8-9-2014

Effects of Language Predictability and Speech Rate on Speech Entrainment Performance in Healthy Individuals

Honey Isabel Hubbard
University of South Carolina - Columbia

Effects of Language Predictability and Speech Rate on Speech Entrainment Performance in Healthy Individuals

by

Honey Isabel Hubbard

Bachelor of Art
University of Tennessee, 2008

Master of Science
University of Texas at Dallas, 2010

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Communication Sciences and Disorders

The Norman J. Arnold School of Public Health

University of South Carolina

2014

Accepted by:

Julius Fridriksson, Major Professor

Al Montgomery, Committee Member

Paul Fillmore, Committee Member

Christopher Rorden, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

DEDICATION

I dedicate my dissertation work to my family. Especially to my father, Rex Hubbard, for his support and inspiration; to my mother, Helen Tinnerman, for teaching values of perseverance and the importance of hard work.

To Auntie Doris Cheshire for countless phone chats full of news, weather, movie reviews, and happy thoughts.

To my siblings, Heather, Hilary, Stefan, Haley, Sterling, and Happy, for their continual cheerleading spirit and reminding me to play as hard as I work.

ACKNOWLEDGEMENTS

I would like to thank my mentor Julius Fridriksson for persevering with me as my advisor throughout the time it took me to complete my degree. Further thanks to Chris Rorden for supporting me and continuing to teach me in my final year. Many thanks to Al Montgomery and Paul Fillmore, who have generously given their time and expertise to better my work. I thank them all for their contribution and their good-natured support.

This work could not be completed without the continual help and guidance from members of the Communication Sciences and Disorders Department. Thanks especially to Karen Mullis and Teresa Boyett for keeping me in line. Thank you, Hiram McDade, for your sense of humor and making me smile. Further thanks goes to Taylor Hanayik for his help, support and contributions in Matlab. My research would not be the same without the help of Angela Hancock and Jennifer Newman for their contributions to the development and recording of stimuli, and to Kaitlin Krebs for her organization and formatting skills.

Finally, I would like to acknowledge my friends: Emilie Blevins, April Doyle, and Christina Sellers for keeping me grounded. Jan Lougeay for her encouragement and insight. My friends at the Mad Platter, for art and creativity. Sarah Grace Hudspeth for helping me forge a path. Thanks for your superpowers. Kimberly Ann Graham Smith for planning and scheming future career paths. Your drive is contagious. Jessica Richardson for being a constant source of feedback and celebrating even small victories. You work hard and you do it with grace.

These are the people who make life rich. A million thanks.

Abstract

Speech entrainment, a paradigm in which a participant shadows the speech of an audiovisual model in real time, has been show to benefit individuals with non-fluent aphasia. A study examining the effects of language predictability and speech rate was conducted to understand factors that influence speech entrainment performance.

A recent study by Fridriksson and colleagues (2012) demonstrated that training with speech entrainment significantly increased the number of words participants with non-fluent aphasia were able to produce. Perhaps even more remarkably, these effects showed generalization. As a result, speech entrainment could be used to rehabilitate speech impairment in stroke. However, there is very limited understanding of the factors that drive and influence speech entrainment performance. In order to better understand the implications of speech entrainment as a therapeutic technique, the current study examines some of the factors we predict may influence speech entrainment performance. This performance will be measured by comparing the delay (in time) between an audiovisual model and the speech of the healthy participants during a speech entrainment task at specific points in a sentence.

Forty participants were recruited from the study body at the University of South Carolina. During the study, normal participants completed either: 1) a speech rate experiment: an experimental paradigm that manipulates the rate of

the audiovisual presentation or 2) a language predictability experiment: an experimental paradigm that manipulates the predictability of upcoming words in a sentence. Twenty participants completed the language predictability task, and twenty participants completed the speech rate task. Error analysis, effects of cloze probability, and speech entrainment accuracy were also explored.

Table of Contents

List of Tables

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAC .... Augmentative Alternative Communication

AOS ....Apraxia of Speech

CMU.. Carnegie Mellon University

DTI .... Diffusion Tensor Imaging

fMRI .... functional Magnetic Resonance Imaging

iPhOD .... Irvine Phonotactic Online Dictionary

LSA ....Latent Semantic Analysis

MIT...Melodic Intonation Therapy

ms .... milliseconds

PH .... Phonological

PS ....Pseudoword

SE ....Semantic

TH ....High Target

wpm .... words per minute

wps ....words per second

CHAPTER 1

INTRODUCTION

Aphasia is a language disorder caused by damage or disconnection of the cortical language areas and typically includes impairments in naming, speech fluency, auditory comprehension, speech repetition, reading, and writing. Stroke is thought to account for 85% of all cases of aphasia, although this percentage fails to account for aphasia as a result of progressive neural degeneration (LaPointe, 2005). Aphasia primarily results from strokes in the left middle cerebral artery territory and damage often involves regions such as Broca's area (Brodmann's areas BA 44 &45), Wernicke's area (BA 22), and the arcuate fasciculus, a major white matter tract connecting the aforementioned cortical areas (LaPointe, 2005).

Although many people are unfamiliar with aphasia, it is a common disorder, affecting over one million people in the United States (LaPointe, 2005). Speech language pathologists diagnose and classify aphasia according to the symptoms presented by individual patients, who can be largely divided as having either non-fluent or fluent aphasia (National Aphasia Association, 2010). There are three types of non-fluent aphasia: Broca's aphasia, global aphasia, and transcortical motor aphasia. As the name implies, each type of non-fluent aphasia is characterized by reduced verbal output and irregular prosody. Although auditory comprehension is relatively spared in patients with Broca's or

Transcortical motor aphasia, it is typically not normal. In contrast, patients with global aphasia have severely impaired auditory comprehension (Benson, 1996; Brookshire, 2007; Davis, 2000). It is also important to emphasize that auditory comprehension ability varies within the different aphasia types.

In many cases of aphasia, patients will continue to improve in regard to both speech and language, even years after brain injury (Fridriksson, 2012; Pulvermüller, et al., 2001; Meinzer, et al., 2005). However, if symptoms persist until the chronic phase (approximately 6-12 months post-stroke), the probability for a complete recovery is low (Pedersen, et al., 1995; Poeck, Huber & Willmes, 1989).

Although much neuropsychological research in aphasia has focused on understanding normal brain function, many studies of aphasia have highlighted the differences between aphasic and normal language functioning as a means to better understand mechanistic accounts of aphasia (Fridriksson et al., 2010; Fridriksson et al., 2012). These clinical insights are important, as they may inform rehabilitation strategies. For example, in a recent study Fridriksson et al. found that audiovisual feedback enabled some individuals with Broca's aphasia to produce fluent speech. Fridriksson and colleagues refer to this kind of on-line feedback as 'speech entrainment.'

The goal of the current study is to investigate speech entrainment in individuals with normal speech and language in order to better understand the fluency-inducing mechanism seen in speech entrainment for individuals with non-fluent aphasia. In the following chapters, I will discuss the current prevailing model of language processing as well as past research of speech entrainment and its origins.

1.1 The Dual Stream Model of Language Processing

During the past 150 years, several neuropsychological models of speech processing have been proposed (Lichtheim, 1885; Hughlings-Jackson, 1879; Head, 1921; Geschwind, 1965; Damasio, 1992.). The current prevailing model owes its origins to visual processing literature that observed two functionally and neuro-anatomically distinct dorsal (which encodes spatial information which is crucial for visually guided reaching) and ventral (which is crucial for visual object recognition) pathways (Mishkin, Lewis, & Underleider, 1982; Mishkin, Underleider, & Macko, 1983; Milner & Goodale, 1995; Hickok & Poeppel, 2000, 2004, 2007; Saur & Hartwigsen, 2012). Analogously, the Dual Stream Model of language, first introduced in 2000 by Hickok and Poeppel, describes two pathways that work in parallel to perceive and understand speech (ventral stream) and to implement and monitor speech production (dorsal stream). The dorsal and ventral streams connect anterior and posterior cortical language areas. The ventral stream maps speech sounds to concepts, connecting Broca's area with the auditory cortex and superior Planum Temporale (sPT), via white matter tracts that pass through the extreme capsule (Saur et al., 2008; Saur & Hartwigsen, 2012). The ventral stream is bilaterally organized with computational differences between hemispheres (Hickok et al., 2008). This bilateral representation is supported by Wada studies (Wada & Rasmussen, 1960) demonstrating that individuals with an anesthetized left hemisphere retain the ability to comprehend speech (McGlone, 1984; Hickok et al., 2008). The bilateral organization of the ventral stream is also evidenced in the recovery of different aphasia types. For example, auditory comprehension, mediated by the ventral stream, seems to improve more than speech fluency, processed by the dorsal stream (Basso, 1992). This phenomenon may be supported by bilateral organization of the ventral stream that allows compensation for damage to the left ventral stream, although not necessarily returning it to pre-morbid ability. The dorsal stream, however, is strongly left hemisphere dominant, as evidenced by the persistence of speech fluency deficits after stroke-induced brain damage to the dorsal stream as well as the profound speech production deficits observed when the left hemisphere is anesthetized. The dorsal stream acts as a sensorimotor interface, which binds auditory speech sounds to articulatory motor maps. The dorsal stream connects portions of Broca's area, precentral gyrus, and postcentral gyrus to area sPT, an area thought to be crucial for sensorimotor integration, via the longitudinal and arcuate fasciculi (Saur et al., 2008). This model provides a framework for testing the role of each of these streams in speech-language production tasks, such as speech shadowing.

1.2 Speech Shadowing

Speech shadowing is a task that requires a participant to simultaneously produce speech as they hear it. The objective is to clearly and accurately reproduce the message as closely as possibly to the speech model. By observing deviations and errors during speech shadowing tasks, it is possible to compare the speed of processing of speech across different individuals or conditions. A crucial measurement used in such experiments is the response latency, which is the difference between the presentation of a word in a speech model and the production of the same word by the participant (Marslen-Wilson, 1973). Shorter latencies indicate better performance than prolonged latencies. To clarify, response latency is the difference between the model and subject's production of a specific word. Since the subject is naive to the message, the model is slightly ahead of the subject's production. By measuring this difference, the response latency of the subject has been used to report on the comprehension and production processes of the shadowed message (Marslen-Wilson, 1973, 1975; Marslen-Wilson & Tyler, 1981). Response latency is reported as a positive value when the subject is slower than the model; reflecting the timestamp of the onset of the word minus the model's onset of the same word.

In his seminal paper in *Nature*, Marslen-Wilson was the first to describe response latency and provide evidence that speech shadowing is not a passive speech task, but requires active language processing (1973). Marslen-Wilson examined 65 neurologically healthy individuals shadowing 300 word passages and reported that shadowing performance could classify 2 distinct groups: close shadowers and distant shadowers. Specifically, only 7 individuals were able to shadow at very close latencies (350 milliseconds (ms) or less) while the rest shadowed at more distant latencies (500-800 ms). In follow-up studies, Marslen-Wilson more formally defined close shadowing as response latencies of 250 ms or less (Marslen-Wilson, 1978, 1985). At a normal speaking rate (approximately 150 wpm), the mean syllable duration is 200 ms (Huggins, 1964; Marslen-Wilson, 1973; Marslen-Wilson & Tyler 1981) Since close shadowing occurs at latencies of 250 ms, participants are actually producing words before all the word's acoustic information is available to them, a little more than a syllable's length behind the model. Marslen-Wilson sought to examine whether this repetition based on partial word information affected the integration of the entire message available to close shadowers. In a second experiment, Marslen-Wilson tested participants' recall (7 close, 7 distant) of a 600-word passage. Despite the short latency, close shadowers demonstrated no difference on recall tasks than distant shadowers, and shadowers performed no differently than a listening control group. Other research has also shown no difference between close and distant shadowers on varying levels of recall (Shattuck & Lackner, 1975). These findings indicate that speech shadowing requires active language processing, and that even close shadowers process content information as it is presented, rather than following passively.

Although there were no differences in recall between close and distant shadowers, an error analysis for both groups was completed to further examine differences in processing. Errors were classified as 'delivery,' 'omission', or 'constructive.' Delivery errors included slurring, hesitations, prolongations, and other errors in production that did not change the word. Errors of omission included words that were completely left out during shadowing. Constructive errors occurred when a participant added or changed entire words or changed part of the word to make it a new word. In his 1973 study, Marslen-Wilson found a total of 402 errors made by the 14 participants during the second experiment. Over half of the errors were classified as delivery or omission errors. The remaining 132 constructive errors were analyzed to determine if there was a difference in the grammaticality of the errors, which might suggest different levels of processing between the two groups (close and distant shadowers). No difference was found between the grammaticality of errors. Although close shadowers made more errors, this reduced accuracy was chiefly driven by the significantly larger number of delivery errors, rather than the difference in the number of construction errors. Marslen-Wilson's findings suggest that on-line shadowing, including active analysis of content, allows participants to predict the upcoming message. Even close shadowers are able to use language processing to aid in rapid comprehension and repetition of the message.

1.3 Speech Shadowing And Language Processing

Based on his observations during speech shadowing (1973, 1975), Marslen-Wilson postulated that sentence perception was processed on at least four distinct levels: phonetic, lexical, syntactic, and semantic (Marslen-Wilson, 1975). He described his theory:

> "[S]entence perception is most plausibly modeled as a fully interactive parallel process: that each word, as it is heard in the context of normal discourse, is immediately entered into the processing system at *all* levels of description, and is simultaneously analyzed at all these levels in the light of whatever information is

available at each level at that point in the processing of the sentence."

In a follow-up study, Marslen-Wilson (1975) tested this parallel model by combing two levels of perturbation in a speech shadowing task. The first level of perturbation included an intentional disruption of semantic and syntactic constraints, which would affect high-level sentence processing. The second level of anomaly included an intentional disruption of lexical integrity of individual words to affect lower level processing. One hundred and twenty sentence pairs containing a tri-syllabic word in the second sentence were manipulated along these two levels. For the semantic-syntactic perturbation, the 120 sentence pairs were broken up equally into 3 groups: congruent, semantic, and syntactic. For the congruent condition, the tri-syllabic word in the second sentences was left unchanged. In the semantic perturbation condition, the tri-syllabic word was changed to a word that was syntactically appropriate, but semantically anomalous. For example:

> The new peace terms have been announced. They call for the unconditional *universe* of all the enemy forces.

In the syntactically anomalous sentences, the trisyllabic word was changed to a word that was both semantically and syntactically anomalous. For example:

> He thinks she won't get the letter. He's afraid he forgot to put a stamp on the *already* before he went to post it.

A second manipulation was made to all 120 sentences to examine lower level processing. Within each group of 40 sentences, the tri-syllabic word of the second sentence was either left intact (no word disruption) or was perturbed within one of its three syllables to make a pseudoword. Therefore, within each condition (congruent, semantic, or syntactic), there were 10 unperturbed sentences, 10 with perturbations to the first syllable, 10 with perturbation to the second syllable, and 10 with perturbation to the final syllable.

Thirteen individuals of varying shadowing latencies completed the shadowing task. The results revealed that restoration of the word occurred most often in congruent sentences (no semantic or syntactic adjustment) when the first syllable of the target word was not perturbed. He argues, therefore, that these data support parallel processing. If processing were serial in a bottom-up order, then participants would have shown no difference in performance regardless of perturbation (semantic, syntactic or pseudoword). Bottom-up processing would also make close shadowing particularly difficult since a shadower would only be able to rely on acoustic information as it comes into the system. Conversely, assuming serial processing in a top-down order, participants would have difficulty completing the task with such speed and accuracy because there is little time for top-down processing before the shadower must make a semantic or syntactic decision regarding the message. Incoming speech signals are automatically entered into processing at all levels (phonetic, lexical, semantic, and syntactic) so that information at each level can constrain and guide simultaneous processing and prediction of the upcoming message.

