elaborate upon the potential attractors, they do specify that it is the gestural stroke that is synchronized with the lexical affiliate.

In contrast, Tuite does not make any statements about developmental progression of gesture and speech production, but does hypothesize that a kinesic base unites the speech and gesture in time by way of coupling pulses within the two systems. He further hypothesizes that the pulse peak corresponds to the stroke portion of the gesture (or peak of another movement like an eyebrow raise) and intonation peaks of the prominent syllables in the speech stream. Yet, he stops short of explaining the actual origin of a kinesic base and the details of the corresponding pulses or exactly what the units of the pulse peaks are for speech.

More recent theorizing by Port (2003) may not address gesture specifically, but does spell-out the idea of coupling of pulses for external and internal oscillators. Importantly for the current investigation, Port conjectures that prosodic structure, as realized in vowels, is the periodic pattern that organizes not just speech, but the coordination of multiple modalities with the speech system. He states that they pulses are associated with neurocognitive oscillations that arise from "major neural patterns somewhere in the brain" (p. 609).

Amalgamation of the theoretical work by Iverson and Thelen (1999), Tuite (1993), and Port will offer 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. As the strength and stability of one motor behavior increases, so will the likelihood that two systems will entrain.
2. The entrainment and subsequent synchronization of speech and gesture is dependent upon prominent neurocognitive pulses that act as periodic attractors within two rhythmic systems.

3. A neurocognitive pulse for a gesture system corresponds to the purposeful stroke and vowel with the greatest motoric effort within a unit of speech.
4. Variability of synchrony will be least for the strongest attractors.
5. Neurocognitive pulses can be coordinated between multiple internal oscillators manifested in oral, manual, facial, and other body movements.

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

APPENDIX A

RECRUITMENT FLYER

PARTICIPANTS NEEDED

Adults between the ages of 18 and 40 are needed to participate in a research study looking at the production of speech and gestures.

Requirements:

- **Right-handed**
- **English spoken as primary language**
- **No history of speech, language, or hearing problems**

Participation will require one 90-minute visit to

Forbes Tower at the University of Pittsburgh

For more information, please call or Email

Heather Rusiewicz

412-396-4205

hrusiewicz@gmail.com

APPENDIX B

SCREENING QUESTIONS

Date of Birth:_____ Today's Date:_____ Age (years):_____

.	Are you fluent in any language other than English? If you speak another language other than English, how often do you speak in this language	YES NO Daily Weekly Monthly Rarely
.	Were you ever diagnosed with a speech or language disorder (e.g., articulation/phonological disorder, stuttering, etc.)?	YES NO
.	Were you ever diagnosed with a neurological disorder (e.g., epilepsy, cerebral palsy, etc.)?	YES NO
.	Do you have normal or corrected-to-normal vision?	YES NO
.	Which hand do you use to perform most one-handed tasks such as writing?	RIGHT LEFT
.	Please describe your ethnic/racial background (circle all that apply).	Caucasian African American Hispanic Asian Pacific Islander Other:_____
.	Please list any previous locations you have resided for one year or more other than Pittsburgh/Western Pennsylvania.	1._____ 2._____ 3._____

APPENDIX C

STIMULI SET 1; SHARE FIRST SYLLABLE, CONTRASTIVE STRESS ON SECOND SYLLABLE

bathrobe	bathtub
football	footprint
grapevine	grapefruit
icecube	icecream
lifeboat	lifeguard
blackbird	blackboard
toothpaste	toothbrush
thumbtack	thumbprint
birdcage	birdbath
lighthouse	lightbulb
eggnog	eggplant
fishbowl	fishhook
seagull	seahorse
snowball	snowflake
teapot	teacup

APPENDIX D

STIMULI SET 2; SHARE SECOND SYLLABLE, CONTRASTIVE STRESS ON FIRST SYLLABLE

suitcase	briefcase
jukebox	mailbox
birdhouse	doghouse
toothbrush	paintbrush
football	baseball
bluebird	blackbird
lighthouse	greenhouse
cupcake	pancake
stoplight	flashlight
footprint	handprint
keyboard	surfboard
notebook	matchbook
wheelchair	highchair
horseshoe	snowshoe
hottub	bathtub

APPENDIX E

EXAMPLE STIMULI ILLUSTRATIONS

6.0 BIBLIOGRAPHY

Abercrombie, D. (1967). Paralanguage. *International Journal of Language and Communication Disorders*, *3*, 55-59.

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. *Journal of Speech and Hearing Research, 19*, 19-35.

Abbs, J.H. & Gracco, V.L. (1982). Evidence for speech muscle functional compartmentalization: Theoretical and methodological implications. *The Journal of the Acoustical Society of America, 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. *Journal of Neurophysiology, 51*, 705-723.

Acredolo, L. & Goodwyn, S. (1985). Symbolic gesturing in language development. *Human Development, 28,* 40-49.

Acredolo, L. & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development, 59,* 450-466.

Acredolo, L. P., & Goodwyn, S. W. (1990). Sign language in babies: The significance of symbolic gesturing for understanding language development. In R. Vasta (Ed.), *Annals of Child Development, Vol. 7* (pp. 1-42). London: Jessica Kingsley Publishers.

Acredolo, L.P. & Goodwyn, S. (2002) *Baby Signs: How to Talk with Your Baby Before Your Baby Can Talk.* New York, New York: McGraw-Hill.

Adams, C. & Munro, RR. (1978). In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. *Phonetica, 35,* 125-56.

Arbib, M.A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28,* 105-124.

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition, 48,* 66-79.

Attwood, A., Frith, U. & Hermelin B. (1988). The understanding and use of interpersonal gestures by autistic and Down's syndrome children. *Journal of Autism and Developmental Disorders,18*, 247-251.

Avikainen, S., Forss, N. & Hari, R. (2002). Modulated activation of the human SI and SII cortices during observation of hand actions. *NueroImage, 15,* 640-646.

Balog, H., & Brentari, D. (2008). The Relationship Between Early Gestures and Intonation. *First Language, 18*, 141-163.

