UNIVERSITY OF CALIFORNIA
RIVERSIDE

Talker-Specific Accent: The Role of Idiolect in the Perception of Accented Speech

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Rachel Marie Miller

August 2011

Dissertation Committee:
Dr. Lawrence D. Rosenblum, Chairperson
Dr. Christine Chiarello
Dr. Curt Burgess

The Dissertation of Rachel Marie Miller is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

Like any other defining moment in life, the completion of a dissertation cannot happen without the wisdom, support, and love of many, many people. For my part, I can thank those who have been most helpful, while retaining in memory others who have been there along the way. Of course, my graduate advisor, Lawrence Rosenblum, has been my scientific guide and academic mentor all these years. It is with deepest gratitude that I thank him for his input, output, countless rewrites, and continued support. I would also like to acknowledge Christine Chiarello and Curt Burgess, for their advice and help during the process. I would not have had inspired ideas without my colleagues: Kauyumari Sanchez and James Dias. The statistics would still be unclear, if not for the support of Russell Pierce, David Vazquez, Theresa Cook, Justin Estep, and Michael Erickson. The experiments would never have been completed without the many undergraduate research assistants who have come and gone over the years. The paperwork would not have been filed without the folks in the office: Faye Harmer, Dianne Fewkes, Conrad Colindres, and Renee Young. I may not have even started this journey without my undergraduate mentors: David Nalbone and Arlene Russell. And, I may not have made it through without the grant and fellowship support of CHASS and the Graduate Division, respectively. I must also thank Sabrina Sidaras, Jessica Duke Alexander, Lynne Nygaard, and Midam Kim for their input and instructions.

Mostly, I would like to acknowledge my wonderful family and friends who have been with me through the good (e.g., significant results, inspiration) and not so good times (e.g., statistical debacles, writer's block). Thank you to my friends here and far away. Thank you to my father, Fred, for being a financial foundation. Thank you to my sister, Mindy, who often was on the other end of the phone. Thank you to Theresa and David for picking up the slack and helping me to rethink things that I couldn't quite get right. Thank you to my mother-in-law, Kay, for being a constant help when time was running out and for the great editorial reviews. Thank you to Margaret and Jim for kind words and constancy. Thank you to my beautiful and brilliant step-daughters, Angelica and Erica, who have endured limited fun-time until this was finished. Thank you especially, to my wonderful husband, Randall, who is the most patient, kind, and self-sacrificing person I know.

Dedication

To
my husband,

Randall Cook,

for a future filled with adventures.

And to,

Nancy, Jill, and Brian,

who could not finish this journey with me,
but are forever in my heart.

“I am in a constant process of thinking about things...”
-- Richard Brautigan (n.d.)

ABSTRACT OF THE DISSERTATION

Talker-Specific Accent: The Role of Idiolect in the Perception
of Accented Speech
by

Rachel Marie Miller

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, August 2011
Dr. Lawrence D. Rosenblum, Chairperson

Strong evidence suggests that familiarity with talker-specific (idiolectic) information benefits speech perception (e.g., Nygaard & Pisoni, 1998). However, it seems that talker familiarity does not influence the perception of *accented* speech (Sidaras et al., 2009). This dissertation assesses whether idiolect and accent are both encoded (Goldinger, 1998); or instead, if the perception of accented speech involves a process of normalization (e.g., Halle, 1985). Two sets of experiments examined talker-specific and accent-general influences on the perception of accented speech using a speech alignment methodology. Speech alignment is the tendency of individuals to subtly imitate the speech of a person with whom they are speaking and occurs to native (unaccented) speech (Goldinger, 1998). During Experiment Series 1, native English subjects shadowed a Chinese- or Spanish-accented model producing English words. In Experiment 1a, raters judged whether a model's tokens sounded more similar to the shadowed token or to a different subject's token shadowed after a same-accented model. Results revealed significant talker alignment. In Experiment 1b, raters judged whether shadowed tokens were more similar in *accent* to models

with the same or a different accent, neither of whom was shadowed. Accent alignment results were inconclusive due to a response bias, which seemed to be related to the magnitude of a model's accent as measured in Experiment 1c. Generally, the finding of talker alignment suggests that talker-specific information is encoded during accented speech perception. Experiment Series 2 investigated potential causes for a lack of these findings in the experiments of Sidaras et al. (2009). In Experiment 2a, listeners were trained to shadow or transcribe Spanish-accented models and were tested on either the same or different models. No effects of training were found. In Experiment 2b, listeners were instead trained and tested on native, English talkers. There was a significant effect of training, but no influence of familiar talker and no difference in accuracy between shadowers and transcribers. These overall findings suggest that talker-specific information is encoded during the perception of accented speech, supporting an episodic account of speech perception. However, the nature and interaction of talker-specific and accent-general information remains unresolved.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

A thorough understanding of speech perception requires examination of factors that generate variability in the speech signal and their potential influences on a listener's processing and production of speech. Listeners are faced with speech that commonly varies due to the state of a talker (e.g., emotion), due to characteristics specific to a particular talker (e.g., idiolectic information), and sometimes due to speech signals that contain more systematic variation, such as that caused by a foreign accent. As a result, perceivers should have an incredibly difficult time when it comes to their ability to process and understand speech. Yet, normal listeners are typically able to resolve the linguistic content of a given message.

At one time, it was suggested that the ability to perceive speech was driven by a process of normalization which removed distortions from the speech signal in order to allow easy retrieval of linguistic content (Shankweiler, Strange, & Verbrugge, 1977). This process involved the "stripping away" of the surface characteristics of speech, leaving very basic, abstract linguistic units to be analyzed (e.g., Halle, 1985; Joos, 1948; Neary, 1989; Summerfield & Haggard, 1975). Yet, this view did not seem able to account for the fact that familiarity with idiolectic (or talker-specific)

information was found to influence and often facilitate perception and recognition of the linguistic message (e.g., Goldinger, 1996; Palmeri, Goldinger, & Pisoni, 1993; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994). For example, Goldinger (1996) and Palmeri et al. (1993) showed that same-voice repetitions improved speed and accuracy of recognizing words (see also Church & Schacter, 1994; Schacter & Church, 1992).