Speech shadowing has proven to be a useful approach for understanding speech recognition and high-level language processing. Marslen-Wilson (1973, 1985) demonstrated that close shadowers were able to mimic speech in real time at the level of the syllable (less than 250 ms response latency). In reaction time studies, 50-75 ms are typically required for a participant to initiate a response. In word recognition and speech shadowing studies (Grosjean 1980; Morton & Long, 1976; Marslen-Wilson & Tyler, 1981), participants demonstrate recognition of a word in context 200 ms after onset. Since 200 ms is typically not enough time for all the acoustic-phonetic information to be available to recognize the word, Marslen-Wilson and Tyler postulated that close shadowers mimic words with syllable level resolution aided by prediction of ongoing high-level processing of the message. The research by Marslen-Wilson (1973, 1975) shows that although some individuals can shadow speech at sublexical rates, they are not merely ‘parroting’ the model based on minimal acoustic-phonetic analysis but are actually understanding and anticipating the sentence as they repeat it. Furthermore, speech shadowing does not support bottom-up processing of lexical items; instead parallel processing of both incoming acoustic information and ongoing contextual analysis aides in word recognition. Marslen-Wilson & Tyler (1981) found that response latencies increased by 60 ms when subjects shadowed semantically anomalous material (jabberwocky) compared to semantically congruent prose. Further, response latencies increased by 95 ms when subjects shadowed scrambled word strings (Marslen-Wilson & Tyler, 1981). These findings provide further evidence that speech shadowing utilizes the context of the sentence to aid in word recognition. Speech shadowing has demonstrated that online language comprehension utilizes both bottom-up and top-down processing.

Marslen-Wilson saw speech shadowing, particularly close speech shadowing, as a unique opportunity to analyze the speech comprehension system (1985). He, along with others, outlined evidence from speech shadowing to suggest a distributed model of language comprehension and word recognition (Marslen-Wilson, 1985; Marslen-Wilson & Tyler 1981). Marslen-Wilson and Tyler (1981) made observations during speech shadowing to support the idea that word recognition is critical for the process of translating the speech signal into a comprehended message and that it is accomplished by "obligatory and automatic central processing" (Marslen-Wilson & Tyler, 1981). In other words, individuals are able to comprehend speech as the semantic and acoustic information becomes available, utilizing both top-down and bottom-up processing. The use of both types of processing allows for language prediction as well as rapid processing and the ability to identify likely motoric/articulatory combinations.

1.4 Speech Shadowing And Rate

Another aspect of speech shadowing that should be evaluated is how varying the presentation characteristics of the message (i.e. the signal quality) may affect participants' performance. Manipulating characteristics such as rate, may give insight into speech shadowing performance by placing increased demands on language processing. Several studies have attempted to determine what effect rate of presentation has on response latency and speech errors. Most speech shadowing studies presented the message between 150 and 160 words per minute (wpm). Lackner and Levine (1975) found normal subjects are able to shadow sentences significantly faster than non-syntactic word lists when the stimuli are presented at 2 words per second (wps). However, they reported an interaction with this difference becoming accentuated as rate increased from 3 through 6 wps. These findings suggest context becomes more critical for speech shadowing at faster rates.

In another study that manipulated speech rate, Carey (1971) introduced the shadowing facilitation hypothesis, which postulates that speech shadowing could actually aid retention as long as the subject monitored his or her speech output and it was identical to the input. To test this hypothesis, Carey compared the retention scores of 36 listeners, who only listened to passages, and 36 shadowers, who shadowed passages at varying rates. Results showed that increasing rate had a negative impact on both listeners' and shadowers' ability to recall words and semantic information from the passages, but shadowers' scores were significantly worse than listeners' at faster rates. Importantly, Carey noted, shadowing becomes a substantially more difficult task at 3 wps than merely listening to a rapid presentation of a passage. This was confirmed by error analysis, which showed that distortions, omissions, and substitutions increased significantly as rate increased. Therefore, speech shadowing performance (judged by errors), and recall (measured by the recall of lexical and semantic information), deteriorated with increasing rate.

Since shadowing performance was negatively correlated to presentation rate, it is important to consider the effects of speech rate on speech comprehension. Carey (1971) was unable to confirm the shadowing facilitation hypothesis as the comprehension differences between listening and shadowing were not sufficiently pronounced. A similar study by Gerver (1974) asked foreign language translators to listen, shadow, and simultaneously interpret. He found that recall was best after the listening condition and poorest after simultaneous interpretation. This led Gerver and others to suggest that simultaneous listening and speaking could negatively impact recall. However, Carey (1971) and Marslen-Wilson (1973) found that meaning can be stored during a speech shadowing task within the same language, at least when presentation rate is relatively slow.

1.5 Speech Shadowing And Neuroimaging

Peschke and colleagues (2009) recently investigated the neural correlates of speech shadowing. The purpose of their study was to establish whether the dorsal stream was engaged in speech shadowing. Twenty individuals participated in functional magnetic resonance imaging (fMRI), during which individuals performed speech shadowing of auditory pseudowords along with either one or many speakers. Activation patterns included bilateral activation of the perisylvian region along with the thalamus and basal ganglia. Although the dorsal stream is thought to be strongly left lateralized, the authors conclude that their results support bilateral involvement of the dorsal stream structures for immediate repetition of pseudowords. It is questionable whether these findings actually support an argument for the bilateral organization of the dorsal stream. There could be several other potential explanations. For example, this experiment relied on the use of pseudowords without context, which may have increased the task difficulty, taxing a broader cortical area. Despite this bilateral activation, these data implicate speech shadowing operates via at least some dorsal stream structures. Further, these imaging studies merely demonstrate bilateral areas are involved during these tasks, but do not identify whether these regions are required by the task (e.g. the right hemisphere activations may be an epiphenomenon of the left activation).

Evidence from Fridriksson and colleagues (2012) also supports the involvement of the dorsal stream in speech shadowing tasks. Diffusion tensor imaging (DTI) and fMRI were collected on 20 healthy control participants. Regions of interests were based on areas of highest activation during fMRI scans while participants completed a speech shadowing task, and probabilistic tractography was performed to elucidate likely connections between the areas. Again, the dorsal stream was implicated, although here, consistent with previous research, only within the left hemisphere.

1.6 Speech Shadowing in clinical populations

Speech shadowing performance has rarely been investigated in individuals with speech disorders. Two important exceptions include studies of adults with penetrating head wounds (Lackner & Shattuck-Hufnagel, 1982) and stroke-induced aphasia, (Fridriksson et al., 2012). This section is devoted only to the former of the two studies as the latter is discussed in more detail along with corresponding neuroimaging data.

Lackner and Shattuck-Hufnagel (1982) compared speech shadowing response latencies during sentence and unrelated word list conditions in controls and individuals who had suffered a penetrating head wound. Although Lackner and Shattuck-Hufnagel stated that none of their participants had obvious persisting language disorders, individuals with left hemisphere and bilateral brain injuries were more impaired on the shadowing tasks compared to controls and the right hemisphere damage group, as judged by longer response latencies. Further, individuals within the left hemisphere or bilateral brain injury groups who experienced initial aphasia after injury showed the most severe impairments during the shadowing task.

For the left hemisphere damage group, it appears that two profiles emerged for individuals as rate increased. To illustrate this pattern, the performances of two individuals were compared. Both individuals had an initial diagnosis of aphasia, although information regarding type and severity was not available and neither had obvious persistent aphasia at the time of study inclusion. One participant, whose initial main difficultly was speech production, experienced decrement in performance as rate of the speech model increased on both sentence and word list measures. His performance on sentences was more resilient to increasing rate than word list performance. This may demonstrate that content facilitated his speech production, highlighting the connection between context and speech shadowing. In contrast, a second participant, whose shadowing performance was approximately the same for both conditions, was more affected by increasing rate on both tasks. Although one could argue that these findings are anecdotal as inferences are drawn from just two individuals, the study illustrates the role of preserved function for each case. It may be that individuals with impaired semantic or syntactic processing perform similarly on both word lists and sentences because they cannot rely on semantics or syntax to the same extent as individuals in whom these functions are relatively preserved. Moreover, it could be the case that patients with damage to the speech comprehension network, the ventral stream, may not be able to utilize content to aid speech shadowing. In contrast, individuals with an impaired speech production network, the dorsal stream, may be able to use speech shadowing to aid production by supplementing the damaged network.

1.7 Unison Speech Production as Treatment

Unison speech production techniques have long been incorporated into treatment approaches for individuals with communication disorders. Unison production of words, phrases, and sentences has been shown to aid speech production in individuals who stutter as well as individuals with apraxia of speech (AOS) or aphasia (Backus, 1945; Max et al., 1997; Kiefte & Armson, 2008; Rosenbeck et al., 1973, Fridriksson et al., 2012). It is important to clarify that speech shadowing was not developed along with these concepts for individuals with communication disorders. Instead, unison speech production is a basic, natural speech production aid during treatment. Speech shadowing was a task that was studied in order to explain the speed and processing of speech. However, these two fields converge, and therefore it is important to discuss each one. Having discussed speech shadowing in detail above, the different techniques of unison speech productions to facilitate fluent speech production in clinical populations will be discussed below.

Unison speech production has been explored in the stuttering literature (Max et al., 1997; Kiefte & Armson, 2008; Kalinowski & Saltuklaroglu, 2003). Choral speech is a speech facilitation technique where an individual who stutters speaks in unison with another person or even with their own voice. Choral speech has been well documented as a fluency inducing condition for individuals who stutter (Kiefte & Armson, 2008). Why exactly choral speech aids fluent speech production in individuals who stutter is not clear (Kiefte & Armson, 2008). In another implementation of unison speech production to facilitate disordered speech, Rosenbeck and colleagues (1973) developed a hierarchical, 8-step treatment for AOS, a motor speech disorder that commonly co-occurs with non-fluent aphasia. They included online mimicry of a speech therapist's production of a word to serve as the highest support level of a detailed cueing hierarchy.

With respect to aphasia, speech shadowing was first studied by Fridriksson et al. (2012), who found that audiovisual speech shadowing aids speech production of individuals with non-fluent aphasia following stroke. Fridriksson et al.(2012) called this effect 'speech entrainment' because the audiovisual model seemed to pull along, or entrain, the Broca's aphasic patients' speech production. Importantly, the mechanisms that permit fluent speech among individuals who stutter and individuals with non-fluent aphasia may be similar (Fridriksson, et al., 2012). The areas for speech fluency include areas of articulation and motor execution (the motor cortex), motor planning (Broca's areas and pre-motor cortex) and timing complex speech motor movements (basal ganglia). The similarity of these two deficits is evidenced by the improvement of speech fluency in both populations by unison speech techniques.

Many studies have focused on improving speech production in non-fluent aphasia (Sparks, Helm, & Albert, 1974; Fridriksson et al., 2012; Szaflarski et al., 2008; Naeser et al., 2005). There have been numerous observations that even individuals with severe aphasia can articulate normally during singing, but cannot produce propositional speech (Hughlings-Jackson, 1864; Backus, 1945; Gertsman, 1964). Backus (1945) reported on group singing as an effective means to improve patients' connected speech. Based on this evidence, Melodic Intonation Therapy (MIT; Sparks, Helm, & Albert, 1974) was developed as a therapy approach to improve speech production for individuals with non-fluent aphasia and relatively spared auditory comprehension. MIT is an intensive, hierarchically structured program that uses intoned (sung) patterns to exaggerate normal melodic content of speech. MIT employs intonation with continuous voicing (singing) of common, short phrases or simple sentences, and during speech production the patient is instructed to move the left hand/arm to tap out the rhythm of the speech (Sparks, Helm, & Albert, 1974). This tapping serves as a pacer, keeping time with the rhythm of the intoned phrase, and may encourage plastic changes in the brain (Schlaug et al., 2008; Vines et al., 2011). Another treatment approach for non-fluent aphasia, script training, is a functional approach that seeks to practice connected speech on a specific topic. The goal of script training is to facilitate participation of individuals with aphasia in personally relevant activities. Scripts are short passages that describe an event, as in telling a story, or facilitate social communication, such as salutations or a chance to advocate on the best way for a communication partner to communicate with an individual with aphasia (Holland et al., 2002; Cherney & Halper, 2008; Cherney et al., 2008). Script training therapy relies on automatization of speech movements using mass practice of a script to increase verbal production of the script (Cherney & Halper, 2008). Most script training studies emphasize using visual and auditory feedback to increase speech production; however, it is unclear how or why this feedback works. As briefly mentioned above, Fridriksson et al. (2012) used fMRI and DTI to support their theory that audiovisual feedback appears to compensate for the damaged dorsal stream in stroke patients, thereby increasing speech output.

Speech entrainment aids in speech production for some individuals with non-fluent aphasia, possibly via a neural-mechanistic process (Fridriksson et al., 2012). This mechanism is particularly apparent among individuals with non-fluent aphasia, where speech production is entrained along with the message of an audiovisual model. This effect does not seem to require intense practice, but generalization after training suggests there is a continuous benefit of training (Fridriksson et al., 2012). In contrast, studies of choral speech in individuals who stutter have shown that choral speech does not provide therapeutic effects after the speech model is withdrawn. Moreover, the effects of choral speech are transient and degrade with continued use (Max et al., 1997). Importantly, speech entrainment may actually provide more central support for individuals with aphasia to compensate for and train the damaged language system. Therefore, the term speech entrainment describes the neural-mechanistic process of inducing speech fluency and the automatic and obligatory processing of the language in a speech shadowing task.

Fridriksson and colleagues (2012) found that speech entrainment increased fluency in some individuals with non-fluent aphasia. In their study, the authors sought to better understand this effect in a series of experiments. Audiovisual speech entrainment was found to provide more benefit to participants with Broca's aphasia than audio-only speech entrainment. Fridriksson et al (2012). compared the number of words participants with Broca's aphasia were able to say under three conditions: spontaneous speech, audio-only speech entrainment, and audiovisual speech entrainment. Individuals were more accurate during the audiovisual condition compared to the audio-only condition. This suggests that the visual component is beneficial and should not, therefore, be underestimated. It is possible that the visual aspects of speech entrainment allow incremental feedforward information of the movement and place of articulators, much like visual cueing in traditional aphasia treatment.

After 6 weeks of treatment, not only were individuals with Broca's aphasia able to say more words, but also the effects of speech entrainment training generalized to spontaneous speech measures (Fridriksson, et al., 2012). Increased fluency was determined based on the number of words produced with speech entrainment compared to the number of word produced during a spontaneous speech measure.

Importantly, individuals with severe AOS did not respond as well to speech entrainment. Fridriksson and colleagues found a negative correlation between AOS severity rating and benefit from speech entrainment. However, individuals whose main deficit was non-fluent aphasia (greater than AOS) were able to speak more fluently during and after training. This suggests that the motor commands for speech and lexical retrieval are relatively intact in these individuals, although not within normal limits. Specifically, it could be that speech entrainment does not compensate for degraded articulatory maps. Speech entrainment may support fluent speech in several ways. The incoming linguistic code is analyzed in real time. High level linguistic processing takes place and allows for prediction of upcoming words in the sentence. This prediction continues to drive the speech mechanism forward by providing a tentative blueprint in the direction of the sentence. If the prediction is correct, its corresponding speech movements have more time to be planned and executed. If the prediction is incorrect, the system adjusts to the perturbation, which, in cases where the adjustment is unsuccessful, may result in a speech error. Subtle, rapid movement of model articulators may also provide useful information for predicting the upcoming word. Speech entrainment may help individuals with Broca's aphasia by not only aiding in linguistic cueing, but likely also by providing audiovisual, prosodic, and temporal cues that entrain speech production.

Fridriksson and colleagues (2012) theorized that speech entrainment works by providing temporal gaiting, like a metronome, for language as well as visual feedback of articulatory movements to increase fluency. The speech entrainment stimuli were produced at a slow yet natural rate and included natural prosody. The authors suggest that this temporal component may have aided the functions of the basal ganglia, which have been shown to play a role in the timing of speech (Stahl et al., 2011). Further, speech entrainment is thought to provide support to the portion of the language network that is commonly damaged in non-fluent aphasia, including Broca's area and underlying white matter tracts (Fridriksson et al., 2012) as well as the superior temporal gyrus, part of Wernicke's area (Fridriksson et al., submitted). In a healthy individual, the dorsal stream (articulation) pulls along the language code of the ventral stream. When the dorsal stream is damaged, as is common in non-fluent aphasia, non-fluency is a product of the loss of entrainment between the two streams. Fridriksson et al. (2012) described this as “language inertia”. However, speech entrainment is able to train fluency by supplementing the damage to the dorsal stream and yoking the ventral stream along with the audiovisual model.

Currently no other research has shown such dramatic restoration of speech fluency to individuals with non-fluent aphasia. While our understanding of the neural mechanisms behind this restoration of fluency is severely limited; the potential to improve the lives of individuals with non-fluent aphasia, given this finding, is immense. Speech entrainment is thought to supplement the damaged dorsal stream by providing individuals with non-fluent aphasia with external temporal gaiting that yokes lexical, semantic, and syntactic processing of the ventral stream and allows for more fluid speech production. Further observations of control subjects may more clearly reveal the processes that are necessary to aid individuals with non-fluent aphasia.