Barbosa, P. A. (2001). *Generating duration from a cognitively plausible model of rhythm production.* Paper presented at the Eurospeech 2001, Ålborg, Denmark.

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

Bard, C, Turrell, Y, Fleury, M, Teasdale, N, Lamarre, Y, & Martin, O. (1999). Deafferentation and pointing with visual double-step perturbations. *Experimental Brain Research, 125*, 410-416.

Bates, E. & Dick, F. (2002). Language, gestures and the developing brain. *Developmental Psychobiology, 40,* 293-310.

Beattie, G. & Shovelton, H. (2006). A critical appraisal of the relationship between speech and gesture and its implications for the treatment of aphasia. *International Jounral of Speech-Language Pathology, 8,* 134-139.

Beckman, M.E. (1986). *Stress and non-stress accent.* Dordrecht, The Netherlands : Foris Publications.

Beckman, M.E. & Cohen, B. (2000). Modeling the articulatory dynamics of two levels of stress contrast. In M. Horne, ed., *Prosody, Theory and Experiment: Studies Presented to Gösta Bruce,* pp. 169-200. Springer.

Beckman, M.E. & Elam, G.A. (1997). *Guidelines for ToBI Labeling, Version 3.*

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P.A. Keating, ed., *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, pp. 7-33. Cambridge University Press.

Bernardis, P. & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia, 44*, 178-190.

Bernstein, N. (1967). *The Co-ordination and Regulation of Movements*. Oxford, UK: Pergamo.

Birdwhistell, R.L. (1952). *Introduction to kinesics. An annotation system for analysis of body motion and gesture*. Louisville: University of Louisville Press.

Birdwhistell, R.L. (1970). *Kinesics and Context: Essays on body motion communication*. Philadelphia: University of Pennsylvania Press.

Bishop, D.V.M. (1990). *Handedness and Developmental Disorder.* Cambridge, MA: Cambridge University Press.

Bishop, D.V.M. (2002). The role of genes in the etiology of speech and language impairments. *Journal of Communication Disorders, 35*, 311-328.

Bishop, D.V.M. & Edmundson, A. (1987). Specific language impairment as a maturational lag: evidence from longitudinal data on language and motor development. *Developmental Medicine and Child Neurology, 29*, 442-459.

Blake, J., Myszczyszyn, D., Jokel, A., & Bebiroglu, N. (2008). Gestures accompanying speech in specifically language-impaired children and their timing with speech. *First Language, 28*, 237-253.

Bluedorn, A. C. (2002). *The human organization of time: Temporal realities and experience.* Stanford, CA: Stanford University Press.

Blute, M. (2006). The evolutionary socioecology of gestural communication. *Gesture, 6,*177-188.

Bolinger, D.L. (1958). Stress and information. *American Speech, 33,* 5-20.

Bolinger (1972). Accent is predictable (If you're a mindreader). *Language,48,* 633-644.

Bonda, E., Petrides, M. Frey, S., & Evans, A.C. (1994). Frontal cortex involvement in organized sequences of hand movements: Evidence from a positron emission tomography study. *Society Neuroscience Abstracts, 20,* 152-156.

Brady, N.C., Marquis, J. Fleming, K. & McLean, L. (2004). Prelinguistic predictors of language growth in children with developmental disability. *Journal of Speech, Language and Hearing Research, 47,* 663-677.

Bull, P. & Connelly, G. (1985). Body movement and emphasis in speech. *Journal of Nonverbal Behavior, 9,* 169-187.

Burke, B.D. (1975). Susceptibility to delayed auditory feedback and dependence on auditory or oral sensory feedback. *Journal of Communication Disorders, 8,* 75-96

Burnett, T.A., Freedland, M.B., Larson, C.R. & Hain, T.C. (1998). Voice *F*0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America, 103,* 3153-3161.

Butcher, C. & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (ed.), *Language and gesture*. New York: Cambridge University Press.

Butterworth, B., & Beatty, G. (1978). Gesture and silence as indicators of planning in speech. In R. Campbell & G.T. Smith (Eds.), *Recent advances in the psychology of language: formal and experimental approaches*. New York: Plenum Press.

Butterworth, B. & Hadar, U. (1989). Gesture, speech, and computational stages: a reply to McNeill. *Psychological Review, 96*, 168-174.

Campbell, N., (2000), Timing in Speech: A Multi-Level Process, in M. Horne (ed), *Prosody: Theory and Experiment* (pp.281-335). Dordrecht: Kluwer Academic Publishers.

Capone, N.C. & McGregor, K. K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language and Hearing Research, 47,* 173-186.

Capone, N.C. & McGregor, K. K. (2005). The effect of semantic representation on toddler's word retrieval. *Journal of Speech, Language, and Hearing Research, 48,* 1468-1480.

Capone, N.C. (2007). Tapping toddlers' evolving semantic representation via gesture. *Journal of Speech, Language, and Hearing Research, 50,* 732-745.

Caselli, C.M., Vicari, S., Longobardi, E., Lami, L., Pizzoli, C., & Stella, G. (1998). Gestures and words in early development of children with down syndrome. *Journal of Speech, Language, and Hearing Research, 41,* 1125-1135.

Chang, P. & Hammond, G.R. (1987). Mutual interactions between speech and finger movements. *Journal of Motor Behavior, 19,*265-275.

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics, 37,* 871-887.

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.

Cohen, A.A. (1977). The communicative functions of hand illustrators. *Journal of Communication, 27,* 54-63.

Condon & Ogston (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease, 143*, 338-347.

Connaghan, K. P., Moore, C. A., Reilly, K. J., Almand, K. B., and Steeve, R. W. (2001): *Acoustic and physiologic correlates of stress production across systems*. Poster session presented at the American Speech-Language-Hearing Association, New Orleans, L.A.

Connaghan, K.P., Skinder, A.E., Strand, E., Hodge, M., Steeve, R.W. (1999). Stress production in children with speech disorders and typically developing peers. Poster presented at the American Speech-Language-Hearing Association Annual Conference, San Francisco, CA, November.