A number of episodic (exemplar-based) accounts exist that attempt to explain the effects of talker familiarity on speech perception (Goldinger, 1996, 1998; Hintzman, 1986; Johnson, 1997; Pierrehumbert, 2002). These models suggest that speech perception involves the storage of detailed traces of a speech event, including information about the talker's voice. These memory traces are activated by incoming stimuli based on similarity to the stored trace, thereby aiding the processing of speech (Goldinger, 1998; Jacoby & Brooks, 1984). In this way, episodic models are able to explain how idiolectic information can improve perceptual recognition of words, enhance memory, and influence speech productions (e.g., Goldinger, 1996, 1998; Goldinger & Azuma, 2004; Miller, Sanchez, & Rosenblum, 2010; Namy, Nygaard, & Sauterteig, 2002; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993; Pardo, 2006; Sanchez, Miller, & Rosenblum, 2010; Shockley, Sabadini, & Fowler, 2004).

Other sources of variability, such as foreign accent, provide initial difficulties for the perceiver. For instance, the intelligibility of non-native speech is equivalent to native speech being reduced by several decibels (Lane, 1963; van Winjgaarden,

2001). Additionally, when first encountering an unfamiliar, non-native accent, listeners have problems identifying words and recognizing when mispronunciations have occurred (e.g., Lane, 1963; Schmid & Yeni-Komshian, 1999). Comprehending accented speech is also more difficult for older listeners, especially those with lower hearing acuity, than younger listeners (Adank & Janse, 2010).

Nonetheless, recent research has uncovered patterns of familiarity-based improvement for accented speech that are similar to those of talker familiarity (e.g., Bradlow & Bent, 2008; Clarke, 2000; Clarke & Garrett, 2004; Flege, Bohn, & Jang, 1997; Sidaras, Alexander, & Nygaard, 2009). Such results might also be explicable by episodic models. However, very recent evidence exists showing that talker-specific learning is potentially overridden by accent-general learning (Sidaras et al., 2009; Bradlow & Bent, 2008).

These results are problematic since a particular nonnative talker will present to the listener, not only information about their native language in the form of an accent, but also their specific idiolectic information. Intuitively, these two sources of information should both influence speech perception and processing. In other words, a listener should be better able to recognize speech that is accented when they are familiar with this type of accent; *and* they should improve even more when this accented speech is produced by a familiar talker.

1.1 Dissertation Overview

This dissertation investigates how foreign accent bears on the influences of talker-specific information by using a speech alignment methodology. The goal of the present dissertation is to answer the following questions: *Does the perception of accented speech involve a process of normalization or is talker-specific information encoded? If talker-specific information is stored during the perception of accented speech, is it somehow 'masked' by accent-general information? Will using a more immediate and productive encoding task reveal the influence of talker-specific information in the perception of accented speech?*

This dissertation is organized in the following way: Chapter 1 presents a brief introduction of the questions addressed and an outline of the dissertation. Chapter 2 examines speech literature on talker-specific and accent-general learning. Chapter 3 considers applicable theoretical literature in the area of speech perception. Chapter 4 discusses speech alignment methodology and relevant literature. Chapter 5 introduces the current study. Chapter 6 covers Experiment Series 1. Chapter 7 details Experiment Series 2. Chapter 8 provides a general discussion of the experimental results and addresses theoretical and practical implications.

Chapter 2

Speech Perception Literature

2.1 Talker-Specific Learning

Individual talkers produce lexical items that both vary from production to production and vary from the same lexical item produced by another talker. These between-talker differences in the acoustic (or visible) composition of speech utterances can be driven by numerous idiolectic characteristics of the talker (e.g., age, positioning of articulators; Abercrombie, 1967; Ladefoged, 1980). Yet, listeners seem to have little difficulty in comprehending these varying speech signals.

Idiolectic (or talker-specific) information is unique to the individual and helps listeners identify a talker by voice alone. This information includes acoustic factors such as spectral structure, formant frequencies, and pitch (Doddington, 1985), as well as articulatory rate, intonation, and vocal intensity (Giles, Coupland, & Coupland, 1991; Natale, 1975). Talker-specific information can also be conveyed through visible articulatory style (e.g., Lachs & Pisoni, 2004; Rosenblum, Niehus, & Smith, 2007; Rosenblum et al., 2002).

Early models of speech perception considered talker characteristics to be a *problem* that needed solving by the listeners (Gerstman, 1968). For example,

Mullennix, Pisoni, & Martin (1989) showed that multiple-talker lists are more detrimental to spoken word recognition than single-talker lists. The authors suggest that these results are due to high processing demands on the perceptual system, which is required to restructure for each new voice being heard.

Alternatively, the processing costs incurred by presenting multiple talkers could be due to memory interference that occurs when memories containing talker-specific information are matched to a lexicon (Martin, Mullennix, Pisoni, & Summers, 1989; Mullennix et al., 1989). Extending these findings, Goldinger, Pisoni, and Logan (1991) showed that the effects of talker variability were dependent on the rate of stimuli presentation. When provided a slow presentation rate, words in early positions of multiple-talker lists were recalled more accurately than words in single-talker lists suggesting that talker-specific information is an integral part of speech processing.