Speech entrainment training may be beneficial in two major ways. First, speech entrainment has been shown to increase fluency as well as number of words spoken in individuals with non-fluent aphasia (Fridriksson et al., 2012). Moreover, this effect generalizes to spontaneous speech. Therefore, speech entrainment may have potential as a therapeutic intervention that allows patients with non-fluent speech to actually practice producing fluent speech. Second, for even those individuals in whom speech entrainment does not generalize to connected speech, speech entrainment could be used as an on-line speech aid. Individuals may benefit from speech entrainment as an augmentative or alternative communication (AAC) aid, for example using videos on their mobile phone to enable speech. In many regards, speech entrainment used as an AAC device is more natural than other approaches since it enables individuals to communicate using their own voice and individualized messages and not computer-generated speech. Many individuals with aphasia have some weakness in their dominant hand; therefore, speech entrainment is easier to implement than relying on other high-tech AAC devices, which can be large and require two hands to carry. Further research with speech entrainment may not only provide therapeutic approaches for those with aphasia, but may also shed new light on normal language functioning.

1.8 Potential Aphasia Profiles

Speech entrainment training is likely most beneficial for individuals with non-fluent aphasia. Fridriksson and colleagues (2012) suggested that individuals with non-fluent aphasia, particularly Broca's aphasia, who have only mild to moderate AOS, may benefit the most from treatment with speech entrainment. They suggest that this may be due to the fact that their articulatory motor maps are generally relatively intact although not necessarily within normal limits. While individuals with Broca's aphasia demonstrated increased fluency after speech entrainment training, it is possible that speech entrainment may even aid some individuals with global aphasia. However, this may depend on the amount of damage to the ventral stream, as indicated above that context and word prediction aid in speech shadowing studies of normal individuals. Fridriksson et al. (2012) suggested that speech entrainment allowed individuals with Broca's aphasia to overcome speech inertia, an inability to bind a language code to articulatory maps. This may also be true of individuals with transcortical motor aphasia, and therefore, they may benefit from speech entrainment.

On the other hand, speech entrainment may not benefit individuals with some fluent aphasia types, such as conduction and Wernicke's aphasia. In these profiles, interruption of phonological encoding may affect the auditory feedback and thus disrupt speech entrainment. It is unlikely that speech entrainment would aid in reducing paraphasic errors or neologisms. For example, conduction aphasia is a syndrome that is characterized by fluent speech with phonemic errors and relatively good auditory comprehension, but an inability to repeat (typically considered to be a disconnection disorder seen following injury to the left arcuate fasciculus). Pilot testing with an individual with conduction aphasia revealed that she was able to produce speech along with the model, but her speech sounded dysarthric with slurring and imprecise articulation, and speech entrainment did not eliminate her language errors. Importantly, for individuals with anomic aphasia speech entrainment treatment could aid in lexical retrieval, potentially strengthening lexical retrieval in sentences. However, there likely would be less emphasis in increasing speech fluency for these individuals, particularly if the dysfluencies in an individual with anomic aphasia are hesitations due to word-finding difficulties (Helm-Estabrooks & Albert, 2004).

1.9 Rationale

Much work is needed to better understand how and why speech entrainment aids individuals with non-fluent aphasia. However, before building upon the effects of the above-mentioned studies, it is necessary to corroborate the findings from studies from 40 years ago with the current evidence on audiovisual speech entrainment. New understanding of speech processing, its neural correlates, and typology of aphasic syndromes may lead to greater insights into the speech entrainment mechanism.

The purpose of this study was to further investigate the speech and language processes that underlie speech entrainment performance in normal individuals. Specifically, the following questions were addressed in two separate experiments: First, to what extent does speech shadowing rely on semantic knowledge versus phonetic/phonological processing? Second, what effect does speech rate play in speech shadowing? This study compared the response latency of participants under different conditions in two experiments by either manipulating the final word in the sentence (Language Predictability Experiment) or speech rate of the speech entrainment model (Speech Rate Experiment). Close speech shadowing performance has been clearly established (response latency 250 ms or less; Marslen-Wilson 1978, 1985). Therefore, for this study response latencies were interpreted in the following manner: A shorter latency reflects complete processing of the sentence stimulus, where prediction of the final word was complete without perturbation. On the other hand, a longer response latency reflects that sentence processing and/or the prediction of the final word was perturbed. Since the stimuli were carefully designed so that the final word is predictable by most college-aged, English speaking participants, the longer processing time of the production of the final word can be directly attributed to the condition of the experiment. We expected longer response latencies to occur when a strong prediction has been violated.

First, it was important to establish what influences the ability to predict the upcoming words in the speech stream. Specifically, to what extent do participants rely on semantic or phonological knowledge for speech entrainment? The currently dominant theory of lexical retrieval is outlined in the Dell Model of Lexical Retrieval (Dell & O'Seaghdha, 1992; Dell et al., 1997), accounting for speech and language errors in individuals with normal language as well as individuals with aphasia. Lexical retrieval is the process by which a semantic concept is translated into lexical and phonetic code for speech production.

Unsuccessful attempts at lexical retrieval result in errors that are generally related to the target by sound or meaning. The Dell model postulates that lexical retrieval is a two-step interactive process. First, the concepts and semantic features of the target word are activated along with competitors that are semantically related to the target word. The activation spreads to the phonological stage *before* a semantic target is selected. Therefore, the phonological forms of semantic competitors are also activated. The semantic concepts with the highest activation cascade down to the phonological level, which further activate the corresponding phonological forms of the target word, the semantically related word, and phonologically related words. Using this model, Dell and colleagues can account for all error types seen in aphasia, including semantic, phonological, mixed paraphasias and neologisms. In the current study, participants will complete a speech entrainment task where the predictability of the last word of a sentence is controlled order to determine the influences of meaning-based and sound-based processing. Selected stimuli either end with a highly predictable word (Block & Baldwin, 2010; Bloom and Fischler, 1980), a semantic foil, a phonological foil, or a pseudoword foil. The main dependent measure is the response latency. Due to the spreading activation from concept to word form, we predict that the latency response for sentences ending with semantic foils will be longer than those of phonological foils, since the perturbation is occurring upstream in the process.

Second, in regards to speech rate, speech entrainment performance has been shown to decrease as rate increases. Observed anecdotally by clinicians, it seems that individuals with aphasia benefit from speech entrainment presented at a slightly slower than normal rate, but that these benefits deteriorate as the rate of speech is further reduced. Since individuals with aphasia likely show a preference for a certain rate (that is not too fast, but not too slow) it is important to further quantify the influence of speech rate on speech entrainment performance. Evidence from therapies in which rate is controlled, such as MIT, suggest that prosody may also play an important role in speech fluency. A recent study by Stahl and colleagues (2011) points to the importance of rhythm, rather than melody, in a speech shadowing task. Seventeen individuals with non-fluent aphasia completed a study comparing speech production under three different conditions: melodic (sung), rhythmic (spoken), and arrhythmic (spoken) speech. Evidence from this study showed no effect for singing greater than rhythmic speech on speech production. Further, evidence showed an effect for rhythmic greater than arrhythmic speech on speech production, particularly for individuals with lesions to the basal ganglia, highlighting the basal ganglia's contribution to speech production. Therefore, rhythm is a contributing characteristic that aids speech production in non-fluent aphasia. A study by Bailly (2003) reported longer response latencies for text-to-speech versus natural prosodic language. These findings suggest that what may be important is not only the content of the speech model but also the manner in which the speech is presented. Therefore in a second experiment, participants engaged in speech entrainment with a speech model where speech rate was manipulated. We predicted that performance would match our anecdotal observations in individuals with non-fluent aphasia, where performance decreases at very slow or very fast rates.

1.10 Hypothesis

In regard to whether semantic and/or syntactic prediction facilitates speech shadowing, we compared response latency of highly predictable sentences with sentence carrier phrases ending in a foil. We hypothesized that language prediction will facilitate speech shadowing performance, demonstrated by a shorter response latency for congruent highly predictable sentences.

Response latency of the congruent sentence served as the control measure and the semantic, phonological, and pseudoword foils can be viewed as perturbations of the control sentence. Therefore, the foil condition with the longest latency compared to the control was assumed to contribute more to speech shadowing performance than the other conditions. Shorter response latency reflected less violation of the predicted target and therefore a lesser role in speech shadowing performance. We predicted that semantic information and the ventral stream would play a stronger role in speech shadowing than phonological information, demonstrated by slower response latencies when semantic predictions are perturbed. If sentences with semantic foils take longer to process, this give evidence that prediction is driven by semantics more strongly than by articulatory gestures.

The second purpose of the study was to determine the relationship between rate and speech shadowing performance. Sentences involving high and low predictability were presented at varying speech rates. We predicted that speech rate would impact speech shadowing. ‘Optimal' rate of presentation were judged as a balance between accuracy and response latency. For individuals with aphasia, it seems that a slow but natural rate is necessary for successful speech entrainment. It is possible that there is an optimal speech rate that minimizes errors and response latency. Anecdotally, this was observed at a stroke center that uses speech entrainment frequently for therapy. During the development of this treatment, a speech-language pathologist observed that at extremely slow rates, individuals with non-fluent aphasia were no longer able to produce fluent speech. This supports the idea that speech entrainment bolsters fluent speech by providing an external mechanism that aids in the temporal propagation of speech movements along the dorsal stream, allowing the language message of the ventral stream to be expressed. This phenomenon needs to be replicated in order to provide more support for the anecdotal evidence from the clinical arena. Furthermore, due to the inherent speech difficulties in non-fluent aphasia and concomitant AOS, it is assumed that the presentation rate of the script should not be faster than normal rate of speech, and indeed may need to be much slower. This study examined the rate of production during speech entrainment for fast and slow rates, which, for this study, was slightly less than a factor of two above (240 wpm) and below (60 wpm) normal speaking rate (150 wpm). There is likely an optimal balance between the time necessary to view the articulators of the model, execute speech motor movements, and maintain the natural prosodic elements that aid in speech perception and prediction.

CHAPTER 2

Methods

2.1 PARTICIPANTS

Forty participants were recruited from the student body at the University of South Carolina. All participants self-reported normal hearing and vision and were native speakers of English with normal speech and language abilities. Participants were psuedo-randomly assigned to one of the two experiments. In all, 20 participants completed the Language Predictability Experiment and 20 participants completed the Speech Rate Experiment.

The study was judged as "exempt" by the University of South Carolina Internal Review Board and therefore, of no inherent risk for participants. Participants were informed about the study procedures and read an IRB-approved letter detailing study procedures prior to completing the study.

2.2 STIMULI

The Bloom and Fischler (1980) and Block and Baldwin (2010) sentence sets were used as a starting point, and from the set, a subset of sentences were selected for each of the study experiments. The stimulus sets for each experiment included sentences with normalized cloze probability (from Block & Baldwin, 2010; Bloom & Fischler, 1980). Cloze probability is defined as the probability of the target word completing a particular sentence stem (Kutas &

Hillyard, 1984). The Bloom & Fischler and Block & Baldwin sentences were developed by surveying college-age students, and asking participants to complete each sentence stem with one word. Cloze probability was determined by the number of responders who agreed on the final word for each sentence stem. Therefore, normed measures were collected for the sentences, ranging from high to low cloze probability. By selecting these items as stimuli, we are able to control the extent to which the sentence endings are predictable by large samples of college-age students. During sentence development Bloom & Fischler and Block & Baldwin reported that clichés were deliberately avoided. Approximately 400 students were surveyed on 398 sentences (Bloom & Fischler study), and 400 students were surveyed on 498 sentences (Block and Baldwin study). The purposes of the Block and Baldwin study included generating additional sentences to expand the stimuli set and to compare the cloze probabilities of the normed sentences from 1980 to those collected in 2010 to verify that cloze probability had not changed in the subsequent 30 years. Block and Baldwin did not find significant changes in the cloze probability from 1980 to 2010.

There were several benefits to selecting these sentences for the current study. First, the Bloom and Fischler sentences have been used in psycholinguistic research for over 30 years. Selecting sentences with normed cloze probability provided normed data on the final word of the sentences, making cloze probability a factor that can be manipulated. Additionally, strong sentence context made it possible to use sentences rather than longer passages in the current experimental tasks. Therefore, more trials could be collected during the experiment. Response latency measurements were recorded at precise points in each sentence offline in an in-house MATLAB graphical user interface (at the onset of the sentences, and onset/offset of the final word of each sentence). For the language predictability experiment, foils of the final word in high cloze probability sentences were generated, including a semantic, phonological, and pseudoword foils.

2.3 Procedure

Participants were pseudo-randomly assigned to either the Language Predictability Experiment or the Speech Rate Experiment. Both experiments were nearly identical; for each task the participant was given the same instructions and practice items. The only difference between the two experiments was the conditions presented. Experimental stimuli were presented using MATLAB and Psychtoolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). During the experiments, participants watched and listened to videos of a model's mouth on a laptop computer situated for optimal viewing and audio comfort for each individual. The stimulus laptop, open to exactly 85 degrees was positioned approximately 24 inches in front of the subject. The experimenter measured the distance of the subject's eyes from the surface of the table. The laptop was then raised until the center of the screen was 15 degrees lower than the height of the subject's eyes. This placed the stimuli directly in the center of line of vision (for guidelines see: http://www.ors.od.nih.gov/). All participants had normal hearing by self-report. The stimuli were presented in a sound booth (IAC

Acoustics Model 40A) over headphones at a constant volume across all participants. The volume was loud but comfortable for individuals with normal hearing. Participants were instructed to speak aloud at the same time as the model and to shadow as closely as possible. The participants completed five, self-paced practice items (not included in the experimental stimulus set) and received feedback from the experimenter after each item. After five training items, all participants were able to follow the experimental instructions to speak along with the model in real time. As sentence stimuli were partially shared between the two experiments, participants were recruited to complete only one experiment in order to eliminate any training effects by previous exposure to the stimuli. Regardless of experimental assignment, all participants completed a self-paced speech entrainment task. All data collection sessions were completed in a soundproof booth to reduce environmental noise. A Dell webcam built-into the presentation laptop (a Dell Precision) recorded participant performance, which was saved and scored offline for response latency, accuracy, and error coding.

2.4 Language Predictability Experiment

The stimuli for this experiment were created by manipulating the final words of sentences from Bloom and Fischler (1908) and Block and Baldwin (2010); this method will now be discussed. In the first experiment, the goal of the foil set was to create 3 foils for each congruent sentence; the congruent sentence was manipulated so that the final word endings were completed by either a semantic, phonological, or pseudoword foil. These foils were designed to reflect the most common error types in aphasia; semantic and phonological paraphasias, and neologisms. Stimulus foils were developed from only the first 124 out of 498 congruent sentences, which constituted the sentences with a cloze probability of .9-1.00. Therefore, this subset reflects sentences for which nearly all people (90-100%) agreed on the final word choice to complete the sentence.

Guidelines for dividing stimulus items into high to low predictability categories based on cloze probability have not been firmly established in the literature. For example, Block and Baldwin (2010) simply ranked probability in equal thirds (high predictability: .66-1.00 cloze probability, medium predictability: .33-.65, low predictability: 0-.32), while Federmeier and Kutas (2001) designated high and low predictability on their own scale based on event-related potential data, where high predictability ranged from .784 to 1.00 cloze probability and low predictability ranged from .17-.784. For this study, the sentences were selected based on cloze probability from the Block and Baldwin stimulus set, and they were selected only within high predictability sentences (.66-1 cloze probability). Through the process of developing the stimulus set, it was necessary to eliminate some sentences from the list of potential stimuli (see rationale, below). After the creation of the stimulus set, the 100 sentences with the highest cloze probability were selected for the Language Predictability Experiment. From the final sentence stimuli chosen, 91% or greater of responders agreed on the final ending of the word. Therefore, even by the standard of the more conservative Federmeier and Kutas classification, all sentences in the Language Predictability Experiment were considered highly predictable.

Highly predictable sentences comprised that final stimulus set, and the final words of each sentence were also replaced with either a) a semantic foil, where the meaning is related but sound structure of the word is changed; b) a phonological foil, where the meaning is changed but the sound structure of the word is similar to the target; or c) a pseudoword foil, where both the semantic and the phonological relatedness are minimized. For the purpose of clarity, the final word of a high cloze probability sentence is referred to as a high target word. An example of a complete set of foils in four identical sentences, with each one ending in a specific manipulation of the high cloze probability target word, is presented below:

She could tell he was mad by the tone of his *voice*. (target, correct)

She could tell he was mad by the tone of his *shout*. (semantic foil, correct)

She could tell he was mad by the tone of his *choice*. (phonological foil)

She could tell he was mad by the tone of his *gaidge*. (pseudoword foil)

It is important to note that the semantic foil represents a possible correct ending to the sentence, but that the cloze probability of the semantic foil is lower than that of the congruent target (represented in the Block and Baldwin stimulus set). Therefore, although the semantic foil was a grammatically correct foil, it was not the most predictable one, based on the Bloom and Fischler (1980) and Block and Baldwin (2010) studies and served its purpose as a semantic paraphasia-type foil for this experiment. The steps taken in the development of these foils are discussed below.