Cooper, W.E, Eady, S.J. & Mueller, P.R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America, 77,* 2142-2156.

Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton: University Press.

Corballis, M.C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences, 26,*199–260.

Corballis, M.C. (2010). Mirror neurons and the evolution of language. *Brain and Language, 112*, 25-35.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex, 45,* 119-130.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Crystal, D. (1975). *The English tone of voice*. London: Edward Arnold.

Crystal, D. (1979). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 33–48). Cambridge: Cambridge University Press.

Crystal, T. H. & House, A.S. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America, 83*, 1553-1573.

Cummins, F. (2002a). Speech rhythm and rhythmic taxonomy. In the *Proceedings of Speech Prosody 2002*, Aix-en-Provence.

Cummins, F.(2002b). On synchronous speech. *Acoustic Research Letters Online, 3,* 7-11.

Cummins, F. & Port, R.F. (1996). Rhythmic commonalities between hand gestures and speech. In *Proceedings of the eighteenth meeting of the Cognitive Science Society*, pp. 415-419. London: Lawrence Erlbaum

Cummins, F. & Port, R.F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26,* 145-171.

de Jong, K. (1991). The oral articulation of English stress accent. Unpublished Ph.D. Dissertation, Ohio State University.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics, 32*, 493-516.

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 D. McNeill (Ed.), *Language and Gesture* (pp. 284-311). Cambridge, UK: Cambridge University Press.

de Ruiter, J.P. (2006). Can gesticulation help aphasix people speak, or rather, communicate? *International Journal of Speech-Language Pathology, 8*, 124-127.

Dittman, A.T., & Llewellyn, L.G. (1969). Body movement and speech rhythm in social conversation. *Journal of Personality and Social Psychology, 11,* 98-106.

Dromey, C., & Bates, E. (2005). Speech interactions with linguistics, cognitive, and visuomotor tasks. *Journal of Speech, Language, and Hearing Research, 48,* 295-305.

Dromey, C., & Benson, A. (2003). Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance. *Journal of Speech, Language, and Hearing Research, 46,* 1234-1246.

Dromey, C.,& Ramig, L.O. (1998). Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation. *Journal of Speech, Language, and Hearing Research, 41,* 1003-1018.

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

Efron, D. (1941). *Gesture and environment.* New York: King's Crown Press.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1,* 49-98.

Ekman, P., & Friesen, W.V.(1972). Hand movements. *Journal of Communication, 22,* 353-374.

Emmorey, K.D. (1987). The neurological substrates for prosodic aspects of speech. *Brain and Language, 30,* 305-320.

Engel, A.K., Konig, P., Kreiter, A.K., Schillen, T.B., & Singer, W. (1992). Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends in Neuroscience, 15,* 218-226.

Erhard, P., Kato, T., Strupp, J.P., Anderson, G., Adriany, G. & Strick, P.L. et. al. (1996). Functional mapping of motor activity in and near Broca'a area. *Neuroimage, 3,* 1053-8119.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development , 59* (PAGES???)

Ferrari, P.F., Gallese, V., Rizzolatti, G., Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17,* 1703-1714.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological review, 100,* 233-253.

Feyereisen, P. (1983). Manual activity during speaking in Aphasic subjects. *International Journal of Psychology, 18,* 545-556.

Feyereisen, P. (2000). Representational gesture as action in space: Propositions for a research program. *Brain and Cognition, 42,* 149-152.

Feyereisen, P. (2006). How could gesture facilitate lexical access?. *Advances in Speech Language Pathology, 8,* 128-133.

Feyereisen & deLannoy (1991). *Gesture and speech: Psychological investigations.* Cambridge: Cambridge University Press.

Finney, S.A., & Warren, W.H. (2002). Delayed auditory feedback and rhythmic tapping: Evidence for a critical interval shift. *Perception & Pyschophysics, 64,* 896-908.

Folkins, J.W., & Abbs, J.H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research, 18,* 207-220.

Folkins, J.W., & Zimmermann, G.N. (1981). Jaw-muscle activity during speech with the mandible fixed. *The Journal of the Acoustical Society of America, 69,* 1441-1445.

Franks, I. M., Nagelkerke, P., Ketelaars, M., & van Donkelaar, P. (1998). Response preparation and control of movement sequences. *Canadian Journal of Experimental Psychology*, *52*, 93-101.

Franz, E.A., Zelaznik, H.N. & Smith, A. (1992). Evidence of common timing processes in the control of manual, orofacial, and speech movement. *Journal of Motor Behavior, 24,* 281-287.

Fried, I., Katz, A., McCarthy, G., Sass, K.J., Williamson, P.,Spencer, S.S. et al. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience, 11,* 3656-3666.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 30*, 765-769.

Garcia, J.M. & Cannito, M.P. (1996). Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. *Journal of Speech and Hearing Research, 39,* 750-760.

Garcia, J.M. Cannito, M.P. & Dagenais, P.A. (2000). Hand gestures: Perspectives and preliminary implications for adults with acquired dysarthria. *American Journal of Speech-Language Pathology, 9,* 107-115.

Garcia, J.M. & Dagenais, P.A. (1998). Dysarthric sentence intelligibility: Contribution of iconic gestures and message predictiveness. *Journal of Speech, Language and Hearing Research, 41,* 1282-1293.

Garcia J.M., Dagenais P.A., & Cannito M.P. (1998). Intelligibility and acoustic differences in dysarthric speech related to use of natural gestures. In M. P. Cannito, K. M. Yorkston, & D. R. Beukelman, (Eds.), *Neuromotor Speech Disorders: Nature, assessment, and management* (pp. 213-227). Baltimore, MD: Brookes Publishing.

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: Kinematic study on healthy subjects. *The Journal of Neurophysiology, 86,* 1685-1699.

Gentilucci, M. & Dalla Volta, R. (2007). The motor system and the relationship between speech and gesture. *Gesture, 7*, 159–177.