However, other evidence reveals that familiarity with talker-specific information is beneficial to perception and recognition of the linguistic content of speech (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993). This *talker-specific learning* has been shown to improve recognition of repeated words (e.g., Craik & Kirsner, 1974; Goldinger, 1996; Palmeri et al., 1993), enhance perceptual identification of speech (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994), and influence speech productions in the form of speech alignment (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Namy et al., 2002; Pardo, 2006; Shockley et al., 2004; Sanchez et al., 2010). For example,

using a novel voice learning paradigm, Nygaard & Pisoni (1998) found that listeners' experience with a particular talker can later aid the retrieval of linguistic content from that same talker even when words were not the same from training to test. In fact, these talker-specific learning effects can even occur cross-modally (e.g., Rosenblum, Miller, & Sanchez, 2007). Talker-specific learning is said to be evidence for the encoding of highly-detailed traces of speech events, which will be discussed in greater theoretical detail later (see 2.1).

On the whole, a great deal of evidence reveals the important role that talker-specific information can play in the perception of native (unaccented) speech (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993). However, the speech signal can also contain variation that is more systematic in nature. This systematic variation can be due to deficits of a listener (e.g., cochlear implant patients; Chang & Fu, 2006) or due to differences in pronunciation of a particular *group* of individuals: i.e., accent (e.g., Clarke & Garrett, 2004; Bradlow & Bent, 2008; Sidaras et al., 2009). Variation in the speech signal due to accent is addressed in the following section.

2.2 Accent-General Learning

Accented speech occurs when the structure of a talker's native language (e.g., phonetic system, phonological rules) interacts with the non-native language they are attempting to produce (e.g., Tarone, 1987; Flege et al., 1997; Sidaras et al., 2009). For example, when producing English, a native Spanish talker will often add epenthetic

schwas before fricative plus stop structures (e.g., [ə]store vs. store); produce full vowels which would be reduced in English (e.g., reas[o]ns vs. reas[ə]ns); and have shorter voice onset times for syllable initial voiceless stops (e.g., [p]at vs. [p^h]at)(Magen, 1998).

Accent variation is considered systematic because general features of a native language (e.g., Spanish) will induce similar deviations on the productions of a given non-native language (e.g., English). In fact, non-native accents are often defined in terms of how they deviate from native speech production norms. For example, Cunningham-Andersson and Engstrand (1989) showed that there was a significant correlation between the number of features that deviated from standard Swedish productions and the perceived strength of an accent. Phonetically trained talkers were recorded producing foreign deviations (e.g., unaspirated initial voiceless plosives; vowel replacements) typical of Finnish or British natives while reading a piece of prose. Native Swedish listeners were asked to indicate whether or not they heard (1) a foreign accent, (2) a regional accent, or (3) if the reading was 'merely strange'. They were also asked to grade the readings on level of deviation. Results indicated that deviation of certain features or combinations of features led to the perception of a foreign accent for 50% of the listeners. In addition, subjective accentedness and number of deviations were correlated, where certain combinations of deviations give a greater impression of foreign accent.

As expected, the deviations from native speech that make up an accent differ with native language background. For instance, acoustical analyses show that native

Brazilian-Portuguese talkers produce English with shorter voice-onset-times (VOTs) for plosives and differences in the realization of consonant clusters (Major, 1987); while ratings of accentedness for native Spanish talkers are influenced most by suprasegmental factors (e.g., syllable structure, phrasal stress; Magen, 1998).

Perceptual ratings of accentedness can also be influenced by factors other than acoustic or articulatory information (e.g., Flege & Fletcher, 1992; Levi, Winters, & Pisoni, 2007). For instance, Levi et al. (2007) showed that high frequency words were rated as less accented than low frequency words. Moreover, these ratings differences were due to listener effects, rather than production differences for these types of words. Acoustical measurements showed that differences in production of words at varying levels of lexical frequency were not completely predictive of accentedness rating. Listener differences also affect intelligibility ratings of accented speech (e.g., native background; Bent & Bradlow, 2003) and accented speech comprehension (e.g., age; Adank & Janse, 2010).

Regardless of their form, the deviations from native speech that make up an accent can cause a listener to encounter numerous difficulties. Research on accented speech has shown decreases in word identification accuracy (Lane, 1963); difficulty in identifying mispronunciations produced by an accented talker (Schmid & Yeni-Komshian, 1999); reduction in intelligibility equivalent to reducing native speech by several decibels (Lane, 1963; van Winjgaarden, 2001), and poorer performance in voice identification when compared to unaccented speech (Irwin & Thomas, 2006).

Still, similar to the influence of talker familiarity, the ability to perceive accented speech improves in situations where a listener is familiar with a foreign accent. Thus, *experience* perceiving accented speech can improve later perception of accented speech (Adank, Hagoort, & Bekkering, 2010; Bradlow & Bent, 2008; Clarke, 2000; Clarke & Garrett, 2004; Flege et al., 1997; Sidaras et al., 2009).

For example, Clarke and Garrett (2004) tested the amount of experience one needs with a particular accent in order to improve recognition of speech with that same accent. They presented listeners with English sentences spoken by Spanish- or Chinese-accented talkers. The accented sentences ended in a probe word not predictable from the sentence content. Listeners judged if a word presented orthographically was the same or different from the final probe word in each sentence. Results revealed that after one minute of exposure to accented speech, there was improvement in the processing efficiency (reaction time) for responding to both the Spanish- and Chinese-accented speech. These results suggest that the speech processing system is flexible enough to adapt to deviations from native speech quickly. However, this research is limited as it only tested the accented speech of one talker per accent.

In a more recent study using multiple talkers, Bradlow and Bent (2008) tested whether training listeners with English sentences spoken by native Chinese talkers would improve transcription accuracy for Chinese-accented speech. Results revealed that listeners in multiple-talker and talker-specific (trained and tested on same model) conditions showed significant improvement in transcription accuracy at test when

compared to listeners in single-talker and control conditions (English-talker, no training). However, when compared to each other, the multiple-talker and talker-specific groups did not differ significantly at test. The authors propose that this generalized adaptation to accented speech could be due to the range of stimuli available in these conditions. In this case, small amounts of accent information from multiple talkers or large amounts of accent information from a single talker are similarly beneficial to later accent perception. In addition, Sidaras et al. (2009) found that studying the speech of Spanish-accented talkers improved listeners' ability to identify novel Spanish-accented speech at test.