2.5 Development of Semantic Foils

For each of the first 124 sentences, related words were generated for the final word in each sentence. In an attempt to create as many foil options as possible, world knowledge, online word association generators, and thesauri were used to generate a list of potential semantic foils for each sentence. Once the word list had been generated, the semantic relationship of each word was obtained using Latent Semantic Analysis (LSA, http://lsa.colorado.edu/; Folts, Kintsch, & Landauer, 1998), an online tool that measures the semantic association between words. The words with the highest association that were also judged to be syntactically congruent were presented in an online survey using Premier SurveyMonkey where individuals were asked to rank the word options they were most likely to choose to end each sentence. A value of 1 indicated an individual was most likely to use a foil to end that particular sentence, whereas any value greater than 1 indicated they were increasingly less likely to select that word to complete the sentence. Forty individuals completed the online survey. This procedure not only ensured that the semantic foil was grammatically plausible but also provided an approximate cloze probability for each semantic foil option. After the survey was completed, individual responses were compiled. The choice with the lowest average ranking, and therefore most often indicated as the appropriate choice, was selected as the semantic foil (mean: 1.6255, SD: .03632).

Next, all the potential semantic foils were compared to their congruent targets from the original sentences. Transcriptions of the congruent targets and the semantic foils were measured using an online Levenshtein calculator (http://planetcalc.com/1721/). A Levenshtein distance is the minimum number of edits needed to transform one string of variables into another by operations of insertion, deletion, or substitution of a single character (Reichel & Schiel, 2005). Specifically, for this project, Levenshtein distance is the minimum number of phoneme edits required to change the transcription of one word into the transcription of another. As the Levenshtein distance was based on phoneme edits it was possible to calculate a Levenshtein distance for all foils, including pseudowords. A ratio was created of Levenshtein distance over the number of phonemes in the semantic foil (mean: 1.0214, SD: 0.2229), therefore creating a meaningful measurement of percent change between the two words (Reichel & Schiel, 2005). Since the semantic foils were not controlled for number of phonemes or syllable length, it was possible to have greater than 100% change. Semantic foils were primarily based on their semantic relationship to the target, as determined by survey responses; therefore a number of other factors including word length and number of phonemes were not controlled. Any foils with a Levenshtein distance ratio less than .5 (50%) change automatically constituted removal of that sentence and its foils from the list of potential stimuli for the experiment. Further, if any semantic foils were repeated across sentences, one of the sentences was removed from further consideration.

2.6 Development of Phonological Foils

Each congruent target's neighborhood density was collected using the Irvine Phonotactic Online Dictionary (iPhOD; http://www.iPhOD.com/), an online phonotactic dictionary and calculator (Vaden, Halpin & Hickok, 2009). Using the calculator function it was possible to generate a complete list of words that are one phoneme different, or one Levenshtein distance, from the congruent target. Since the purpose of including phonological foils is to provide a close phonological neighbor of the target, only words in the immediate neighborhood density were selected. Further, as the sentences have highly predictable final words (congruent targets), any words that began with the same initial phonemes were removed from the list of potential targets. The rationale behind this was that beginning the word with the same initial phoneme as the congruent target would further lead the participant to make a speech error in favor of the congruent target word. Therefore, the neighborhood density list was reduced to only words with the same rhyme and coda as the congruent targets. To further reduce the list, any neighbor that began with a vowel was removed as a potential phonological foil. The rationale behind this was that only the rhyme and coda were remaining, and therefore these words were actually closer than other neighbors to the congruent target. The remaining foils consisted of items with only substitutions or insertions appearing as the initial phoneme. Any congruent targets that either did not have enough phonological neighbors or only had phonological neighbors that did not meet the criteria set forth above were eliminated, along with their corresponding congruent sentences, from the list of potential stimuli.

To ensure that the phonological foils had little or no semantic association to the congruent targets (and therefore the corresponding semantic foils); word associations were collected for each congruent target and the remaining phonological neighbors (mean: .0355, SD: .0625). The phonological neighbors that had the least association were selected. In the case that two different sentences yielded the same (but less associated) phonological neighbor, the phonological neighbor with the next lowest association was selected. This was done to ensure that no phonological foils were repetitions of congruent target words or other phonological or semantic foils. The selection of phonological foils was significantly less associated to the congruent targets than the semantic foils, *t*-test (two-tailed, paired), $p<0.001$). Therefore, all phonological foils are unique (not repeating any semantic foils or congruent targets) and one Levenshtein distance from congruent targets. Essentially, the phonological foils are perfect rhymes of the congruent targets and have little semantic relevance in the sentence.

2.7 Development of Pseudowords

The previous foils discussed reflect both semantic and phonological errors that might be typical in the speech production of patients with aphasia. Every attempt was made to create foils that singly reflected either a semantic relationship or phonological relationship to the high cloze sentences. However, it is impossible to completely uncouple phonology from semantics, and semantics from phonology, in the case of real words. Therefore, as an attempt to compare the perturbations of phonological and semantic foils and to reflect the neologistic errors made by individuals with aphasia, a set of pseudoword foils was made. Pseudowords were made using the bank of pseudowords from the iPhOD dictionary. Changing one phoneme in an English word created these pseudowords, therefore these foils follow English phonology conventions. The Levenshtein distance was employed to compare the pseudoword foils and the congruent targets (mean: .97, SD: 0.0867). Therefore, the pseudoword foils make up a set of words with unique phonology from highly predictable congruent targets, and absolutely no semantic association to the congruent targets. However, there was one similarity between congruent target words and their corresponding pseudoword foils. All pseudowords matched the C-V structure of the targets as an attempt to control articulatory complexity. In other words, targets and pseudowords matched in number of phonemes and syllable length. However, there is no way to predict the pseudoword foil prior to its audiovisual presentation.

2.8 Speech Rate Experiment

In this portion of the study, individual sentences were presented at five rates: 60, 105, 150, 195, 240 wpm. Normal speech rate is approximately 150-160 wpm (Marslen-Wilson, 1973). Audiobooks are recorded at approximately the same rate (Williams, 1998), and most speech shadowing studies use this rate as well (Marslen-Wilson, 1973, 1975; Marslen-Wilson & Tyler, 1981; Lambert, 1992). During recording of the speech models, the speaker produced the sentences at a consistent, normal rate. After recording, individual wpm were calculated for each sentence. Using FFMPEG (www.ffmpeg.org), a computer software program, the audio and video files were then manipulated to achieve the other four rates while maintaining fundamental frequency. The program utilized the *waveform*-similarity-based *synchronized overlap-add* (WSOLA) algorithm to manipulate the playback of the audio in the stimuli, and we used presentation time stamp adjustments to manipulate the playback of the video stream in the stimuli. This method provided a wide range of temporal manipulations with little to no degradation in the stimulus set. The program manipulated rates at precise factors and, therefore, allowed for the creation of scripts including almost any possible speech rate. For example, if, during recording, the model spoke at a rate of 150 wpm, the speed of the audio and video files were multiplied by a factor of 1.6 to create a rate of 240 wpm. The sound signal became distorted at extremely fast or slow rates, making our upper and lower bounds 240 wpm and 60 wpm, respectively. At extremely slow rates the model's voice was distorted by an unnatural pitch that could not be corrected. At extremely fast rates, the model's speech sounded clipped and short, functional words began to lose important acoustic information, while some were lost altogether beyond 270 wpm. During data collection, participants were asked to shadow the model as closely as possible. Unlike the previous experiment, half of the sentences had high cloze probability, and half had low cloze probability in the Speech Rate Experiment. According to the results of Bloom and Fischler (1980) and reproduced in Block and Baldwin (2010), half of the sentences had greater than 94% agreement of final word response, and the remaining half had 48% or less agreement on the final word. The only manipulation of these sentences was the rate of their presentation. Therefore, the sentence stimuli for the Speech Rate Experiment were the same sentences as presented in the Block and Baldwin paper (2010) and the language was not manipulated. Accordingly, the lower cloze probability sentences reflected a near chance probability of predicting the last word of the sentence. The data were analyzed both for effect of rate as well as cloze probability on speech entrainment performance.

2.9 Scoring Responses

During each experiment, subject performance was recorded from a webcam on the laptop. Using an in-house MATLAB program, the audio files were extracted from the video. Time points for the sentence onset and final word onset and offset were marked on the audio file presented in another in-house MATLAB program. This scoring program allowed for a direct comparison between the subject's speech production and the speech model. The audio files were presented visually as a waveform alongside the waveform of the model sentence used in the experiment. The difference of the onset and offset of the final word was calculated to obtain a duration value of the final word. As well as displaying time point data, the scoring program provided opportunity for the scorer to transcribe the subject's final word production in a free-text box, code errors, and record the correct number of words produced during the subject's production of the sentence. This final measure was used as a global measure of how well the participant reproduced the sentence. The scoring program was used to record and save these scores that were later entered into Microsoft Excel and SPSS (SPSS, Inc., Chicago, III). All transcripts were written using the Carnegie Mellon University Pronouncing Dictionary code (CMU; Weide, 1995), to account for the lack of many International Phonetic Alphabet symbols on a standard keyboard.

Twenty participants completed the Language Experiment while another twenty participants completed the Rate Experiment. Participants' performance was collected via auditory recordings. Data scoring involved marking the onset in time of the final word of the sentence. The main dependent measure, response latency of the final word of the sentence, was calculated by subtracting the onset time of the final word of the participant from the final word onset time of the model, which was displayed during the experiment.

For each sentence trial, each participant's performance was recorded and stored as an audio data file. Individual audio files were loaded into a MATLAB graphical user interface program (see Figure 2.1). Colored markers on the participant waveform were manipulable to allow the scorer to mark specific points. Markers indicated the sentence onset, final word onset, and final word offset. The model sentences had been scored previously, and displayed the markers at each location described above. The scorer was able to move the makers along the subject waveform and listen to the entire audio file or clips of the audio between the markers. After positioning the markers into their proper position, the scorer coded the correctness of the final word of the sentence. If the word was incorrect, the scorer could check boxes to indicate the type of error made and transcribe the erroneous production. Finally, the scorer listened to the participant's full production of the sentence stimulus and then counted the number of correctly produced words in each participant's attempt during the sentence.

After scoring in the graphical user interface was complete, the data was processed with Microsoft Office Excel to sample times as well as calculating response latency. The response latency was calculated for each subject for each sentence trial by subtracting the model's final word onset from the subject's final word onset. Mean response latency, excluding incorrect final word trials, was calculated for each subject for each condition. Global sentence accuracy was determined by counting the number of words that were said correctly for each sentence. A ratio was calculated in Excel for global sentence accuracy by the quotient of number of correct words over number of words in the sentence, this was a way to represent the accuracy of the participant's ability to speak along with the model. The mean global sentence accuracy was calculated by block by experimental trial order as well as by condition. Finally, when the last word of the sentence was said incorrectly, the participant's response was error coded and transcribed. Error patterns were collected for individuals by condition and summed to reflect the error patterns of the group by condition

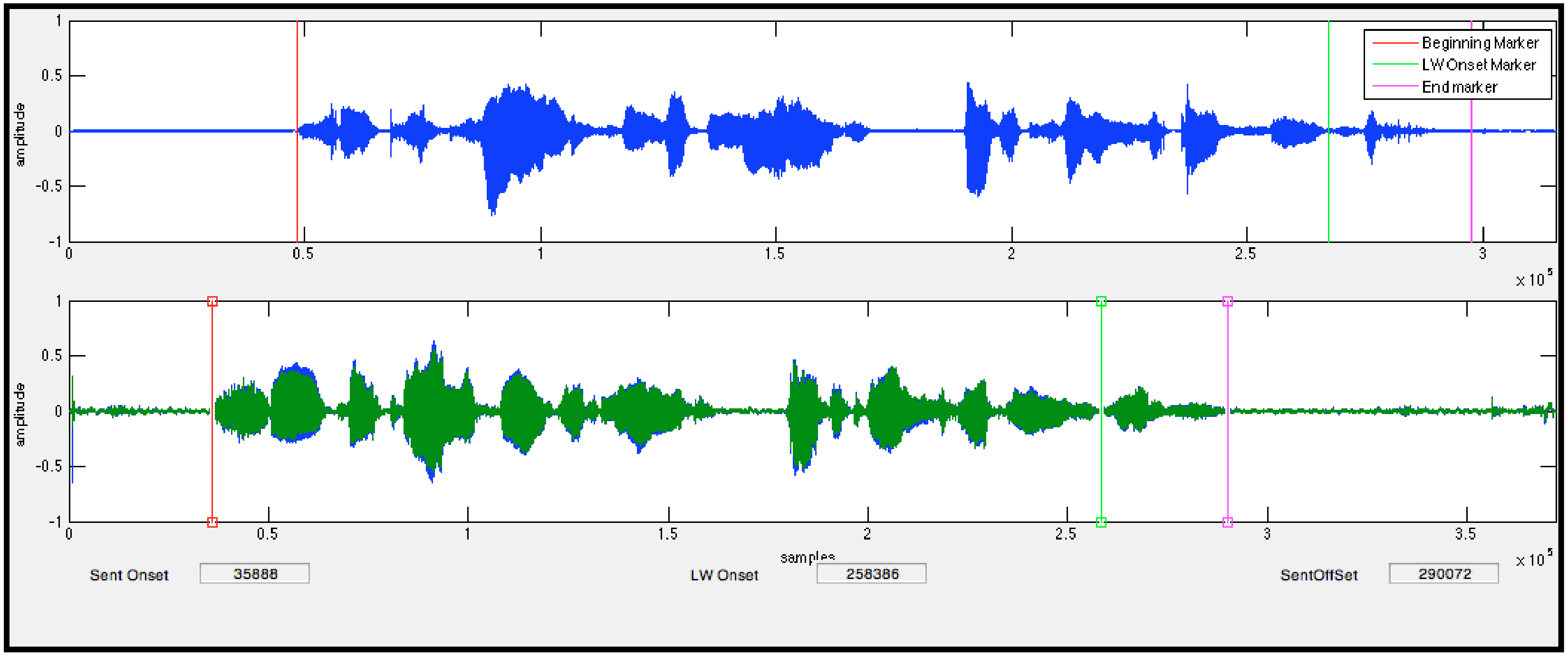

Figure 2.1: A screenshot of the MATLAB graphical user interface that was used to score the audio data. The top pane is the waveform of the model, which was presented in the experiment. The bottom pane is the performance of the participant during the trial. The buttons to the right of each pane would play all or portions of the audio file. The red, green, and magenta bars on the lower pane were manipulable, and the scorer positioned them to mark the sentence onset, final word onset, and sentence offset respectively; these bars were fixed on the model sentence. Error coding for the final word was recorded by checking appropriate box(es) to describe the error. If the final word was correctly produced then “LW correct” was the only box checked. Global sentence accuracy was recorded in a free text box, where the scorer inputted the number of words correctly produced in the sentence. Lastly, the final word was transcribed in CMU transcription code in another free text box.

CHAPTER 3

RESULTS

3.1 LANGUAGE PREDICTABILITY

In order to determine the role of language predictability on speech entrainment performance, the mean response latency of high target sentences was compared to the mean response latency of semantic, phonological, and pseudoword foils. The means and standard deviations are presented in Table 3.1. To analyze the associations between response latency among the four language predictability conditions, a correlation matrix was computed using the mean response latencies for each condition for each participant.

Significant correlations were found between all four conditions, all with $p<.001$ (Table 3.2), meaning that individuals who had faster response latencies in one condition, tended to have faster response latencies in the other conditions. Conversely, those who had slower response latencies in one condition, tended to be slower in other conditions.

Table 3.1: The response latency mean (in seconds) and standard deviation for conditions.

	N	Mean	Std. Deviation
TH	20	0.277	0.141
SE	20	0.364	0.177
PH	20	0.348	0.163
PS	20	0.374	0.182

High Target (TH), Semantic (SE), Phonological (PH), and Pseudoword (PS).

Table 3.2: Pearson correlation matrix between conditions.