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

Gerken, L.A. (1991). The metric basis for children's subjectless sentences. *Journal of Memory and Language, 30,* 431-451.

Gerken, L.A., Juszyk, P.W., & Mandel, D.R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition, 51,* 237-265.

Gerken, L.A., & McGregor, K. (1998). An overview of prosody and its role in normal and disordered child language. *American Journal of Speech-Language Pathology, 7,* 38-48.

Gleick, J. (1987). *Chaos. Making a New Science*. New York, NY: Penguin Books USA Inc.

Goffman, L., & Malin, C. (1999). Metrical effects on speech movements in children and adults. *Journal of Speech, Language and Hearing Research, 42,* 1003-1015.

Goldin-Meadow, S. (1998). The development of gesture and speech as an integrated system. *New Directions for Child Development, 79,* 29-42.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3,* 419-429.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help usthink*. Cambridge, MA: Harvard University Press.

Goodwyn, S. W., & Acredolo, L. P. (1993). Symbolic gesture versus word: Is there a modality advantage for onset of symbol use? *Child Development, 64*, 688-701.

Goodwyn, S.W., Acredolo, L.P., & Brown, C.A. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior, 24,* 81-103.

Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics, 26,* 129-143.

Grabowski, T.J., Damasio, H., & Damasio, A.R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *NeuroImage, 7,* 232-243.

Gracco, V.L. (1994). Some organizational characteristics of speech movement control. *Journal of Speech and Hearing Research, 37*, 4-27.

Gracco, V.L., & Abbs, J.H. (1985). Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Nueropathology, 54,* 418-432.

Gracco, V.L., & Abbs, J.H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research*, *65,* 155-166.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable centric perspective. *Journal of Phonetics, 31,* 465-485.

Grézes, J., Armony, J.L., Rowe, J., & Passingham, R.E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: An fMRI study. *NeuroImage, 18,* 928-937.

Grosjean, F. & Gee, J.P. (1987). Prosodic structure and spoken word recognition. *Cognition, 25,* 135-155.

Gross-Tsur, V., Manor, O., Joseph, A., & Shavlev, R.S. (1996). Comorbidity of developmental language disorders and cognitive dysfunction. *Annals of Neurology, 40,* 338-339.

Guenther, F.H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105,* 611-633.

Gussenhoven, C. (1991). The English rhythm rule as an accent deletion rule. *Phonology, 8,* 1-35.

Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge: Cambridge University of Press.

Hadar, U. (1989). Two types of gestures and their role in speech production. *Journal of Language and Social Psychology, 8,* 221-228.

Hadar, U., Wenkert-Olenik, D., Krauss, R., & Soroker, N. (1998). Gesture and the processing of speech: Neuropsychological evidence. *Brain and Language, 62,* 107-126.

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

Halle, M., & Vergnaud, J.R. (1987). Stress and the cycle. *Linguistic Inquiry, 18,* 45-84.

Hamblin, E. (2005). *The effects of divided attention on speech motor, verbal, fluency and manua l task performance* (Master's Thesis, Bringham Young University 2005). Provo, UT: Bringham Young University.

Hammer, D.W. (2006). *Treatment strategies for childhood apraxia of speech [*DVD]. Pittsburgh, PA: CASANA.

Hanlon, R.E., Brown, J.W. & Gerstman, L.J. (1990). Enhancement of naming in nonfluent aphasia through gesture. *Brain Language, 38,* 298-314.

Harrington, J., Fletcher, J., & Roberts, C. (1995). Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics, 23,* 305-322.

Hashimoto, Y., & Sakai, K. L. (2003). Brain activations duringconscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping, 20*, 22–28.

Hayes, B. (1985). Iambic and trochaic rhythm in stress rules. In *Proceedings of the thirteenth meeting of the Berkley Linguistics Society,* ed. by M. Niepokuj, et. al., 429-446. Berkley, CA: Berkley Linguistic Society.

Hayes, B. (1986). Inalterability in CV phonology. *Language, 62,* 321-351.

Hayes, B. (1989). The prosodic hierarchy in meter. In *Rhythm and Meter*, ed. by Paul Kiparsky and Gilbert Youmans, 201-260. Orlando, FL: Academic Press.

Hayes (1995). *Metrical Stress Theory.* Chicago: University of Chicago Press.

Hayes. B., & Lahiri, A. (1991). Bengali intonation phonology. *Natural Language and Linguistic Theory, 9,* 47-96.

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J.C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience, 17,* 1123-1128.

Higgins, C. & Hodge, M. (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology, 10*, 271-274.

Hill, E.L. (2001). Nonspecific nature of specific language impairment: A review of the literature in regards to concomitant motor impairments. *International Journal of Language & Communication Disorders, 36,* 149-171.

Hill, E.L., Bishop, D.V.M., & Nimmo-Smith, I. (1998). Representational gestures in developmental coordination disorder and specific language impairment: Error-types and the reliability of ratings. *Human Movement Science, 17,* 655-678.

Higgins, C. & Hodge, M. (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology, 10*, 271-274.

Hird, K., & Kirsner, K. (1993). Dysprosody following acquired Neurogenic impairment. *Brain and Language, 45,* 46-60.

Hirsh-Pasek, K., Kemler-Nelson, D. G., Jusczyk, P. W., Wright-Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269-286.

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. *Perception and Psychophysics, 36,* 296-302.

Howell P. & Dworzynski K. (2001) Strength of German accent under altered auditory feedback. *Perception & Psychophysics, 63,* 501–513.

Howell P. & Sackin S. (2002) Timing interference to speech in altered listening conditions. *Journal of the Acoustical Society of America, 111*, 2842–2852.

Hustad. K.C., & Garcia, J.M. (2005). Aided and unaided speech supplementation strategies: Effect of alphabet cues and iconic hand gestures on dysarthric speech. *Journal of Speech, Language and Hearing Research, 48,* 996-1012.

Iverson, J.M., Capirci, O., Longobardi, E., & Caselli, M.C. (1999). Gesturing in mother-child interactions. *Cognitive Development, 14,* 57-75.