The abovementioned studies show that familiarity with the systematic variation of an accent can improve later perception of novel speech with that same accent (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009), much the way that familiarity with talker-specific information improves later perception of novel speech from the same talker (e.g., Nygaard & Pisoni, 1998; Nygaard et al., 1994). Clearly, these two sources of variation do not occur independently. A given non-native talker will provide the listener not only general information about their native language background, but also their unique idiolectic information.

It seems that both talker-specific and accent-general information should influence speech perception and processing, yet there is evidence that this is not the case, as will be discussed in the next section.

2.3 Talker-Specific Accent

Sidaras et al. (2009) were the first authors to examine how the role of talker familiarity would influence the perception of Spanish-accented speech. Many of the details and interpretations of this study are the impetus for the questions being addressed in the current dissertation, so the study will be discussed in some detail.

To examine the influences of talker-specificity on the perception of accented speech, Sidaras et al. (2009) used a high variability training and test paradigm; i.e., they presented multiple, accented talkers at *both* training and test. This allowed them to test how listeners adjust to an accent (e.g., Spanish-accented English), as well as to particular talkers.

Native English listeners were trained to transcribe sentences and words spoken in English by a series of six Spanish-accented talkers. The listeners were then tested on their ability to transcribe novel, Spanish-accented sentences (or words). However, half of the subjects heard the novel sentences (or words) produced by the same talkers heard during training, while the other half heard a set of new Spanish-accented talkers. Sidaras and colleagues (2009) reasoned that if listeners were learning both the lawful variation of accented speech *as well as* talker-specific information, then those individuals who were trained and tested with the same talkers would show the greatest amount of learning. This finding would be consistent with the results of talker familiarity effects in non-accented speech recognition (Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993). However, if listeners were simply incorporating accent-general information and not using

talker-specific characteristics, being trained with accented speech should improve performance similarly, regardless of whether the same or different talkers were heard from training to test.

The results of Sidaras et al. (2009) indicated that subjects improved in transcription accuracy of accented speech comparably, regardless of whether they heard the same or different talkers heard during training. This suggests that listeners may have been becoming familiar with the lawful variation of accented speech without regard to the talker-specific information.

The results seem somewhat at odds with the findings of previous studies for which the use of talker-specific (or idiolectic) information aided later perception of (unaccented) speech (Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993). However, there is some evidence that suggests that an unfamiliar accent makes it more difficult to identify talkers. Kerstholt, Jansen, Van Amelsvoort, & Broeders (2006) showed that an unfamiliar accent reduces the likelihood of correctly identifying an individual as a perpetrator in a voice line-up. In addition, listeners were not as accurate at identifying voices in a line-up that were produced with a Spanish-accent than those produced in unaccented English (Thompson, 1987). Taken together, these findings suggest that accent information may somehow override talker characteristics, which could result in the lack of talker effects reported by Sidaras et al. (2009). The theoretical implications regarding the individual and/or combined influences of talker-specific and accent-general information are addressed in more detail in the next chapter.

Chapter 3

Theoretical Considerations

Several theoretical accounts are relevant to the finding that the influences of talker-specific information seem to be overridden during the perception of accented speech (Sidaras et al., 2009).

The following sections present different theories of talker-specific influences on speech perception and suggest how they might account for the influence—or lack of influence—of talker-specific information during the perception of accented speech. Further theoretical considerations will also be discussed.

3.1 Speech Normalization

In trying to explain the way the human perceptual system deals with the "problem" of variability in the speech signal, early speech theorists suggested that a process of *normalization* occurred (Shankweiler et al., 1977). The normalization process is thought to passively filter phonetically-irrelevant information (e.g., voice quality, speech rate, emotional tone) from speech, while retaining information about acoustic patterns that reveal linguistic content (e.g., Halle, 1985; Joos, 1948).

Through normalization, the end result of speech processing is matching of a prototypical and symbolic representation that possesses none of the surface characteristics of the original signal to ideal templates (e.g., Halle, 1985; Joos, 1948; Neary, 1989; Summerfield & Haggard, 1975). The accounts of normalization in speech are similar to those proposed in computational vision where pattern matching occurs to an ideal template, regardless of the size or positioning of a visual input pattern (e.g., Roberts, 1965).

Normalization is thought to allow listeners to understand speech content, regardless of who produced the speech. Potentially, this process could also allow listeners to understand the linguistic content of accented speech.

Recall however, that evidence points to the actual encoding of talker-specific information during the perception of native speech (e.g., Goldinger, 1996; 1998; Nygaard & Pisoni, 1998; Palmeri et al., 1993). For example, Goldinger (1996) found that subjects are faster and more accurate at identifying words repeated in the same vs. a different voice. Still, it is possible that talker normalization occurs when listeners perceive accented speech; a possibility supported by the results of Sidaras et al. (2009).

Inherent in normalization accounts is the proposition that talker and other sources of variability are processed separately from linguistic content. In fact, this separate processing may increase the load on the cognitive system and slow speech perception when perceiving multiple talkers (e.g., Creelman, 1957; Mullennix et al., 1989; Martin et al., 1989), though others argue that sources of variability are

processed simultaneously (see Nusbaum & Magnuson, 1997). Accent is another source of variability that may increase cognitive load and, in the process, make the processing of talker variability more tenuous.