	TH	SE	PH	PS
TH	1	--	--	--
SE	.977**	1	--	--
PH	.980**	.991**	1	--
PS	.971**	.996**	.989**	1

High Target (TH), Semantic (SE), Phonological (PH), and Pseudoword (PS).

**. Correlation is significant at the 0.01 level (2-tailed).

A repeated measures ANOVA was calculated to assess whether there were differences between the average response latency of the four sentence types. Assumed sphericity and Greenhouse-Geisser correction for sphericity revealed no change in significance. Therefore, although nonhomogeneity and unequal variance violations were present and since the sphereicity corrections did not change significance, these violations were judged not sufficiently serious to warrant further investigation. Results indicated that participants performed at varying response latencies on different sentence types, F (3,19)=52.193, p<.001. Visual examination of the response latency means, reported in Table 3.1, suggest that participants had shorter response latencies during high target and phonological foil sentences and longer response latencies during semantic or pseudoword foil sentences. To analyze whether there was a significant difference between the last word response latency across the four sentence types, a Bonferroni pairwise comparison was completed (p=.05/6=.0083; See Table 3.3 and Figure 3.1). Only four comparisons were found to be statistically significant,

Table 3.3: Results of the pairwise comparison for mean response latency by sentence type.

(I) Response Latency	(J) Response Latency	Mean Difference (I-J)	Sig.
TH	SE	-0.088*	.000
	PH	-0.071*	.000
	PS	-0.098*	.000
SE	PH	0.017	.076
	PS	-0.010	.107
PH	PS	-0.027*	.008

High Target (TH), Semantic (SE), Phonological (PH), and Pseudoword (PS).

Asterisk (*) indicates significance following Bonferroni correction (p=.05/6=.0083).

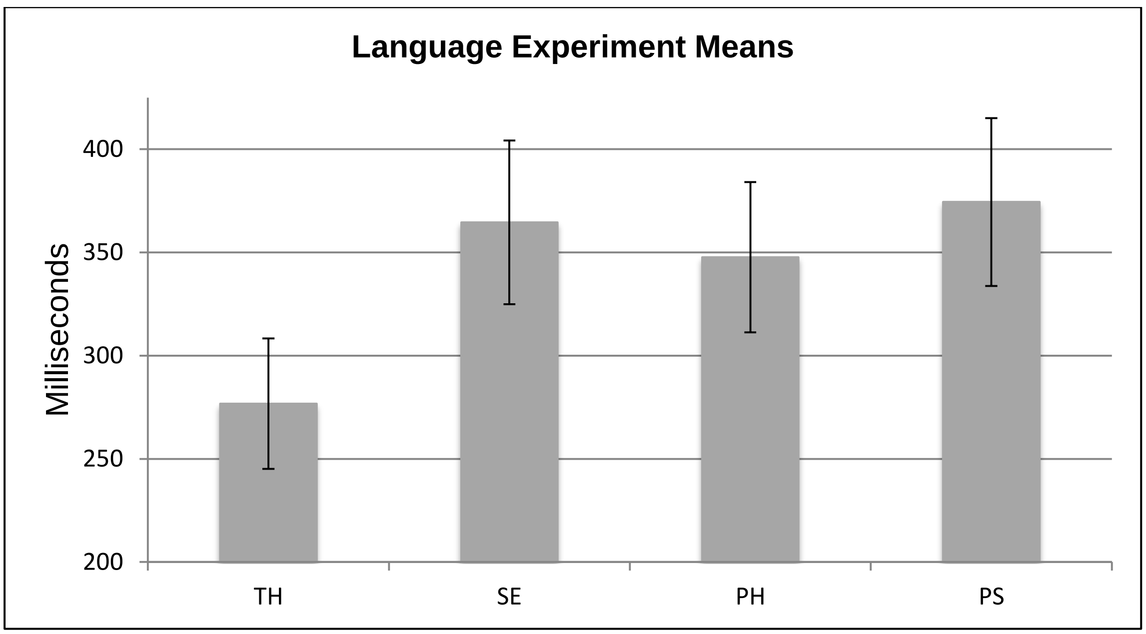

Figure 3.1: The mean response latency for each condition, error bars represent standard error.

high target (TH) was faster than all other conditions ($p<.001$) as well as phonological (PH) faster than pseudoword (PS) ($p=.008$).

3.2 Effects of Transformations

Since the response latency data were positively skewed and not normally distributed, a logarithmic transform was performed in order to normalize the data and thus increase the power of the statistical analysis. However, the transform did not reveal any change in statistical significance between conditions in the pairwise comparison. Therefore, the results for the untransformed data will be discussed for ease of explanation.

3.3 Secondary Hypothesis

An *a priori* planned comparison between the semantic (SE) and PH conditions was carried out. As stated in the Introduction, the literature postulates a two-step, interactive lexical retrieval model (Dell, et al., 1992;1997). To test the if there was a difference in response latency between the two theoretical steps (semantic and phonological) from the Dell model of lexical retrieval, a paired-samples t-test was conducted to compare the response latencies in the SE and PH conditions. There was a significant difference in the response latency for SE (0.365s; SD 0.177) compared to PH (0.3476s; SD 0.1630) conditions, $t(19)= 2.753$, $p = .013$.

3.4 Analysis of Errors

In order to compare the types of errors in each condition, the error codes for each trial were collected. During the entire experiment, including all four conditions, 1788 errors were recorded. Errors could be divided into either a language error, including semantic, phonemic, or neologism; or any articulation error, such as substitution, omission, or addition. Since it was possible to restore the TH word to the sentence, restoration was also included as a language error type. Additionally, since it was possible to make more than one type of error for a word, combinations were also included in the articulation errors. Thus, errors fell into six different error types within two categories: 1. Restoration and semantic errors (language errors); and 2. addition, omission, substitution, and mixed errors (articulation errors). For each participant, errors were counted for each condition and also summed for total errors made during the experiment. Then, the individual profiles by condition were added together across participants, so that a total error count for each condition was tallied.

The High Target (TH) condition was viewed largely as a control condition, where only 110 total errors (6%) were made. Fifteen of the twenty participants made four or less errors during the TH condition. Because of the low number of errors, the TH error patterns were not considered, leaving the three experimental conditions, accounting for 94% of the errors made during the experiment: Semantic, Phonological, and Pseudoword. The number of errors produced in each condition is presented in Figure 3.2. Individual error profiles were constructed by counting the number of times a specific error occurred in a given condition across participants. To compare the error types by condition, language and articulation errors were investigated in individual mixed ANOVAs.

For the language errors, a mixed ANOVA where the between-subject factor was error type (restore, semantic) and the within-subject factor was foil

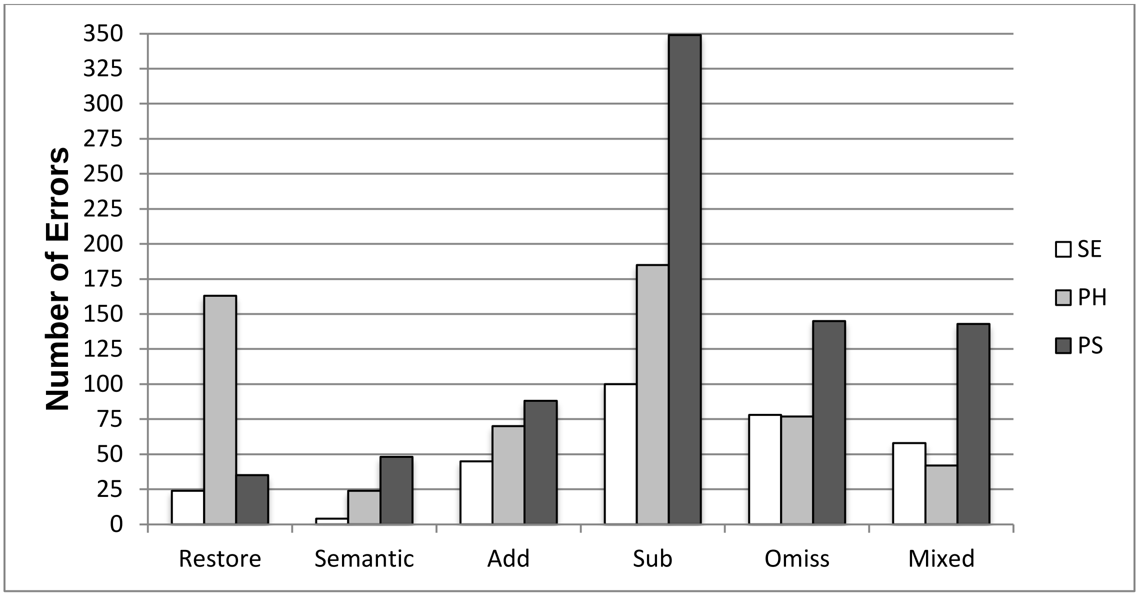

Figure 3.2: The count of error type by condition; Semantic (SE), phonological (PH), and pseudoword (PS).

condition (SE, PH, PS), a found main effect of error type $F(1,19)=8.850$, $p=.008$, and foil $F(2,38)=23.029$, $p<.001$ was found. The interaction was also significant, $F(2,38)=19.759$, $p<.001$. Figure 3.3 displays the means of the language ANOVA. Paired t-tests, with Bonferroni correction (.05/3= .0167), were run to determine if there was a significant difference between the foil types within the error types. Significance differences were found among the conditions in the restore error type. Specifically, the PH condition was statistically different from both the SE and PS conditions, with $t(38)=4.63$, $p=.01$ and $t(38)=4.31$, $p<.001$, respectively.

To investigate the articulation errors, a mixed ANOVA where the between-subject factor was error type (addition, omission, substitution, and mixed) and the within-subject factor was foil condition (SE, PH, PS) was performed. A main effect of error type $F(3,48)=11.881$, $p<.001$, and foil $F(2,32)=62.664$, $p<.001$ was revealed. The interaction was also significant, $F(6,96)=12.941$, $p<.001$, driven

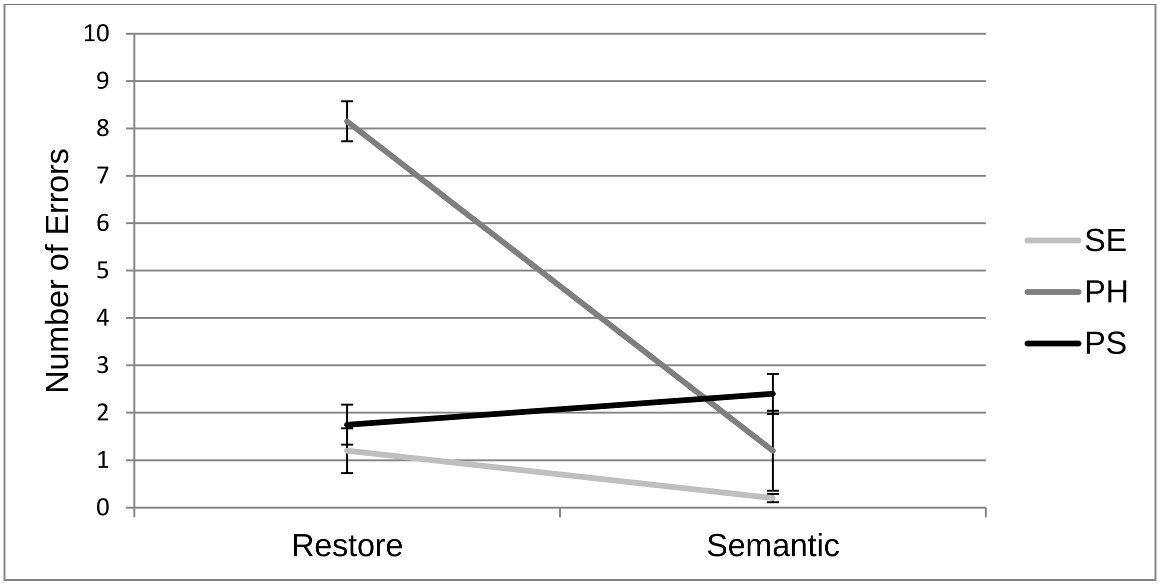

Figure 3.3: Mean and standard error for the language error ANOVA.

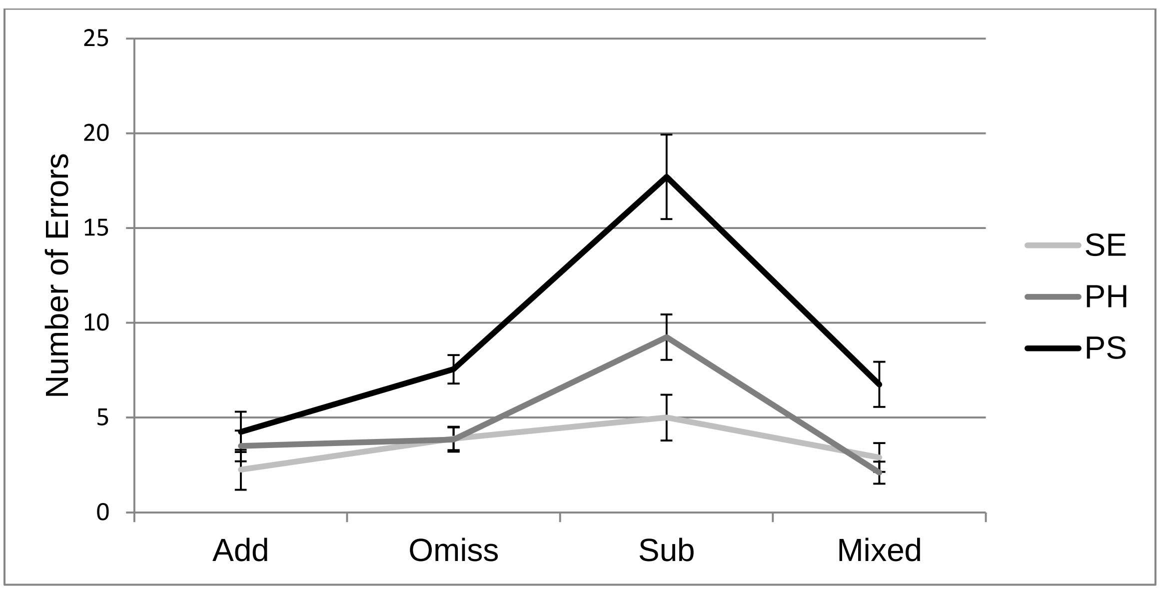

Figure 3.4: Mean and standard error for the articulation error ANOVA.

mostly by the substitution errors. Figure 3.4 displays the means of the articulation ANOVA. Pairwise comparisons of foil type revealed that significantly more errors were made in the PS condition than in SE or PH, with $p<.001$ and $p<.001$, respectively. The comparison of the other two conditions was not significant.

3.5 SPEECH RATE

In order to determine the role of speech rate on speech entrainment performance, the response latency of varying speech rates was compared. The speech rates investigated were; 60 wpm, 105 wpm, 150 wpm, 195 wpm, and 240 wpm. The mean response latency for each speech rate condition is presented in Table 3.4.

To analyze the associations between response latency among the five speech rate conditions, a correlation matrix was computed using the mean response latencies for each condition for each participant. Significant correlations were found across all five conditions, all with $p<.001$. The r values are presented in Table 3.5. Pearson correlation coefficients revealed a strong relationship between response latencies across all five conditions, meaning that those who had faster response latencies in one condition, tended to have faster response latencies in the other conditions. Conversely, those who had slower response latencies in one condition, tended to be slower in other conditions.

Table 3.4: The response latency (in seconds) mean and standard deviation for each condition, 60, 105, 150, 195, and 240 wpm.

	N	Mean	Std. Deviation
60	20	.275	.245
105	20	.238	.227
150	20	.225	.211
195	20	.222	.192
240	20	.253	.211

Table 3.5: Pearson correlation matrix between conditions, 60, 105, 150, 195, and 240 wpm.

	60	105	150	195	240

60	Pearson Correlation	1	--	--	--	--
105	Pearson Correlation	.996**	1	--	--	--
150	Pearson Correlation	.991**	.995**	1	--	--
195	Pearson Correlation	.971**	.971**	.984**	1	--
240	Pearson Correlation	.915**	.909**	.930**	.977**	1

**. Correlation is significant at the 0.01 level (2-tailed).

A repeated measures ANOVA was conducted to assess whether there were differences between the average response latency of the five speech rate conditions. Assumed sphericity and Greenhouse-Geisser correction for sphericity revealed no improvement to statistical significance. Therefore, although nonhomogeneity and unequal variance violations were present, they were judged not serious enough to warrant further investigation. Results indicated that participants did perform differently on different speech rates, $F(4,19)=4.474$, $p=.003$. The means and standard deviations are presented in Table 3.4. Examination of these means suggest that participants responded faster to sentences presented near normal speaking rate, and slower to both slower and faster rates. To analyze whether there was a significant difference between the five speech rate conditions, a Bonferroni pairwise comparison was carried out ($p=.05/10=.005$; See Table 3.6 and Figure 3.5). The only means that had a significant difference were, 60-105 ($p<.001$) and 60-150($p=.001$).