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

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

Iverson, J.M, Longobardi, E., & Caselli. M.C. (2003). Relationship between gestures and words in children with Down's syndrome and typically developing children in the early stages of communicative development. *International Journal of Language & Communication Disorders, 38,* 179-197.

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

Johnston, J.C., Durieux-Smith, A., & Bloom, K. (2005). Teaching gestural signs to infants to advance child development: A review of the evidence. *First Language, 25,* 235-251.

Johnston, R.B., Stark, R.E., Mellits, E.D., & Tallal, P. (1981). Neurological status of language impaired and normal children. *Annals of Neurology, 10,* 159-163.

Jones, M.R., & Boltz, M. (1989). Dynamic attending and response to time. *Psychological Review, 96,* 459-491.

Jones, J. A., & Striemer, D. (2007). Speech disruption during delayed auditory feedback with simultaneous visual feedback. *Journal of the Acoustical Society of America, 122*, 135-141.

Kalinowski, J., Stuart, A., Sark, S., & Arnson, J. (1996). Stuttering amerlioration at various auditory feedback delays and speech rates. *International Journal of Language and Communication Disorders, 31*, 259-269.

Keele, S.W., Jennings, P., Jones, S., Caulton, D. & Cohen, A. (1995). On the modularity of sequence representation. *Journal of Motor Behavior, 27*, 17-30.

Kelso, J.A. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 15,* R1000-R1004.

Kelso, J.A., & Tuller, B. (1983). "Compensatory articulation" under conditions of reduced afferent information: A dynamic formulation. *Journal of Speech and Hearing Research, 26,* 217-224.

Kelso, J.A., & Tuller, B. (1984). Converging evidence in support of common dynamical principles for speech and movement coordination. . *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 246,* R938-R935.

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

Kelso, J., Tuller, B., & Harris, K. (1983). A ''Dynamic Pattern''perspective on the control and coordination of movement. In P. MacNeilage (Ed.), *The production of speech* (pp. 137–173). New York: Springer Verlag.

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. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 812-832.

Kelso, J.A., & Zanone, P.G. (2002). Coordination and dynamics of learning and transfer across different effector systems. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 776-797.

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

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

Kent, R.D. (1984). Psychobiology of speech development: Coemergence of language and a movement system. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 246,* R888-R894.

Kent (1997). Gestural phonology: Basic concepts and applications in speech-language pathology. In M. J. Ball & R. D. Kent (Eds.), *The new phonologies: Developments in clinical linguistics* (pp. 247-268). London: Singular Press.

Kent, R.D., & Read, C. (2002). *The Acoustic Analysis of Speech.* San Diego, CA: Singular.

Kinsbourne, M., & Cook, J. (1971). Generalized and lateralized effects of concurrent verbalization on a unimanual skill. *The Quarterly Journal of Experimental Psychology, 23,* 341-345.

Kinsbourne, M., & Hiscock, M. (1983). Asymmetries of dual task performance. In J.B. Hellige (Ed.), *Cerebral hemisphere asymmetry: method, theory and application* (pp. 255-334). New York: Praeger.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48,* 16-32.

Klapp, S.T., Porter-Graham, K.A., & Hoifjeld, A.R. (1991). The relation of perception and motor action: ideomotor compatibility and interference in divided attention. *Journal of Motor Behavior, 23,* 155-162.

Komilis, E., Pelisson, D., & Prablanc, C. (1993). Error processing in pointing at randomly feedback induced double step stimuli. *Journal of Motor Behavior, 25,* 299-308.

Krahmer, E., & Swerts, M. (2007). Effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57,* 396–414.

Krams, M., Rushworth, M.F., Deiber, M.P., Frackowiak, R.S., & Passingham, R.E. (1998). The preparation, execution, and suppression of copied movements in the human brain. *Experimental Brain Research, 120,* 386-398.

Krauss, R.M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us?. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 389-450). San Diego: Academic Press..

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and Gesture* (261-283). New York: Cambridge University Press.

Krauss, R.M., Dushay, R.A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gesture. *Journal of Experimental Social Psychology, 31,* 533-552.

Krauss, R.M., & Hadar, U. (1999). The Role of Speech related arm/hand gestures in word retrieval. In R. Campbell & L. Messing (Eds.), *Gesture, Speech, and Sign* (pp. 93-116). Oxford: Oxford University Press.

Kugler, P.N. & Turvey, M.T. (1987). *Information, Natural Law, and the Self-Assembly of Rhythmic Movement*. London: Lawrence Earlbaum

LaBarba, R.C., Bowers, C.A., Kingsberg, S.A., & Freeman, G. (1987). The effects of concurrent vocalization on foot and hand motor performance: A test of the functional distance hypothesis. *Cortex, 23,* 301-308.

Ladd. D.R.(1980). *The Structure of Intonational Meaning*. Bloomington: Indiana University Press.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Large, E.W., & Jones, M.R. (1999). The dynamics of attending: How people track time varying events. *Psychological Review, 106,* 119-159.

Lashley, K.S. (1951), The problem of serial order in behavior. In L.A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior.* New York, NY: Wiley.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5,* 253-263.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.

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

Liberman, M.Y. (1975). *The intonational system of English*. Unpublished doctoral dissertation, MIT, Cambridge.

Liberman, M.Y., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8,* 249-336.

Liberman, A.M., & Whalen, D.H. (2000). Of the relation of speech to language. *Trends in Cognitive Sciences, 4,* 187-196.

Lieberman,P. (1998). *Eve Spoke: Human Language and Human Evolution.* New York, NY: WW Norton & Company.

Loehr, D.P. (2004). Gesture and Intonation (Doctoral dissertation, Georgetown University, 2004). *Dissertation Abstracts International, 65* (06), 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. *Journal of Computational Chemistry, 28*, 1384-`399.

McEachern, D. & Haynes, W.O. (2004). Gesture-speech combinations as a transition to multiword utterances. *American Journal of Speech-Language Pathology, 13*, 227-235.