Still, Sidaras et al. (2009) showed that accent training improves the later perception of speech with the same accent, suggesting that accent information is not being removed from the signal as full normalization would suggest (e.g., Halle, 1985; Joos, 1948; Neary, 1989; Summerfield & Haggard, 1975). In order to address this enhancement in the processing of accented speech, a proponent of normalization might contend that the accented-training effects of Sidaras et al. are due to adaptation of normalization processes. In other words, the normalization processes get faster and more efficient as listeners become accustomed to a source of variability (e.g., accent; Kolers, 1979; Kolers & Roediger, 1984).

In fact, one explanation presented by Sidaras et al. (2009) is that when presented accented speech, listeners may engage in routines which unravel variation due to accent, talker, and other sources. From this viewpoint, the reason listeners show improvement after exposure to accented speech may be due to the tuning of procedural memory or normalization operations, which increases processing efficiency (e.g., Kolers, 1979; Kolers & Roediger, 1984).

Nevertheless, normalization accounts cannot clearly explain how the perception of speech content is improved by familiarity with a given talker (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993) or accent (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al, 2009).

This evidence is more supportive of the episodic encoding of the surface characteristics of speech. Episodic accounts will be addressed in the next section.

3.2 Episodic Accounts

In response to the limitations of normalization, episodic accounts (aka exemplar-based models) of speech perception were developed to propose the encoding of speech events as highly-detailed traces in memory (e.g., Goldinger, 1996, 1998; Hintzman, 1986; Johnson, 1997; Pierrehumbert, 2002). In these models, the surface characteristics of a speech event, including talker-specific characteristics, are involved in the activation of stored traces, with more activated traces influencing subsequent perception (e.g., Goldinger, 1998; Johnson, 2008). For example, Goldinger's (1998) episodic encoding theory proposes that traces of heard speech events are present and accessible in lexical memory.

Goldinger (1998) explains this episodic encoding of speech information in the context of MINERVA 2 (Hintzman, 1986). From this account, every speech event a listener encounters forms a trace in memory, which includes the surface characteristics of that event (e.g., linguistic content, talker-specific information). When a new word is presented to a listener, an *analogue probe* is communicated to all stored traces in memory. These traces are activated based on similarity to the probe. A collection of all activated traces constitutes an *echo* that is sent to working memory (WM) from long term memory (LTM). This echo can contain information that is not in the probe (e.g., conceptual knowledge), which associates the echo to past

experience. The summed activation of all traces (i.e., *echo intensity*) increases with greater similarity of the probe to existing traces and the greater number of these traces. The echo intensity is associated with recognition memory (i.e., stronger echoes produce faster reaction times).

As support for an episodic account of speech perception, Palmeri et al. (1993) used a continuous recognition memory task (CRMT) to investigate the effects of talker variability and voice (same vs. different) on the recognition of spoken words. Palmeri and colleagues found that same-voice repetitions were recognized more quickly and with higher accuracy than different voice repetitions over all levels of talker variability. In addition, Goldinger (1996) found that both recognition and perceptual identification of spoken words improves when speech is produced in the same vs. a different voice from training to test (see also Schacter & Church, 1992; Church & Schacter, 1994). If normalization were occurring, talker-specific information would be unavailable to help in the retrieval of linguistic content (e.g., Halle, 1985; Joos, 1948; Neary, 1989; Summerfield & Haggard, 1975).

Other evidence for the encoding of detailed talker information comes in the form of *speech alignment*. Speech alignment is the tendency of individuals to subtly imitate the speech of a person with whom they are speaking (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Namy et al., 2002; Nye & Fowler, 2003; Pardo, 2006; Shockley et al., 2004; Sanchez et al., 2010). Goldinger (1998) has explained that speech alignment occurs when the model's talker-specific information influences the subject's speech productions due to activation of episodic traces.

Consequently, this phenomenon suggests that talker-specific information is not normalized because this information influences subsequent speech productions. Speech alignment and its theoretical implications will be addressed in greater detail in Chapter 4.

Finally, talker familiarity has also been shown to improve perception of novel speech (Nygaard & Pisoni, 1998; Nygaard et al., 1994). Using a novel voice learning paradigm, Nygaard & Pisoni (1998) found that listeners' experience with a particular talker can later aid the retrieval of linguistic content even when words were not the same from training to test. In this case, the encoding of episodic information may be occurring at a sublexical (e.g., phonemic, subphonemic) level (see also Nielsen, 2011). Regardless, these results suggest that both the talker-specific and linguistic information for speech events are preserved in long term memory (LTM).

From the perspective of episodic accounts, the perception of accented speech should involve storage of both talker-specific and accent-general information. If this is the case, then listeners should be better able to recognize accented speech when they are familiar with that particular accent (e.g., Spanish-accented English), *and* they should improve even more when this accented speech is produced by a familiar talker. However, this did not occur in the experiments conducted by Sidaras et al. (2009) and begs the question of what happened to the talker-specific information.

In considering normalization and episodic accounts, several explanations are possible for the lack of talker familiarity results reported by Sidaras et al. (2009). First, although normalization may not occur during the perception of native (unaccented) speech (e.g., Goldinger, 1996; 1998; Nygaard & Pisoni, 1998; Palmeri et al., 1993), the large phonetic differences or added cognitive load when perceiving accented speech could require a process of normalization that removes talker-specific information (e.g., Halle, 1985; Joos, 1948; Nearey, 1989; Summerfield & Haggard, 1975). If this were true, then accented speech perception should not reveal the presence of talker familiarity effects, regardless of the methodology applied. Alternatively, the increased processing costs associated with perceiving accented speech may reduce encoding of talker-specific information during certain types of tasks (e.g., transcription). Finally, as addressed by Sidaras et al. (2009), talker-specific information could be encoded as usual, but the presence of a larger amount of accent-general information (i.e., due to presentation of multiple, accented talkers) may *mask* the influences of talker-specific information.