Table 3.6: Results of the pairwise comparison for mean response latency by sentence type, 60, 105, 150, 195, and 240 wpm.

(I) LWRL	(J) LWRL	Mean Difference (I-J)	Sig.
60	105	0.037*	0.000
	150	0.050*	0.001
	195	0.052	0.054
	240	0.022	1.000
105	150	0.013	0.410
	195	0.016	1.000
	240	-0.015	1.000
150	195	0.003	1.000
	240	-0.027	1.000
195	240	-0.030	0.095

(*) indicates significance following Bonferroni correction (*p*=.05/10=.005).

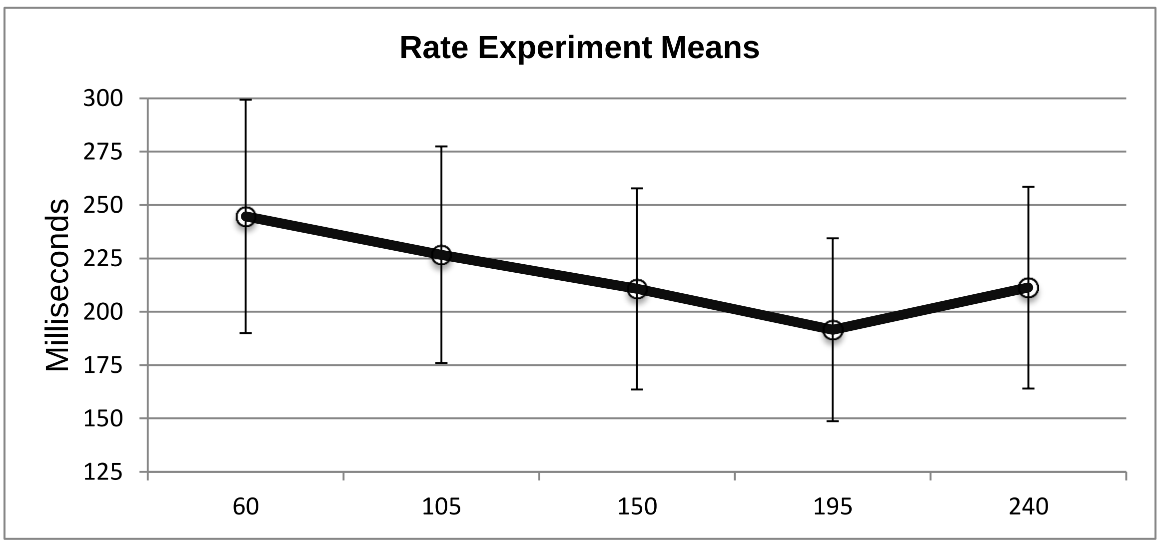

Figure 3.5: The mean response latency and standard error for each condition.

3.6 EFFECTS OF TRANSFORMATION

Because the response latency data were positively skewed and not normally distributed, a logarithmic transform was performed in order to normalize the data and thus increase the power of the statistical analysis. As before, because the transform did not reveal any change to significance or interpretation between conditions in the pairwise comparisons, the results will be discussed using the original data for ease of explanation.

3.7 Secondary Hypothesis

The secondary hypothesis that speech rate would affect global sentence accuracy was also investigated. Since the study was presented in a block design and the first presentation of the sentence within each block was balanced by condition, only the global sentence accuracy of the first sentence within each block was used, totaling 80 sentences per participant.

In order to determine the role of speech rate on speech entrainment performance, the global sentence accuracy of varying presented speech rates were compared. The mean global sentence accuracy for each condition was 0.980 (SD 0.020), 0.983 (SD 0.016), 0.969 (SD 0.026), 0.956 (SD 0.034), and 0.909 (SD 0.051), respectively (See Table 3.7).

Again, a repeated measure ANOVA, with Greenhouse-Geisser correction, was conducted to assess whether there was an effect of global sentence accuracy for the five speech rate conditions. Results indicated that participants did perform differently at different rates, F (4,19)=31.077, p<.001 (Table 3.8). The means and standard deviations are presented in Table 3.8. Examination of these

Table 3.7: The sentence accuracy mean and standard deviation for each condition in wpm.

	N	Mean	Std. Deviation
60	20	.9798	.02006
105	20	.9826	.01629
150	20	.9686	.02579
195	20	.9559	.03363
240	20	.9091	.05136

means suggest that participants more accurately entrained the sentences at slower speech rates. To analyze whether there was a significant difference between the five speech rate conditions, a Bonferroni-corrected pairwise comparison was done. Several mean global sentence accuracies were found to be significantly different (See Table 3.8), which continued to support the overall assumption that increasing speech rate more greatly affected global sentence accuracy (60-195 (p=.015), 60-240 (p<.001), 105-195 (p=.010), 105-240 (p<.001), 105-240 (p<.001), 195-240 (p=.002). These results are presented in Table 3.8 and Figure 3.6.

Effect size was calculated for each comparison. Because the conditions were highly correlated, the effect size calculations included controlling for the correlation (see Cohen, 1988*,* page 49). The resulting effect sizes ranged from small (d<0.40): for the comparisons: 60-240 and105-240, to large (d>0.80) for

Table 3.8: Results of the pairwise comparison for mean sentence accuracy by sentence type, 60, 105, 150, 195, and 240 wpm.

(I) SentAcc	(J) SentAcc	Mean Difference (I-J)	Sig.
60	105	-0.003	1.000
	150	0.011	0.397
	195	0.024*	0.015
	240	0.071*	0.000
105	150	0.014*	0.046
	195	0.027*	0.010
	240	0.073*	0.000
150	195	0.013	1.000
	240	0.059*	0.000
195	240	0.047*	0.002

Asterisk (*) indicates significance following Bonferroni correction(p=.05/10=.005).

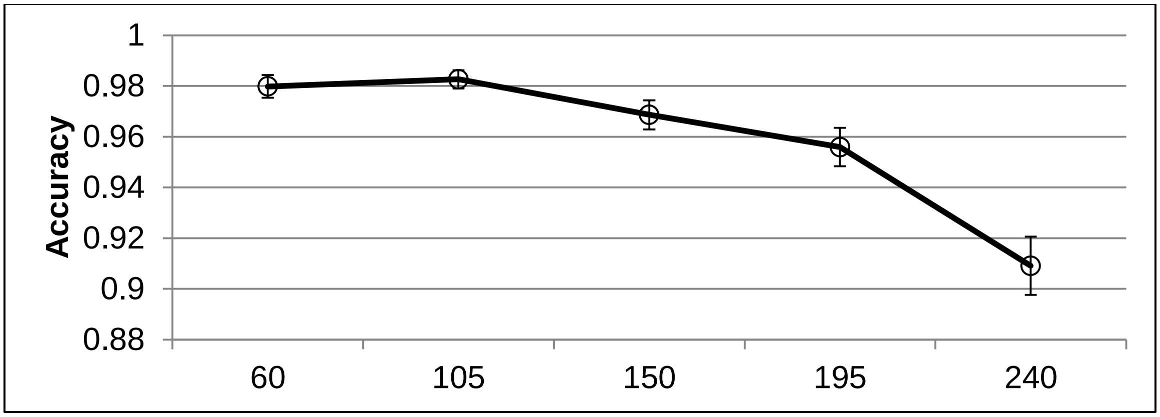

Figure 3.6: The mean sentence accuracy for each condition. Error bars represent standard error.

the comparisons: 60-105, 60-150, 60-195, 105-150, 105-195, 150-240, 195-240. Only one comparison, 150-195, produced an effect size less than 0.20, suggesting that performance, measured by response latency, for these conditions is very similar.

3.8 Time Course Information

Performance during the entire course of the experiment was also monitored. For both experiments, response latency was compared based on the order of stimulus presentation. Although the sentences were presented randomly for each participant, the presentation order was recorded in order to see if there were any changes in participant performance across the length of the experiment (e.g. training or fatigue effects). Figure 3.7 shows the mean response latency for each blocked trial across all 20 participants (for each experiment) in order of stimuli presentation. Note the initial drop in response latency after the initial trial for the Language Experiment, suggesting that performance on the first trial was slower than those following. This improved performance is not visible in the Rate Experiment. This may be due to the speed of the presentation of the Rate Experiment or that the sentences within blocks did not vary. In the Rate Experiment, some first trials were presented at a very slow rate while others were faster. Since our analysis above revealed an effect of speech rate on response latency, it is possible that averaging across these different presentation speeds may have washed out the effect of the first trial. Visual inspection suggested no obvious change in performance across time in either experiment; therefore, any effect from training or fatigue was not further considered. A one tailed *t*-test comparing the response latency of the initial trial block to the second trial block for the Language Experiment was performed, $t(19)=$, $p=.062$. There was not a significant difference between the initial and the following trial, and therefore removing the initial blocked trial from any of the analyses was not warranted.

3.9 Effects of Cloze Probability

To determine whether there was an effect of cloze probability on response latency, a *t*-test was performed on the averaged performance for each sentence across all participants for the normal speaking rate (150 wpm). The normal speaking rate condition was selected for this analysis to reduce any effect of condition on the results. Response latency on the last word for high cloze probability sentences was expected to be faster than for low cloze probability sentences. Surprisingly, the mean last word response latency for low cloze probability (0.209 s) was faster than the mean last word response latency for high cloze probability (.226 s); however, this difference was not statistically significant (p=.086*).*

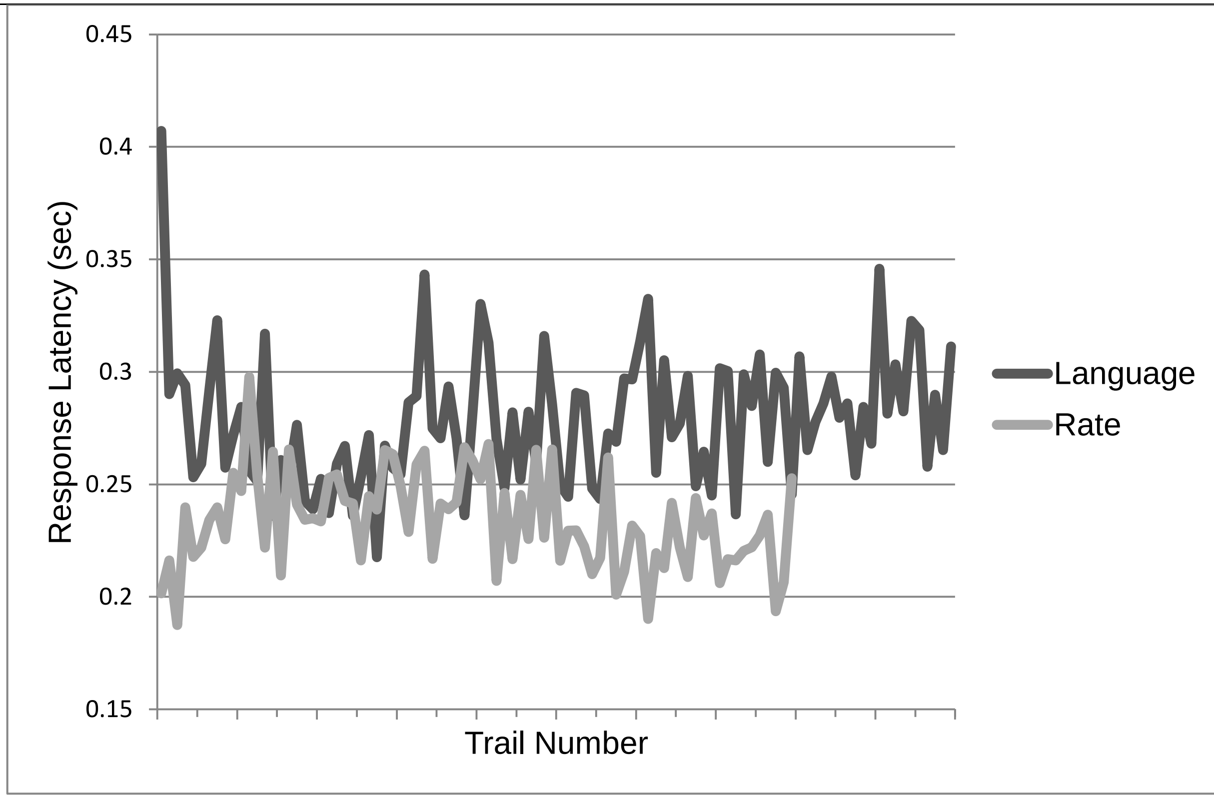

Figure 8: Blocked trial order: The average response latency of trial blocks across all participants for each experiment in order of stimuli presentation. The dark line represents the time course information from the Language Experiment and the light line represents the time course information from the Rate Experiment. Note the Rate Experiment consisted of fewer blocked trials, but more conditions within each block.

CHAPTER 4

DISCUSSION

The results of the Language Experiment support the findings of Marslen-Wilson (1973, 1975), showing that shadowers process the language of incoming messages as they hear and repeat it. First, the TH condition was faster than all experimental conditions suggesting that online processing of syntactic and semantic information facilitates speech entrainment performance. Further, speech entrainment performance is disrupted by violations of feed forward predictions based on meaning, as demonstrated by the experimental conditions. For example, when the feed forward prediction is violated, even with a semantically congruent, but less predictable foil, there is increase in response latency. Semantic foils slowed response latency more than phonological foils, suggesting it was easier to amend to the phonological foil because they share phonological and articulatory features. Therefore, when the foil shares a phonological neighborhood with the predicted word, it results in shorter response latencies. This supports the Dell model of lexical retrieval (1992; 1997), which posits that semantic concepts are initiated first and then the phonological forms of the semantic concept (and its neighbors) are activated. The cascading and interactive effects between these two levels are supported in response latencies of the current experiment.

The phonological foils were also faster than pseudowords foils, which were matched for phoneme structure and articulatory complexity, suggesting that further perturbation to the predictability of the last word in the sentence is increasingly detrimental to speech entrainment performance even when articulatory complexity is controlled. The PS condition is the opposite of the TH condition, in that the PS shares no relationship to the sentences or the mostly likely word choice (or the final word in a sentence. In the case of PS foils, there is no facilitation from online processing of semantic syntactic or phonological information. Therefore the PS condition represents speech entrainment performance that may be comparable to single word repetition, which is based on only acoustic level information. This shows a facilitation effect for words with similar phonology, reflected in the second step of Dell's model of lexical retrieval. This suggests that it is easier to change the phonological forms slightly to an incongruent word than to violate highly anticipated words for other less likely, albeit congruent, words or to an anomalous pseudoword.

Although not significant within the ANOVA, the SE and PH *t*-test comparisons were singled out based on the steps of Dell model for lexical retrieval. The PH condition was faster than the SE condition, suggesting that phonological information is somewhat more vulnerable to perturbation than semantic information. This may be a result of lexical retrieval initiating motor speech plans of the most activated word automatically, which is supplemented by the online prediction of the final word of the sentence. In other words, speech entrainment is driven more by meaning (perturbations to highly predictable targets caused more processing time than phonological perturbations) than by foils that shared acoustic and articulatory/motor plans, and that the motor speech plans are obligatorily activated when selected. Since the PH foil was a perfect rhyme, only its onset was changed in order to accurately complete the PH condition, whereas although meaning was maintained in the SE condition, the predictability of the final word form and structure changed. This supports the Dell model's supposition, that although there is interaction between levels, semantic activation precedes phonological activation, resulting in more flexibility in the PH condition. Therefore, violating a highly predictable word, even with a near semantic match, perturbs performance more than semantically anomalous close phonological foils.