MacNeilage, P.F. & Davis, B.L. (1993). Motor explanations of babbling and early speech patterns. In B.B. Bardies, S. de Schoen, P. Juszyk, P.F. MacNeilage, & J. Morton (Eds.), Developmental Neurocognition: *Speech and Face Processing in the First Year of Life* (pp. 341-352). Dordrect: Kluwer.

Marslen-Wilson, W.D. & Tyler, L.K. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society, Series B, 295*, 317-332.

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

Mayberry, R.I., Jaques, J., & DeDe, G. (1998). What stuttering reveals about the development of gesture speech relationship. *New Directions for Child Development, 79,* 77-87.

McClave, E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research, 23*, 45-66.

McClave, E. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research, 27*, 69-89.

McLean, L.P. & McLean, J.E. (1974). A language training program for nonverbal autistic children. *Journal of Speech and Hearing Disorders, 39*, 186-193.

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

McClean, M.D., & Tasko, S.M. (2004). Correlation of orofacial speeds with voice acoustic measures in the fluent speech of persons who stutter. *Experimental Brain Research, 159,* 310-318.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92,* 350-371.

McNeill (1987). So you think gestures are nonverbal? Reply to Feyereisen. *Psychological Review, 94,* 499-504.

McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.

McNeill, D. (2000). *Language and gesture.* Cambridge: Cambridge University Press.

McNeill (2005). *Gesture and thought.* Chicago: University of Chicago Press.

Meister, I.G., Boroojerdi, B., Foltys, H., Sparing, R., Huber, W., & Topper, R. (2003). Motor cortex hand area and speech: Implications for the development of language. *Neuropsychologia, 4,* 401-406.

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

Morrel-Samuels, P., & Krauss, R.M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology, 18,* 615-622.

Morsella, E.,& Krauss, R.M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology, 117,* 411-424.

Munhall, K.G. (1985). An examination of intra-articulator relative timing. *Journal of the Acoustical Society of America, 78,* 1548-1553.

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

Namy, L.L., Acredolo, L., & Goodwyn, S. (2000). Verbal labels and gestural routines in parental communication with young children. *Journal of Nonverbal Behavior, 24,* 63-79.

Neijt, A.H. (1990). Prosodic structures and phonetic findings-the case of equally stressed adjectives. In B. Reineke & D. Coopmans (Eds.), *Linguistics in the Netherlands* (pp. 113-122). Dordrecht: Foris.

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology.* Dordrecht: Foris Publications.

Nespor, M., & Vogel, I. (1989). On clashes and lapses. *Phonology, 6,* 69-116.

Nobe, S. (1996). Representational Gestures, Cognitive Rhythms, and Acoustic Aspects of Speech: A Network/Threshold Model of Gesture Production (Doctoral dissertation, University of Chicago, 1996). *Dissertation Abstracts International, 57* (07), 4736. (UMI No. 9636827)

O'Dell, M., & Nieminen, T. (1999). *Coupled oscillator model of speech rhythm.* In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences*, *Vol. 2* (pp. 1075–1078). New York: American Institute of Physics.

Ojemann, G.A. (1984). Common cortical and thalamic mechanisms for language and motor function. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 246,* R901-R903.

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday videotapes. *Journal of Autism and Developmental Disorders, 24,* 247-257.

Ostry, D. J., Cooke, J. D., & Munhall, K. G. (1987) Velocity curves of human arm and speech movements, *Experimental Brain Research*, *68*, 37-46.

Pashek, G.V. (1997). A case study of gesturally cued naming in aphasia: Dominant versus nondominant hand training. *Journal of Communication Disorders, 30,* 349-366.

Paulignan, Y., Jeannerod, M., MacKenzie, C., & Marteniuk, R. (1991). Selective perturbation of visual input during prehension movements. *Experimental Brain Research, 87,* 407-420.

Peters, J.M. (1977). *Pictorial communication.* Cape Town: David Phillip.

Peterson, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience, 1,* 153-170.

Pfordresher, P. Q., & Benitez, B. (2007). Temporal coordination between actions and sound during sequence production. *Human Movement Science, 5*, 742-756.

Pierrehumbert, J.B. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, MIT, Cambridge.

Pierrehumbert, J., & Beckman, M. (1988). Japanese tone structure. *Linguistic inquiry monographs, 15,* 1-282.

Pike, K.L. (1945). *The intontation of American English.* Ann Arbor: University of Michigan Press.

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*, 747-754.

Port, R.F. (2003). Meter and speech. *Journal of Phonetics, 31,* 599-611.

Port, R., Tajima, K., & Cummins, F. (1998). Speech and rhythmic behavior. In G. J. P. Savelsburgh, H. van der Maas, & P. C. L. van Geert (Eds.), *The nonlinear analysis of developmental processes* (pp.5-45). Amsterdam: Royal Dutch Academy of Arts and Sciences.

Powell, R.P., & Bishop, D.V. (1992). Clumsiness and perceptual problems in children with specific language impairment. *Developmental Medicine and Child Neurology, 34,* 755-765.

Power, E., & Code, C. (2006). Waving not drowning: Utilising gesture in the treatment of aphasia. *International Journal of Speech-Language Pathology,8,* 115-119.

Prablanc, C., & Martin, O. (1992). Automatic control during hand reaching at undetected two-dimensional target displacements. *Journal of Neurophysiology, 67,* 455-469.

Ragsdale, J.D., & Silvia, C. (1982). Distribution of kinesic hesitation phenomena in spontaneous speech. *Language and Speech Teddington, 25,* 185-190.

Rauscher, F. B., Krauss, R. M., & Chen, Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science ,7*, 226-231.

Ravizza, S. (2003). Movement and lexical retrieval: Do noniconic gestures aid in retrieval?. *Psychonomic Bulletin and Review, 10,* 610-615.

Raymer, A.J. (2007, June 19). Gestures and words: facilitating recovery in aphasia. *The ASHA Leader, 12*, 8-11.