One way to examine if talker-specific information is encoded during the perception of accented speech is to test the question using a different methodology than that used by Sidaras et al. (2009). As stated, Sidaras et al. used a transcription task to assess the influence of talker-specific information of accented speech. Although transcription tasks have revealed evidence of talker familiarity effects for native speech (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993), these tasks may not provide access to talker-specific

information during the perception of accented speech due to, for example, a delay between stimulus presentation and response. This could be the result of a reduction in the encoding of talker-specific information during the perception of accented speech. For this reason, a task methodology that does not involve a delay between the presentation of accented stimuli and a response to this speech could reveal the immediate effects of talker influences. The aforementioned speech alignment methodology is just such a task and will be discussed in greater detail in the next chapter.

3.3 Other Theoretical Considerations

Two additional theories account for talker-specific influences on speech perception and production by discussing the link between these functions, as well as the objects of speech perception (e.g., Fowler, 1986, 2003, 2004; Fowler, Brown, Sabadini, & Weihing, 2003; Liberman, 1983; Liberman & Mattingly, 1985; Sancier & Fowler, 1997; Shockley et al. 2004). These *gestural theories* suggest that the objects of speech perception are articulatory gestures of the vocal tract (e.g., Liberman & Mattingly, 1985; Fowler, 1986; 2003), rather than acoustic or auditory events. However, *motor theory* of speech perception (e.g., Liberman & Mattingly, 1985) and the *direct-realist approach* (e.g., Fowler, 1986; 2003; Gibson, 1979) differ in their explanations of how the perceptual system works with these gestures.

For the motor theory of speech perception, human listeners recover representations of articulatory events (i.e., intended gestures), which are processed by

a specialized module in the brain (Liberman & Mattingly, 1985). The direct-realist approach (e.g., Gibson, 1979) suggests that human listeners (perceivers) do not need representations or a specialized module because gestures lawfully form the actual information in the speech signal (e.g., Fowler, 1986; 2003). According to Fowler (2003), the perceptual system (via the sense organs) is stimulated by structure in media that allows the direct perception of distal objects and events. For example, patterns of light reflected from a chair stimulate the eyes and provide visual information about the chair.

Also common between the motor and direct realist theories is the conception that the speech perception and production functions are linked, potentially due to these functions sharing a *common currency* (Fowler, 2004; Fowler, Brown, Sabadini, & Weihing, 2003; Liberman, 1983; Liberman & Mattingly, 1985; Sancier & Fowler, 1997; Shockley et al. 2004). If the basic units of speech perception and production are the same (i.e., articulatory gestures), then the perception of gestures containing talker-specific information might influence the production of speech that integrates some of that information (Sanchez et al., 2010). Perception is thought to prime productions that are more similar to the perceived talker. In fact, this proposal is commonly addressed in the literature on speech alignment (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Namy et al., 2002; Nye & Fowler, 2003; Pardo, 2006; Shockley et al., 2004; Sanchez, et al, 2010) in which talker-influences on spontaneous production is observed (see the next section).

This concept could also be applied in the context of accented speech, where the consequence of perceiving accented speech might be productions of speech that sounds more accented. In fact, findings of gestural drift– a shifting of articulatory gestures in the direction of an ambient language community – provide initial evidence for this inference (see Sancier & Fowler, 1997).

Thus, if the informational details extracted in speech perception prime production, then talker-specific and accent-general information might both leave their mark on speech productions. For the purposes of the present dissertation, this direct link between perception and production can be helpful in establishing whether or not talker-specific and accent-general information are available during accented speech perception.

Chapter 4

Speech Alignment

Even as newborn babies, human beings have the remarkable capability to imitate facial expressions and novel acts (see Meltzoff & Moore, 1997, for review). Yet, the tendency to imitate is not just limited to infants. Adults also have been found to unconsciously imitate the behaviors and postures of a conversational partner in a social context (e.g., Chartrand & Bargh, 1999; Shockley, Santana, & Fowler, 2003). Originally considered an intentional act mediated by factors such as social desirability (Natale, 1975), Chartrand and Bargh (1999) suggest that this imitation is often passive and can occur without volition. They proposed the *chameleon effect*, a nonconscious tendency toward mimicking facial expressions, body posture and mannerisms of another person.

Yet an individuals' imitative propensity is not just restricted to mimicry of body position and expressions. During dialogue, alignment occurs at numerous communicative levels of speech (e.g., semantic, syntactic, lexical, phonological)(see Pickering & Garrod, 2004, for a review). As mentioned previously, this *speech alignment* is the tendency of individuals to subtly imitate the speech of a person with whom they are talking (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al.,

2010; Namy et al., 2002; Nye & Fowler, 2003; Pardo, 2006; Shockley et al., 2004; Sanchez et al., 2010). In the course of conversational interaction, talkers have been found to partially match each other in intonational contour, speech rate, and vocal intensity (Giles, Coupland, & Coupland, 1991; Natale, 1975). But even in isolation, individuals will align to the speech of a recorded model (e.g., Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Nye & Fowler, 2003; Sanchez et al., 2010; Shockley et al., 2004).

Speech alignment has been demonstrated both to auditory speech (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Namy et al., 2002; Nye & Fowler, 2003; Pardo, 2006; Shockley et al., 2004) and to visual (lipread) speech (Gentilucci & Bernardis, 2007; Miller et al., 2010; Sanchez et al., 2010). Although considered an unconscious tendency (i.e., one that does not necessitate an explicit decision making process), this phenomenon can be influenced by talker-independent factors (e.g., word frequency; Goldinger, 1998) and socio-cognitive biases (Babel, 2009), as well as by phonetic repertoire and language knowledge (e.g., Babel, 2009; Nielsen, 2011; Nye & Fowler, 2003).