The error analysis for the language experiment revealed that different types and frequencies of errors were produced across the different conditions. A total of 1788 errors were made during the experiment. Few errors were made in the TH condition (110, 6.2%). This is understandable as these control sentences included predictable endings and semantic and syntactic congruency. Proceeding from there, more errors were made in the following conditions respectively, SE (309, 17.3%), PH (561, 34.4%), and PS (808, 45.2%). Based on the ANOVA for the language errors, PH errors involved the greatest change between the two error types. Semantic errors were more common in the pseudoword condition and restoration errors more common in the phonological condition. Phonological foils were restored to the TH word more often, reflecting that participants were more likely to violate phonological precepts in the sentence than to select an unpredicted semantically related word based on the incoming acoustic information. This most likely reflects the surface level of processing as in the phonological step expected by the Dell model. The semantic concept is selected based on language prediction, and increasingly constrained word choices are created. The strength of this prediction is resistant to semantic change, but lexical properties, such onset or possibly rhyme, are more likely to be manipulable. In other words, the initial semantic/syntactic prediction is too strong to be deactivated in order to maintain phonological congruency based on the unanticipated initial phoneme of the PH foil and therefore participants make significantly more restoration errors than semantic errors. This again highlights semantics as a driving force in speech entrainment; the acoustic information is not sufficient in the PH foils to overcome the predictive nature of the sentence. Based on the ANOVA for articulation errors, it is clear that over all, participants made more errors during the PS condition. This is likely due to the lack of acoustic, semantic, or real word information. The PS foils represent the most basic mimicry of the acoustic information as it is being presented. Importantly, although there were many errors, this was not an impossible task. Across all PS trails, (100 pseudoword trials * 20 participants= 2000 total trials) participants were able to correctly adjust their speech to accurately mimic a pseudoword approximately 60% of the time, a performance that would have to be considered as reflecting above chance accuracy. Note that a word production error is not binary (correct or incorrect); it is possible to make as many errors in a given word as there are phonemes in the word. Correct production of a single word requires that all phonemes be correctly articulated and sequenced. While it is self-evident why participants make more errors during the PS condition, it is still possible to perform speech entrainment on unrelated and unfamiliar words. The tradeoff is longer response latencies and less accuracy, however, this performance is still greater than chance. Therefore, for individuals with aphasia, where semantic knowledge may be more impacted, it may still possible to perform speech entrainment, but perhaps slower and with more error-full productions.

An important finding from the Fridriksson and colleagues (2012) study showed that individuals with aphasia correctly produced more words in a script while performing speech entrainment during an audiovisual condition than an audio-only condition. The data from the current study, however, highlight the importance of semantic information during speech entrainment production in normal individuals. There may be several reasons that help explain this potential inconsistency. First, perception of motor speech movements has been shown to aid in comprehension (McGettigan, et al., 2010; Schwartz, et al., 2004; Sumby & Pollack, 1954). Originally, Sumby and Pollack (1954) documented the benefit of visual speech perception in noise for understanding. However Remez (2005) showed that the benefit visual input was not limited to conditions with speech in noise, but that listeners were aided by visual input even in better acoustic conditions. Therefore, there is some benefit of visual speech perception even in normal hearing environments. The impact of visual cues on speech perception is perhaps best demonstrated by the McGurk effect, where contradicting visual and auditory information result in a perceptual shift to a new phoneme (McGurk and

MacDonald, 1976). Reisberg and colleagues (1987) demonstrated that comprehension during audiovisual speech conditions is greater than in auditory-only conditions. The Dual Stream Model (Hickok and Poppel, 2000) posits that two streams process different types of information. The dorsal stream maps speech sounds to motor movements, and the ventral stream maps sounds to semantic concepts. The benefit observed in audiovisual speech conditions may demonstrate the synergy of these two streams processing different kinds of information simultaneously for rapid and accurate comprehension and speech production (Hickok and Poppel, 2000; 2004; 2007; Hickok, et al., 2008). The beneficial effect of providing visual information during speech entrainment may simply reflect these same findings. Second, it may be that the benefit of visual information may be stronger in individuals with aphasia and concomitant AOS than in a normal population, something that was not explicitly tested here. Normal individuals may not rely on the visual model as strongly as do individuals with aphasia, possibly because their intact, normal language network may not require extra, redundant input to operate efficiently. Although normal individuals are aided in noisy situations by seeing the speech along with the message, their performance without this visual input is still functional. Thirdly, and most likely, viewing as well as hearing the model aides performance, simply based on the fact that providing more relevant information, incorporating more modalities, is beneficial in general, whether the individual has aphasia or not. Studies have shown that visual distractors can negatively impact understanding (Tiippana, et al., 2004; Alsius, et al., 2005; Campbell, 2008), so it is important that additional information be relevant to the communication for the documented benefit. The Directions Into Velocities of Articulation model (DIVA; Guenther, Hampson, and Johnson, 1998) suggests that children learn speech movements by observing the movements in competent speakers, and through practice with trial and error, develop speech motor movements that they assign to the production of a specific sound. Therefore, a strong connection between sound and speech movement and the visual observation of speech movements are developed. Furthermore, reliance on perception of speech movements likely changes over time. Specifically, as semantic and syntactic information within the message builds creating stronger and more constrained forward prediction of upcoming words, reliance switches from motor speech perception (as observed in PS condition) to semantic and syntactic information (as observed in TH, SE, and PH conditions). In the PS condition when there is no semantic and syntactic congruency available to aid performance, it is likely that participants are relying on subtle articulation cues to help them anticipate what upcoming sounds they may encounter. In the end it is not that semantic and visual speech movement information are in competition for importance, but that one informs the other. For example, the PH foils were commonly restored to the TH foils. This demonstrates that highly predictable word choices, even incoming PH foils, which shared a majority of the words' sound patterns, were likely to be restored to predicted word choices. This was likely based on semantic congruency and against visual and acoustic differences. This changing relationship between reliance on acoustic and semantic information could be explored by calculating response latency measurements for words at the beginning of a sentence (rather than the end) and for words within sentences with lower cloze probability. The findings from the current study specifically show that speech movements and acoustic information can compete, as during the PH condition, in which case the broader semantic information and its forward prediction often influences incoming acoustic and speech motor movements of a single phoneme or word.

The findings from this experiment corroborate previous findings from Marslen-Wilson. Specifically, that sentence processing occurs at all levels of speech (Marslen-Wilson, 1975). Performance on the TH sentences, which represents typical language processing, was associated with the shortest response latencies. Perturbation of congruent sentences in any condition (phonological, lexical, semantic, syntactic) results in a decrement of speed and accuracy. However, not all perturbations are created equal. As demonstrated in this experiment, the TH sentences were entrained significantly faster than any of the other experimental conditions. This is consistent with Marslen-Wilson & Tyler (1985), who found that semantically related word lists are shadowed faster than non-semantic word lists. Further, results from the *t*-test revealed that sentences with PH foils are entrained with less processing lag, or show less decrement in entrainment, compared to SE foils. Semantic information is more heavily weighted during the processing of the incoming speech signal, and therefore also in speech entrainment performance. However, since cloze probability was not definitively shown to impact response latency, speech entrainment performance may not be influenced by cloze probability. This experiment also reaffirms

Marslen-Wilson's findings (1975) that speech entrainment is not solely supported by top-down processes. Bottom-up processing is highlighted by the observation that individuals are able to accurately perform speech entrainment under the PS condition. Although slower without it, speech entrainment is not solely supplemented by semantic/syntactic information, and can be performed (with more errors) with only acoustic information. The use of both types of processing allows for language prediction as well as rapid processing and the ability to identify likely motoric/articulatory combinations, aiding in the rapid response required during speech entrainment. Therefore, these data support Marslen-Wilson's supposition that individuals are not passively parroting during speech entrainment, but actually processing the message along several different levels (1973, 1975).

The results from the Rate Experiment support the findings and clinical observations that rate affects speech entrainment performance. The repeated measures ANOVA revealed that faster response latency times occurred for the conditions closest to normal speaking rate, 105, 150, and 195 wpm. While there was no significant difference between the response latency of these "near normal speaking rate" conditions, the slowest condition (60 wpm), and the faster condition (240 wpm) had significantly longer latencies. These effects confirm that speech rate plays an important role in language processing. When items are too slow or too fast there is a cost to the speech entrainment performance. It could be that at extremely slow rates, the prosodic factors that help provide cohesion in fluent speech are lost.

The slowest average response latency was found at the 60 wpm condition. This condition was significantly slower than the 105 and 150 wpm conditions and approached significance for the 195 wpm condition. It could be that subtle acoustic and prosodic indicators are also lost at rapid speeds. This could reflect that normal speech rate (150 wpm) is the optimal speed for processing acoustic information and implementing motor speech movements. It may be possible that human speech processing is already performing optimally for accuracy and speed during speech production. However, individuals with aphasia and concomitant AOS have slower speaking rates (McNeil, et al., 2009; Wambaugh et al., 2012; Yorkston & Beukelman, 1980). The nonfluency seen in non-fluent aphasia may be due to a loss of synchrony between the dorsal and ventral streams. This nonfluency is improved by speech entrainment possibly because it re-establishes the lost entrainment between those two streams, providing a speech motor plan to ongoing language processing. The ongoing language processing of incoming information that Marslen-Wilson describes as being analyzed at all levels (1981; phonological, lexical, semantic, and syntactic) is most likely not within normal limits for these individuals. Therefore, slowing the presentation rate of speech entrainment for individuals with aphasia could possibly benefit their performance in two ways: slower speech rate to accommodate apraxia of speech impairments and extended time for access to and retrieval from the lexicon. For example, Fridriksson and colleagues (2012) had individuals with non-fluent aphasia performing speech entrainment at approximately 50 wpm. It is possible that the reduced rate of speech for these individuals was beneficial, despite being slower than the slowest recordings for young control participants in the current study. Therefore, it may be possible that although normal individuals do not perform optimally at slower rates, the inherent difficulties for individuals with aphasia and AOS benefit from slower rates.

The other speech rate comparisons (60-240,105-150, 105-195, 105-240, 150-195, 150-240, 195-240) were not statistically significant. However, a specific pattern of performance emerged where there was a trend toward less benefit of speech entrainment performance at the slowest and very high speech rate presentations. The fastest response latencies were found at 150 and 195 speech rate presentations. Caution should be taken when interpreting these results, but it is likely that this finding reflects optimal performance at speech rates that reflect normal speech production.

There was difference in sentence accuracy for different speech rate conditions as well. The speech rate-accuracy tradeoff was explored by comparing participants' ability to correctly articulate the entire sentence at different speech rate presentations. The results show that at increasing speech rate presentations there was a slight decrement to speech entrainment accuracy. These results may not be surprising, but may highlight the importance of temporal information in speech entrainment performance. The explanation for this effect may be twofold: First, the acoustic information needs to be heard and processed for understanding. Second, the articulatory movements of repeating the message back need to be planned, sequenced, and executed. In both instances, increasing speech rate presentation increases the demands on these systems. However, just as with the response latency information, there seems to be an optimal speech rate around 105 wpm where fewer errors are made.

The response latency and sentence accuracy measures balance out the effect of fastest response latency at 195 wpm (although not statistically significant) and global sentence accuracy at 105 wpm. The trade off between fastest response latency and accuracy seem to converge at normal speaking rate (150 wpm). Response latency may be limited or slowed by language processing; such as identification of lexemes from acoustic information or semantic and syntactic incongruency. Priming studies show that lexical recognition occurs approximately 200 ms into the production of the word when semantic priming is available (Grosjean 1980; Morton & Long, 1976; Marslen-Wilson & Tyler, 1981). However, lexical recognition time increases as less priming information is available or in the case of semantic or phonological competitors (Goldinger, et al., 1989; Allopenna, et al., 1998). Thus, lexical retrieval and syntactic processing is limited by processing time. Further, speech accuracy may be limited or slowed by identification and sequencing of phonemes within the selected lexeme and the speed of articulating the given word. For example, motor speech difficulties, such as stuttering or dysarthria, are improved when individuals speak slowly. The reasons for this may include providing more time for planning and articulating the speech motor movements. Reported in the stuttering literature, articulatory planning is thought to begin as soon as lexical concepts are activated and phonological activation is initiated, with normal naming reaction times at approximately 450 ms (Indefrey and Levelt, 2004; van Lieshout, et al., 1996;

Biermann-Ruben, et al., 2004). Speech articulations are rapid, highly formulaic, and practiced motor movements, but even these movements take some time, however small, to execute. Increasing the rate of speech increases the load on these movements, resulting in more speech errors either due to an error in planning or from imprecise or misarticulated speech sounds. Due to these time limitations and the accuracy tradeoff as that time is further reduced, it is clear that normal speaking rate balances efficiency in speed and accuracy. Normal speaking rate, whether by design from practice or speed limitations in language processing or motor movement plans, is preferred for a reason.

Although the idea may seem self-evident, the rate data further support Lackner and Levine’s (1975) findings that increasing speech rate is more detrimental to non-syntactic wordlists than to sentences. Our data show that there is a temporal window of optimal performance. Although cloze probability was not confirmed to have an effect on last word response latency, there was an improving and then falling trend in performance as speech rate increased, with the fastest response latencies supported best at normal and slightly faster speech rates (150 and 195 wpm). Further, the trade off in global sentence production accuracy decreased as rate increased, with the highest accuracy at slightly slower than normal speech rate (105 wpm). These data confirm, corroborate, and highlight the need for appropriate speech rate presentation in speech entrainment and speech therapy in general. Finally, the effect of speech rate presentation on retention needs to be investigated. Although recall of shadowed information was not tested, Carey (1971) also found that increased rate not only affected shadowing performance, but also decreased retention of the message. This fact may have potential implications to clinical populations. It is likely that retention of shadowed information also aides the language processing connections, similar to the theories that support script training. While there may be benefit in speech entrainment on language processing of new scripts, the added potential therapeutic benefit of training and mass practice of the same script cannot be overlooked, as demonstrated in the improvement and generalization found in the patients from the Fridriksson et al. (2012) speech entrainment treatment study.

Looking at the time course information, there was an initial drop in response latency during the Language Experiment but not in the Rate Experiment. This difference is likely due to the change in speech rate presentation during the Rate Experiment. Since the ANOVA for the rate experiment revealed changes in the response latency, it is likely that the different rate presentation could also affect this time course analysis. However, for both experiments, performance across the entire experiment (apart from the first trials in the Language Experiment) was stable. A downward sloping trend would indicate a training effect, and an upward sloping trend a fatigue effect. Although there is a high degree of variability across individuals trials, it is important to note that each participant was experiencing different conditions at each trial. The participant's performance, although variable by condition, is stable across the experiment, and there are no drifts in performance based on fatigue or due to training. To reiterate, the highly correlated performances across conditions further indicate that participant performance is relatively stable. Marslen-Wilson reported similar findings in 1973. In his studies, response latencies were also consistent, and participants were either distant or close shadowers, but did not change groups. Although all subjects received the same amount of practice before starting the experiments, the initial drop in response latency in the Language Experiment suggests that training impacts performance. Although the difference is not statistically significant, this finding underscores the importance of task training before initiating the experimental task.

The test comparing the response latency of high and low cloze probability did not reveal a significant difference. While not a fundamental hypothesis, it would have been expected that the high cloze probability sentences would have been more strongly primed than low cloze probability sentences, resulting in faster response latencies. However, it is possible that the blocked nature of this experiment was not sensitive enough to detect a difference in this secondary hypothesis analysis. Participants repeated the same sentence a total of five times within block, and this could have impacted performance in unpredicted ways. This is not sufficient information to conclude that cloze probability has no effect on response latency. There are two limitations to the current study. First, the block design of the experiments may have induced treatment effects within sentence blocks, particularly in the Rate Experiment. There may have been unanticipated effects repeating the same sentence several times in a row, reducing the reliance on language processing that would have been present in a randomized paradigm. However, a randomized design would have introduced the opportunity for more speech errors since the sentences were unrelated. Using the highly controlled sentences instead of longer scripts, as seen in previous work (Marslen-Wilson, 1973; 1975; Lackner & Levine, 1975), allowed us to look at different attributes of the message, including cloze probability and creating carefully designed foil stimuli. Second, the inclusion of additional foil stimuli could have allowed a more carefully balanced experiment, especially in the Language Experiment. For example, the use of anomalous real word foils could have provided additional contrast in response latency performance between the semantic foil, which was semantically congruent, and the pseudoword foil, which was both semantically unpredictable and a non-English word. The addition of a real, semantically incongruent word may potentially reveal a distinction between reliance on semantic processing and acoustic analysis for speech entrainment performance. Additionally, phonemic onset versus rhyme phonological foils could allow for more fine-grained comparisons between different levels of phonological processing.