Raymer, A.M., Singletary, F., Rodriguez, A., Ciampitti, M., Heilman, K.M., & Rothi, L. (2006). Effects of gesture + verbal treatment for noun and verb retrieval in aphasia. *Journal of the International Neuropsychological Society, 12,* 867-882.

Richards, K., Singletary, F., Rothi, L.J., Koehler, S., & Crosson, B. (2002). Activation of intentional mechanisms through utilization of nonsymbolic movements in aphasia rehabilitation. *Journal of Rehabilitation Research and Development, 39,* 445-454.

Rime, B., & Schiaratura, L. (1991). Gesture and Speech. In R. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (pp. 239-281). Cambridge: Cambridge University Press.

Ringel, R.L. (1970). Oral sensation and perception: a selected review. *Speech and the Dentofacial Complex: The State of the Art, ASHA Reports 5.* Washington D.C.: American Speech and Hearing Association.

Ringel, R.L., & Steer, M.D. (1963). Some effects of tactile and auditory alterations on speech output. *Journal of Speech and Hearing Research, 13,* 369-378.

Rizzolatti, G., & Arbib, M.A. (1998). Language within our grasp. *Trends in Neuroscience, 21,* 188-194.

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

Rose, M.L. (2006). The utility of arm and hand gestures in the treatment of aphasia. *Advances in Speech Language Pathology, 8,* 92-109.

Rose, M.L., & Douglas, J. (2001). The differential faciliatory effects of gesture and visualisation processes on object naming in aphasia. *Aphasiology, 15,* 977-990.

Rose, M., Douglas, J., & Matyas, T. (2002). The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology, 16,* 1001-1030.

Rusiewicz, H.L., Dollaghan, C.A., & Campbell, T.F. (2003). Separating lexical and phrasal Stress. Poster presented at the American Speech-Language-Hearing Association Annual Convention, Chicago, IL, November.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase window and multifrequency rhythms. *Human Movement Science, 19,* 499-526.

Saltzman, E.L., & Munhall, K.G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1,* 330-380.

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology, 19*, 827-830.

Scharp, V. L., Tompkins, C. A., & Iverson, J. M. (2007). Gesture and aphasia: Helping hands? *Aphasiology*, *21*, 717-725.

Schlaug, G., Knorr, U., & Seitz, R.J. (1994). Inter-subject variability of cerebral activations in acquiring a motor skill: a study with positron emission tomography. *Experimental Brain Research, 98,* 523-534.

Schmidt, R.A. & Lee, T.D. (1999). *Motor Control and Learning: A Behavioral Emphasis* (3rd ed.), Champaign, IL: Human Kinetics..

Schmidt, R.C., Shaw, B.K., & Turvey, M.T. (1993). Coupling dynamics in interlimb coordination. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 397-415.

Schulman, R.(1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America, 85,* 295-312.

Schwartz, R.G. & Goffman, L. (1995). Metrical patterns of words and production accuracy. *Journal of Speech and Hearing Research, 38,* 876-888.

Selkirk, E.O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry, 11,* 563-605.

Selkirk, E.O. (1984). *Phonology and Syntax.* Cambridge: MIT Press.

Selkirk, E.O. (1986*). Phonology and Syntax: The relationship between sound and structure.* Cambridge: MIT Press.

Selkirk, E.O. (1995). Sentence Prosody: Intonation, Stress and Phrasing. In John Goldsmith (Ed.) *Handbook of Phonological Theory* (pp. 550-569). Oxford: Basil Blackwell.

Selkirk, E.O. (1996). The prosodic structure of function words. In *University of Massachusetts occasional papers 18: Papers in Optimality Theory* (pp.439-469). Amherst:University of Massachusetts.

Seth-Smith, M., Ashton, R., & McFarland, K. (1989). A dual-task study of sex differences in language reception and production. *Cortex, 25,* 425-431.

Seyal, M., Mull, B., Bhullar, N., Ahmad, T., & Gage, B. (1999). Anticipation and execution of a simple reading task enhance corticospinal excitability. *Clinical Neurophysiology, 110,* 424-429.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *The Journal of the Acoustical Society of America, 86,* 78-88.

Shaiman, S., & Gracco, V.L. (2002). Task-specific sensorimotor interactions in speech production. *Experimental Brain Research, 146,* 411-418.

Shattuck-Hufnagel, S. (1995). The importance of phonological transcription in empirical approaches to 'stress shift' versus 'early accent': Comments on Grabe and Warren. In B. Connell and A. Arvaniti (Eds.), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence* (pp. 128-141). Cambridge: Cambridge University Press.

Shattuck-Hufnagel, S. & Turk, A.E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 193-247.

Shriberg, L. D., Campbell, T. F., Karlsson, H. B., McSweeney, J. L, & Nadler, C. J. (2003) A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Clinical Linguisitics and Phonetics*, *17*, 549-574.

Shumway, S., & Wetherby, A. M. (2009). Communicative Acts of Children with Autism Spectrum Disorders in the Second Year of Life. *Journal of Speech, Language, and Hearing Research, 52*, 1139-1156.

Sluijter, A.M.C. (1995). *Phonetic correlates of stress and accent.* Unpublished doctoral dissertation, Leiden University.

Sluijter, A.M.C. & van Heuven, V.J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America, 100*, 2471-2485.

Smith, A. (1992). The control of orofacial movements in speech. *Critical Reviews in Oral Biology and Medicine, 3*, 233-267.

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

Smith, V., Mirenda, P., & Zaidman-Zait, A. (2007). Predictors of expressive vocabulary growth in children with autism. *Journal of Speech, Language, and Hearing Research, 50,* 149-160.

Snellen, H. (1862). Eye examination chart originally created by Dutch ophthalmologist Hermann Snellen, M.D.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research, 37*, 831-840

Snow, D. (1998). A prominence account of syllable reduction in early speech development: The child's prosodic phonology of tiger and giraffe. *Journal of Speech, Language, Hearing Research, 41,* 1171-1184.