The speech alignment phenomenon is thought to uncover the influences of talker-specific information on subsequent speech productions (e.g., Goldinger, 1998). Evidence regarding a close connection between speech perception and production processes is important for at least two reasons. First, it is consistent with general neurophysiological research on mirror neurons (mirror systems), which are active both during specific motor behavior and during the perception of those motor

behaviors (e.g., Fadiga, Fogassi, Povesi, & Rizzolatti, 1995; Oztop, Kawato, & Arbib, 2006). It is also supported by recent brain imaging and lesion studies suggesting that the brain areas associated with imitation of prosodic and segmental phonetic properties are direct neighbors to a brain area that has been implicated in processing of auditory spatial information and vocal sounds (i.e., posteromedial superior temporal plane; Kappes, Baumgaertner, Peschke, & Ziegler, 2009; Warren, Wise, & Warren, 2005).

Unlike the transcription tests used by Sidaras et al. (2009), alignment is thought to reveal a direct and immediate influence of perception on the production of speech. In this way, a speech alignment methodology may reveal the encoding of talker-specific information during the perception of accented speech. The following sections introduce evidence for different types of speech alignment, the shadowing paradigm, and the rationale for using an alignment methodology.

4.1 Talker Alignment

Alignment to talker-specific information has been shown in numerous experimental contexts (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Miller, Sanchez, & Rosenblum, submitted; Namy et al., 2002; Nielsen, 2011; Pardo, 2006; Shockley et al., 2004). Goldinger (1998) presented some of the first empirical evidence for talker alignment using a shadowing paradigm.

In the basic shadowing paradigm, subjects are first asked to read aloud a series of words (baseline). They are then asked to say these same words as quickly as

possible (shadow) after they hear each word said by a model. Alignment is assessed through an AXB-matching task where naïve raters are asked to judge which of a subject's two words (the baseline or shadowed word) is more similar to (or a better imitation of) the model's word. The subject's words are presented in the A and B position, while the model's word is in the X position. Typically, alignment is said to occur when raters judge the shadowed token as more similar to the model's token at greater than chance levels.[1]

Investigating an episodic model of speech perception, Goldinger (1998) found that subjects align to talkers when asked to shadow isolated words, and that the strength of this alignment is a function of talker-independent factors (e.g., word frequency). Goldinger also had talkers shadow isolated words produced by recorded models either immediately or after a 3 - 4 s delay. Subjects were first asked to read text (baseline) words, then to complete a listening task where they heard models produce 0, 2, 6, or 12 repetitions of these words, and finally to shadow these words. The baseline and shadowed words of each subject were then compared in an AXB perceptual matching task. The results indicated that: a) that subjects in the immediate shadowing condition were judged as showing greater alignment than subjects in the delayed shadowing condition; b) that pooled over these two conditions, low

[1] The use of AXB perceptual measures of alignment are common in the speech alignment literature (e.g., Goldinger, 1998) and are often used in lieu of acoustical analysis for several reasons. For example, it is often difficult to identify the exact phonetically-relevant dimensions of the speech signal that are changing during alignment. Although acoustical analyses can reveal changes in individual phonetic dimensions of speech, it is not clear if alignment is a result of a single change or a change in numerous dimensions. Also, alignment is thought to serve a sociolinguistic function of improving communicative efficiency (Giles et al., 1991; Natale, 1975). Thus, it seems appropriate to take advantage of the perceptually-relevant measures of alignment afforded in AXB-matching tasks (see Miller et al., submitted, for a review).

frequency words invoked a higher degree of alignment than high frequency words, and c) that perceived alignment increased with the number of repetitions of each word the subjects heard during the listening task. According to Goldinger, this evidence suggests that episodic traces of words we hear are present and accessible in lexical memory and alignment emerges as a byproduct of responding. In response to the immediate vs. delayed shadowing results, Goldinger suggests that the delay allows long-term traces to flood working memory. This then reduces the influence of talker-specific information on speech alignment.

In order to uncover the phonetic dimensions of speech alignment, Shockley et al. (2004) had subjects shadow auditory tokens with digitally extended voice-onset-times. Consistent with the results of Goldinger (1998), Shockley et al. showed that subjects align to digitally extended voice-onset-times of a model's auditory tokens. In a similar vein, presentation of audiovisual speech with lengthened auditory VOTs and varying visible syllable rate can invoke changes in a talker's VOT durations (Sanchez et al., 2010). Finally, Nielsen (2011) illustrated that alignment to extended VOTs generalizes, not only to novel words with the same initial stop consonant (e.g., word initial /p/), but to words with a stop consonant that has the same place of articulation (e.g., word initial /k/). These finding suggests that shadowers' articulatory gestures shift in the direction of a model's articulatory gestures, at least for this particular phonetically-relevant dimension of speech.

Alignment to a given talker is not restricted to the shadowing of isolated words and syllables. Using an interactive map task (Anderson et al., 1991), Pardo

(2006) showed that a live interaction could elicit alignment between interlocutors that persists even after the conversation has ended (see also, Kim, Horton, & Bradlow, 2011).

The aforementioned research shows that talker alignment occurs for native (unaccented) speech (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Miller et al., submitted; Namy et al., 2002; Nielsen, 2011; Pardo, 2006; Shockley et al., 2004), but does talker alignment occur in the context of accented speech?

One way to examine if talker-specific information is encoded during the perception of accented speech is to evaluate speech productions for the presence of *talker* alignment. Assessing speech productions shadowed after accented models could help determine whether or not talker-specific information is still encoded (i.e., not normalized) during the perception of accented speech. To do so, the present dissertation enlists a shadowing paradigm.