There are several potential future directions to explore from the results of these experiments. First and foremost the current experiments include a select number of foils in young, normal individuals to inform on behaviors described in clinical populations. There are many questions to be asked in regard to these effects in older adults or clinical populations, such as individuals with aphasia. Particularly, Lackner and Shattuck-Hufnagel (1982) reported the effects of speech shadowing in brain-injured individuals. They described potential distinguishing effects of semantic processing and rate in two individuals with left hemisphere injures. Unfortunately, there is little available information regarding clinical presentation and location of these injuries. Examining the effects of speech entrainment across different aphasia types and lesions locations may contribute to a better understanding the neurological mechanism that supports speech entrainment. For example, if individuals with anomic aphasia were provided enough time during speech entrainment (by reducing the rate of presentation) could speech entrainment provide support in lexical retrieval? By understanding the different factors that aid in speech entrainment performance, individual speech entrainment paradigms could be created that benefit the specific deficits of different aphasia types. Second, the current study made no attempt to further distinguish between distant and close shadowers. This was not an objective of the study, but it is possible that there is further information to be explored by distinguishing among the abilities of close and distant shadowers, as Marslen-Wilson’s work demonstrates. Third, retention of shadowed information was also not recorded as part of the current study. This, however, may have implications for using speech entrainment as a therapeutic exercise. Lastly, neuroimaging studies could further inform on the critical brain regions and connections for speech entrainment performance. Fridriksson et al. (2012) used fMRI and DTI tractography to study speech entrainment related brain activation in normal individuals as well as changes in activation patterns (fMRI) following speech entrainment treatment in individuals with non-fluent aphasia. Future work should include lesion-symptom mapping and potentially exploring brain activation within other clinical populations that may benefit from speech entrainment, such as stuttering, pure and childhood apraxia of speech, and primary progressive aphasia. These groups may continue to inform cortical areas necessary to language, repetition, fluent speech production, and possible therapeutic implications for speech entrainment in other clinical populations.

4.1 Conclusions

These studies indicate that speech entrainment is not strictly a speech production (articulation) exercise, but that linguistic factors are also important. The findings from the Language Experiment demonstrate that the linguistic qualities of the message can aid or impair performance. Importantly, speech entrainment is not impossible to perform without a meaningful message, but that message is slower and more prone to error. The Rate Experiment demonstrated that speech rate presentation also impacts speech entrainment performance. Further, the results suggest that normal speaking rate is a product of optimizing speed and accuracy during speech production. These results are meaningful, in that they inform normal performance during speech entrainment tasks. Especially in light of the work by Fridriksson and colleagues, future work should investigate these effects in individuals with aphasia, so that modifications can be made in order to provide potential therapeutic invention for these individuals.

References

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, 29(2), 130-1.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419-439.

Alsius, A., Navarra, J., Campbell, R. & Soto-Faraco, S. S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*. 15, 839–843.

Backus, O. (1945). The rehabilitation of aphasic veterans. *Journal of Speech and Hearing Disorders*, 10, 149-153.

Bailly, G. (2003). Close shadowing natural versus synthetic speech. *International Journal of Speech Technology*, 6(1), 11-19.

Basso, A. (1992). Prognostic factors in aphasia. Aphasiology, 6(4), 337-348.

Benson, D. F., & Ardila, A. A. (1996). *Aphasia: A clinical perspective*. New York, NY: Oxford University Press.

Biermann-Ruben, K., Salmelin, R., & Schnitzler, A. (2005). Right rolandic activation during speech perception in stutterers: a MEG study. *Neuroimage*,*25*(3), 793-801.

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, 42(3), 665-670.

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition*, 8(6), 631-642.

Brookshire, R. H. (2007). *Introduction to neurogenic communication disorders*. St. Louis, MO: Mosby-Elsevier, Inc.

Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision 10*:433-436.

Brust, J.C., Shafer, S.Q., Richter, R.W.,& Bruun B. (1976). Aphasia in acute stroke. *Stroke*, 7, 167-174.

Carey, P. W. (1971). Verbal retention after shadowing and after listening. *Perception & Psychophysics*, 9(1), 79-83.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1001-1010.

Chapey, R. (Ed.). (2001). *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders* (4th Ed.). Baltimore, MD: Lippincott Williams and Wilkins.

Cherney, L. R., & Halper, A. S. (2008). Novel technology for treating individuals with aphasia and concomitant cognitive deficits*. Topics in stroke rehabilitation*, 15(6), 542-554.
Cherney, L. R., Halper, A. S., Holland, A. L., & Cole, R. (2008). Computerized script training for aphasia: Preliminary results*. American Journal of Speech-Language Pathology*, 17(1), 19.Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25, 975.
Damasio, A. R. (1992). Aphasia. *New England Journal of Medicine*, 326(8), 531-9.
Davis, G.A. (2000). Aphasiology: Disorders and clinical practice. Needham Heights, MA: Allyn and Bacon. 1976(7), 167–174.
Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, 42(1), 287-314.
Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review*, 104(4), 801.
Fay, W. H., & Colleman, R. O. (1977). A human sound transducer/reproducer: Temporal capabilities of a profoundly echolalic child. *Brain and Language*, 4(3), 396-402.
Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 202.
Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse processes*, 25(2-3), 285-307.
Fridriksson, J., Bonilha, L., Baker, J. M., Moser, D., & Rorden, C. (2010). Activity in preserved left hemisphere regions predicts anomia severity in aphasia. *Cerebral Cortex*, 20(5), 1013-9.
Fridriksson, J., Hubbard, H. I., & Hudspeth, S. G. (2012). Transcranial brain stimulation to treat aphasia: A clinical perspective. *In Seminars in speech and language,* 33(3),188-202.
Fridriksson, J., Hubbard, H. I., Hudspeth, S. G., Holland, A. L., Bonilha, L., Fromm, D., & Rorden, C. (2012). Speech entrainment enables patients with Broca's aphasia to produce fluent speech. *Brain*, 135(12), 3815-3829.
Fridriksson, J., Richardson, J. D., Fillmore, P., & Cai, B. (2012). Left hemisphere plasticity and aphasia recovery. *NeuroImage*, 60(2), 854-863.
Gertsman, H. L. (1964). A case of aphasia. *Journal of Speech and Hearing Disorders, 29*, 89-91.
Gerver, D. (1974). Simultaneous listening and speaking and retention of prose. *The Quarterly Journal of Experimental Psychology*, 26(3), 337-341.
Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of memory and language*, *28*(5), 501-518.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4), 267-283.
Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological review*, *105*(4), 611.
Head, H. (1921). Discussion on aphasia. *Brain*, 43(4), 412-450.
Helm-Estabrooks, N., & Albert, M. L. (2004). *Manual of aphasia and aphasia therapy*. Austin, TX: Pro Ed.
Holland, A., Milman, L., Munoz, M., & Bays, G. (2002). Scripts in the management of aphasia. *In World Federation of Neurology, Aphasia and Cognitive Disorders Section Meeting*, Villefranche, France.
Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., ... & Grant, A. (2008). Bilateral capacity for speech sound processing in auditory comprehension: evidence from Wada procedures. *Brain and language*, 107(3), 179-184.
Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, 4(4), 131-138.
Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
Hughlings-Jackson, J. (1864). London Hospital Reports, 1, 448.
Hughlings-Jackson, J. (1879). On Affections of Speech From Disease of the Brain. *Brain*, 2(2), 203-222.
Indegreay, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components, *Cognition*, 92(1). 101-144.
Kalinowski, J., & Saltuklaroglu, T. (2003). Choral speech: the amelioration of stuttering via imitation and the mirror neuronal system. *Neuroscience & Biobehavioral Reviews*, 27(4), 339-347.
Kalinowski, J., & Saltuklaroglu, T. (2005). *Stuttering*. San Diego, CA: Plural Publishing.
Kauhanen M.L., Korpelainen J.T., Hiltunen P., Maatta R., Mononen H., Brusin E., … & Myllyla V.V. (2000). Aphasia, depression, and nonverbal cognitive impairment in ischaemic stroke. *Cerebrovasc Dis*, 10(6), 455– 461.
Kiefte, M., & Armson, J. (2008). Dissecting choral speech: properties of the accompanist critical to stuttering reduction. *Journal of communication disorders*, 41(1), 33-48.
Kleiner M., Brainard D., Pelli D. (2007). What's new in Psychtoolbox-3?. *Perception*. 36 ECVP Abstract Supplement.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161-163.
Lackner, J. R., & Levine, K. B. (1975). Speech production: Evidence for syntactically and phonologically determined units. *Perception & Psychophysics*, 17(1), 107-113.

Lackner, J. R., & Shattuck-Hufnagel, S. R. (1982). Alterations in speech shadowing ability after cerebral injury in man. *Neuropsychologia*, 20(6), 709-714.
Lambert, S. (1992). Shadowing. *Meta: Journal des traducteurs Meta:Translators' Journal*, 37(2), 263-273.
LaPointe L. L. (Ed.) (2005). *Aphasia and related neurogenic language disorders*. New York, NY: Thieme Medical Publishers.
Lichtheim, L. (1885). On aphasia. *Brain*, 7(4), 433-484.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, 19(1), 1.
Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244(5417), 522-3
Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198), 226-228.
Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, 4(1), 55-73.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. Cognition, 25(1), 71-102.
Marslen-Wilson, W. D., Tyler, L. K., & Le Page, R. B. (1981). Central processes in speech understanding [and discussion]. *Philosophical Transactions of the Royal Society of London*, 295(1077), 317-332.
MATLAB 8.0 and Statistics Toolbox 8.1, The MathWorks, Inc., Natick, Massachusetts, United States.
Max, L., Caruso, A. J., & Vandevenne, A. (1997). Decreased stuttering frequency during repeated readings: A motor learning perspective. *Journal of Fluency Disorders*, 22(1), 17-33.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, 18(1), 1-86.
McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., & Scott, S. K. (2012). Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia*, 50(5), 762-776.
McGlone, J. (1984). Speech comprehension after unilateral injection of sodium amytal. *Brain and Language*, 22(1), 150-157.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Perceptions and Psychophysics. *Nature*, 264,746-748.
McNeil, M. R., Robin, D. A., & Schmidt, R. A. (2009). Apraxia of speech: Definition, differentiation, and treatment. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders (2nd Ed.)* (pp. 249–268). New York, NY: Thieme.
Meinzer, M., Djundja, D., Barthel, G., Elbert, T., & Rockstroh, B. (2005). Long-term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy. *Stroke*, 36(7), 1462-1466.
Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. New York: Oxford.

Mishkin, M., Lewis, M. E., & Ungerleider, L. G. (1982). Equivalence of parieto-preoccipital subareas for visuospatial ability in monkeys. *Behavioral brain research*, 6(1), 41-55.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*, 6, 414-417.

Morton, J., & Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, 15(1), 43-51.

National Aphasia Association (2010). *Aphasia Frequently Asked Questions*. Retrieved November 4, 2010 from http://www.aphasia.org/content/aphasia-faq

Naeser, M. A., Martin, P. I., Nicholas, M., Baker, E. H., Seekins, H., Kobayashi, M., ... & Pascual-Leone, A. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca’s area: an open-protocol study. *Brain and language*, 93(1), 95-105.

Ogar, J., Slama, H., Dronkers, N., Amici, S., & Gorno-Tempini, M. (2005). Apraxia of Speech: An overview. *Neurocase (Psychology Press),* 11(6), 427-432.

Pedersen P.M., Jorgensen H.S., Nakayama H., Raaschou H.O., & Olsen T.S. (1995). Aphasia in acute stroke: incidence, determinants, and recovery. *Ann Neurol,* 38(4), 659–666.

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision 10*:437-442.

Peschke, C., Ziegler, W., Kappes, J., & Baumgaertner, A. (2009). Auditory–motor integration during fast repetition: The neuronal correlates of shadowing. *Neuroimage*, 47(1), 392-402.

Poeck, K., Huber, W., & Willmes, K. (1989). Outcome of intensive language treatment in aphasia. *Journal of Speech and Hearing Disorders*, 54(3), 471.

Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a New Neurobiology of Language. *The Journal of Neuroscience*, 32(41), 14125-14131.

Poeppel, D., & Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, 92(1), 1-12.

Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke*, 32(7), 1621-1626.

Reichel, U. D., & Schiel, F. (2005). Using morphology and phoneme history to improve grapheme-to-phoneme conversion. *Proc. Eurospeech*, 1937-1940.

Reisberg, D., McLean, J. & Goldfield, A. (1987). Easy to hear but hard to understand: a lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Cambell (Eds.), *Hearing by eye: the psychology of lip-reading* (pp. 97–113). Hillsdale, NJ: Lawrence Erlbaum Associates.

Remez, R. E. (2005). Three puzzles of multimodal speech perception. In W. Vatikiltis-Bateson, G. Bailly, & P. Perrier (Eds.), *Audiovisual speech* (pp. 12-19). Cambridge, MA: MIT Press.
Rosenbek, J. C., Lemme, M. L., Ahern, M. B., Harris, E. H., & Wertz, R. T. (1973). A treatment for apraxia of speech in adults. *Journal of Speech and Hearing Disorders*, 38(4), 462.
Salmelin, R., Schnitzler, A., Schmitz, F., Jäncke, L., Witte, O. W., & Freund, H. J. (1998). Functional organization of the auditory cortex is different in stutterers and fluent speakers. *Neuroreport*, 9(10), 2225-2229.
Saur, D., & Hartwigsen, G. (2012). Neurobiology of language recovery after stroke: lessons from neuroimaging studies. *Archives of physical medicine and rehabilitation*, 93(1), S15-S25.
Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M. S., ... & Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences,* 105(46), 18035-18040.
Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: Why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music perception*, 25(4), 315
Schlaug, G., Marchina, S., & Norton, A. (2009). Evidence for Plasticity in White-Matter Tracts of Patients with Chronic Broca's Aphasia Undergoing Intense Intonation-based Speech Therapy. *Annals of the New York Academy of Sciences*, 1169(1), 385-394.
Schneider, D. E. (1938). The Clinical Syndromes of Echolalia, Echopraxia, Grasping and Sucking: Their Significance in the Disorganization of the Personality*. The Journal of Nervous and Mental Disease*, 88(1), 18-35.
Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2), B69-B78.
Shattuck, S. R., & Lackner, J. R. (1975). Speech production: Contribution of syntactic structure. *Perceptual and Motor Skills*, 40(3), 931-936.
Siirtola M., Narva E.V., Siirtola T. (1977). On the occurrence and prognosis of aphasia in patients with cerebral infarction. *Scand J Soc Med Suppl*, 14, 128-133.
Sparks, R., Helm, N., & Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy*. Cortex*, 10(4), 303-316.
Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain*, 134(10), 3083-3093.
Stephen, S. C., & Haggard, M. P. (1980). Acoustic properties of masking/delayed feedback in the fluency of stutterers and controls. *Journal of Speech, Language and Hearing Research*, 23(3), 527.
Sumby, W. H., & Pollack, I. (2005). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of America*, *26*(2), 212-215.
Szaflarski, J. P., Ball, A. L., Grether, S., Al-fwaress, F., Griffith, N. M., Neils-Strunjas, J., ... & Reichhardt, R. (2008). Constraint-induced aphasia therapy stimulates language recovery in patients with chronic aphasia

after ischemic stroke. *Medical science monitor: international medical journal of experimental and clinical research*, 14(5), CR243.

Tiippana, K., Andersen, T. S. & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *Eur. J. Cogn. Psychol.,* 16(3), 457–472.

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). *Irvine phonotactic online dictionary*,2.

van Lieshout, P. H., Hulstijn, W., & Peters, H. F. (1996). From planning to articulation in speech production: What differentiates a person who stutters from a person who does not stutter? *Journal of Speech and Hearing Research*,*39*(3), 546.

Vines, B. W., Norton, A. C., & Schlaug, G. (2011). Non-invasive brain stimulation enhances the effects of melodic intonation therapy. *Frontiers in psychology*, 2.

Wada, J., & Rasmussen, T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: experimental and clinical observations. *Journal of Neurosurgery*, 17(2), 266-282.

Wade D.T., Hewer R.L., David R.M. (1986). Aphasia after stroke: natural history and associated deficits. *J Neurol Neurosurg Psychiatry*, 49, 11-16.

Wambaugh, J. L., Nessler, C., Cameron, R., & Mauszycki, S. C. (2012). Acquired apraxia of speech: The effects of repeated practice and rate/rhythm control treatments on sound production accuracy. *American Journal of Speech-Language Pathology*, *21*(2), S5-S27.

Weide, R. L. (1995). The Carnegie Mellon pronouncing dictionary: cmudict. 0.4.[On-line].

Williams, J. R. (1998). Guidelines for the use of multimedia in instruction. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 1447-1451.

Woods, B. T. (1987). Impaired speech shadowing after early lesions of either hemisphere. *Neuropsychologia*, 25(3), 519-525.

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of speech and hearing disorders*, *45*(1), 27-36.