Stefanini, S., Caselli, M.C., & Volterra, V. (2007). Spoken and gestural production in a naming task by young children with Down syndrome. *Brain and Language, 101,* 208-221.

Stone, W.L., Ousley, O.Y., Yoder, P.J., Hogan, K.L., & Hepburn, S.L. (1997). Nonverbal communication in two- and three-year-old children with Autism. *Journal of Autism and Developmental Disorders, 27,* 677-696.

Stuart, A., Kalinowski, J., Rastatter, M.P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *The Journal of the Acoustical Society of America, 111,* 2237-2241.

Studdert-Kennedy (2002). Mirror neurons, vocal imitation, and the evolution of particulate speech. In M.I. Stamenov & V. Gallese (Eds.), *Motor Neurons and the Evolution of Brain and Language.* Amsterdam, The Netherlands: John Benjamins Publishing.

Summers, W.V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustical analyses. *The Journal of the Acoustical Society of America, 82,* 847-863.

Terai, Y., Ugawa, Y., Enomoto, H., Furubayashi, T., Shiio, Y., Machii, K. et al. (2001). Hemispheric lateralization in the cortical motor preparation for human vocalization. *The Journal of Neuroscience, 21,* 1600-1609.

Thal, D. (1991). Language and cognition in normal and late-talking toddlers. *Topics in Language Disorders, 11,* 33-42.

Thal, D., & Bates, E. (1988). Language and gesture in late talkers. *Journal of speech and hearing research, 31,* 115-123.

Thal, D., O'Hanlon, L., Clemmons, M., & Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language and Hearing Research, 42,* 482-496.

Thal, D., & Tobias, S. (1992). Communicative gestures in children with delayed onset of oral expressive vocabulary. *Journal of Speech and Hearing Research, 35,* 1281-1290.

Thal, D., & Tobias, S., (1994). Relationships between language and gesture in normally developing and late-talking toddlers. *Journal of Speech and Hearing Research, 37,* 157-170.

Thal, D., Tobias, S., Morrison, D. (1991). Language and gesture in late talkers: A 1-year follow up. *Journal of Speech and Hearing Research, 34,* 604-612.

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

Thornton, C.D., & Peters, M. (1982). Interference between concurrent speaking and sequential finger tapping: Both hands show a performance decrement under both visual and non-visual guidance. *Neuropsychologia, 20,* 163-169.

Tokimura, H., Tokimura, Y., Oliviero, A., Asakura, T., & Rothwell, J.C. (1999). Speech-induced changes in corticospinal excitability. *Annals of Neurology, 40,* 628-634.

Tuite, K. (1993). The production of gesture. *Semiotica, 93,* 8-105.

Tuller, B., Harris, K.S., & Kelso, J.A. (1982). Stress and rate: Differential transformation of articulation. *The Journal of the Acoustical Society of America, 71,* 1534-1543.

Tuller, B., Harris, K.S., & Kelso, J.A. (1983). Converging evidence for the role of relative timing in speech. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 829-833.

Tuller, B. & Kelso, J.A. (1989). Environmentally-specified patterns of movement coordination in normal and split-brain subjects. *Experimental Brain Research, 75,* 306-316.

Tuller, B. & Kelso, J. A. S. (1990) Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 429–52). Hillsdale, NJ: Erlbaum Associates.

Tuller, B., Kelso, J.A., & Harris, K.S. (1982). On the kinematics of articulatory control as a function of stress and rate. *The Journal of the Acoustical Society of America, 72,* S103-

Turk, A.E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics, 27,* 171-206.

Vangheluwe, S., Suy, E., Wenderoth, N., & Swinnen, S. (2006). Learning and transfer of bimanual multifrequency patterns: Effector-independent and effector-specific levels of movement representation. *Experimental Brain Research, 170,* 543- 554.

van der Merwe, A. (1997). A theoretical framework for the characterization of pathological speech sensorimotor control. In M. McNeil (Ed.), *Clinical Management of Sensorimotor Speech Disorders. New* York, NY: Thieme Medical Publishers, Inc.

van Kuijk, D. & Boves, L. (1999). Acoustic characteristics of lexical stress in continuous telephone speech. *Speech Communication, 27*, 95-111.

van Wijngaarden, S. J., & van Balken, J. A. (2007). Theoretical feasibility of suppressing offensive sports chants by means of delayed feedback of sound. *The Journal of the Acoustical Society of America, 122*, 436-445.

von Holst, E. (1937). On the nature and order of the central nervous system. In R. Martin (Ed.), *The collected papers of Erich von Holst; The behavioral physiology of animal and man*. Coral Gables, FL: University of Miami Press.

von Holst, E. (1973). Relative coordination as a phenomenon and as a method of analysis of central nervous system function. In R. Martin (Ed.), *The collected papers of Erich von Holst; The behavioral physiology of animal and man* Coral Gables: University of Miami Press.

Volterra, V., Caselli, MC., Capirci, O., & Pizzuto, E. (2004). Gesture and the emergence and development of language. In M. Tomasello & D.I. Slobin (Eds.), *Beyond Nature-Nurture : Essays in Honor of Elizabeth Bates.* Mahwah, NJ : Lawrence Erlbaum Associates.

Watt, N., Wetherby, A., & Shumway, S. (2006). Prelinguistic predictors of language outcome at 3 years of age. *Journal of Speech, Language and Hearing Research, 49,* 1224-1237.

Weismer, S.E., & Hesketh, L.J. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research, 36,* 1013-1025.

Weismer, S.E., & Hesketh, L.J. (1998). The impact of emphatic stress on novel word learning by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1444-1458.

Wheeldon, L. (2000). *Aspects of Language.* New York: Psychology Press.

Wouters, J., & Macon, M.W. (2002). Effects of prosodic factors on spectral dynamics. I. Analysis.: *The Journal of the Acoustical Society of America*, *111,* 417-427.

Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S. (2004; June). *The timing of speech-accompanying gestures with respect to prosody.* Paper presented at the meeting the Sound to Sense: 50+ Years of Discoveries in Speech Communication, Cambridge, MA.