The shadowing paradigm was chosen above interactional tasks for several reasons: (1) Use of the shadowing task allows presentation of stimuli and models to be held constant for later comparison purposes. Interaction tasks typically place pairs of naïve subjects together and record their conversational interaction (e.g., Kim et al., 2011; Pardo, 2006). (2) A shadowing task can be implemented into the high variability training and test paradigm (e.g., Sidaras et al., 2009), which allows for comparisons between accent training using transcription vs. shadowing tasks. The immediacy of speech shadowing may make the encoding of talker-specific

information more efficient, and thus demonstrate an influence of talker familiarity on the perception of accented speech. (3) The shadowing task allows for the manipulation of talker-independent variables (e.g., word frequency), which have been shown to have an influence on both alignment (Goldinger, 1998) and the assessment of accentedness (Levi et al., 2007). Additional rationale for the use of the alignment methodology is discussed in more detail in a subsequent chapter (see 4.3).

4.2 Accent Alignment

Empirical evidence for alignment to accented speech is limited, and support for *accent* alignment comes most often in the form of anecdote. For instance, when a person returns from a trip abroad (e.g., France, Canada), they are sometimes described as having 'picked up' the speaking style of the ambient language community.

Still, there is some empirical evidence for a phenomenon akin to accent alignment. For example, *gestural drift,* a shifting of articulatory gestures in the direction of an ambient language, is thought to occur due to unconscious imitation (Sancier & Fowler, 1997). Sancier and Fowler (1997) looked at whether short periods of exposure to either a native or a non-native language community would have a perceptible effect on the speech productions of bilingual talker of Brazilian-Portuguese. Recordings of the talker after a 4.5 month stay in the U.S., a 2 month stay in Brazil, and finally after another 4 month U.S. stay were presented to native Brazilian-Portuguese and native English listeners. Results indicated that Brazilian-

Portuguese listeners noticed an accent after the talker's recent stay in the U.S.; while native American-English listeners did not notice a comparable change in accent after the talker's stay in Brazil. There was also a significant shift of VOT in the direction of most recent language exposure based on acoustical analyses. There was a noticeable shortening of VOTs for unaspirated stops after a stay in Brazil; and a noticeable lengthening of VOTs after a recent stay in the United States**.** These findings suggest that even a short exposure to a given language community can influence a talker's speech productions to perceptibly shift in the direction of that language.

Beyond gestural drift, research also suggests that many years exposure to an ambient language drives long-term changes in speech productions (e.g., Flege, 1987; Major, 1992). Delvaux and Soquet (2007) show that ambient speech characteristics of a dialect influence a listener's speech (see also, Kraljic, Brennan, & Samuel, 2008). Yet, none of these studies directly investigate whether or not subjects will *align to* or subtly imitate a given accent by shifting speech productions towards accent-general information.

In order to test conversational alignment between accented talkers, Kim et al., (2011) had talkers with varying language and dialect backgrounds perform an interactive task. Talkers were asked to perform a picture description task with someone who had the same native language and dialect, the same native language and a different dialect, or a different native language. Kim et al. revealed that *talker* (interlocutor) alignment during an interactive task was more likely for talkers who

share a native language and dialect, than for those who have different dialects or native language backgrounds. Kim and colleagues suggest that one reason for the lack of phonetic convergence between interlocutors of different dialects could be due to a need to maintain intelligibility during the conversation, which might require slower speech rates and more pauses. Alignment between interlocutors with close language distances could be facilitated by shared phonetic categories. Conversely, a lack of shared phonetic repertoire might limit the alignment between interlocutors without similar language backgrounds (see Babel, 2009, for review).

Although the initial results of Kim et al. (2011) do not reveal alignment between interlocutors with differing language backgrounds, this could be due to extra-linguistic factors (e.g., social biases). As well, the authors did not attempt to differentiate the influences of talker-specific and accent-general information in their assessments of alignment, as these influences were confounded (i.e., ratings were not made which could separate talker and accent).

The concept that alignment can occur to a particular accent is extremely pertinent to the current dissertation. Alignment to accent could occur either with or without alignment to talker. The former would suggest the encoding and availability of both sources of information, while the latter would be similar to the findings of Sidaras et al. (2009) suggesting that accent may override idiolect, perhaps during early levels of processing.

4.3 Rationale for an Alignment Methodology

For the present dissertation, the speech alignment methodology was chosen to investigate the encoding of talker specific-information during the perception of accented speech for several reasons.

First, the alignment methodology reveals an immediate influence of perception on the production of speech. Sidaras et al. (2009) use a task where listeners are asked to identify the accented stimuli by transcribing it on a keyboard. Whereas transcription provides evidence for influences on implicit memory (Goldinger, 1996), the effects of speech alignment have been interpreted as evidence for perceptual regulation of speech productions based on input from a given talker (e.g., perception-production link; Fowler, 2004; Pardo, 2006). A link between perception and production has been found for action events (e.g., Bäckman, Nilsson, & Chalom, 1986; Cohen, 1983). For example, subjects are better at remembering action events when they are asked to "act out" (or enact) these events than when they are given a written list of the events (Cohen, 1983). In other words, a subject would better remember the action of someone "opening the door" when they were told to pretend to open the door, than when they simply read about the task. In fact, Bäckman et al. (1986) suggest that qualities of subject-performed events such as motor features (e.g., open, point) and characteristics of the objects (e.g., texture, shape) may automatically be encoded in memory. These findings and those regarding the link between speech perception and production (Fowler, 2004; Pardo, 2006) could suggest

that perceptual regulation via the shadowing task may be more effective at exposing talker effects when perceiving accented speech than the transcription task.

Second, the minimal delay incurred during speech shadowing may provide facilitation of talker effects by allowing talker-specific information to be encoded more efficiently. Recall that Goldinger (1998) found greater alignment when subjects shadowed words immediately than when they shadowed after a 3 - 4 s delay. Goldinger offers that this is due to interference between the word held in working memory and other traces of the same or similar items in long term memory that reduces the efficacy of the details of the original word stimulus. In other words, numerous traces that are similar to the talker's word flood working memory and reduce talker effects. This reduction in the dependability of talker-specific information to aid speech perception after a delay could also occur in the context of accented speech. In fact, a delay of several seconds has also been shown to reduce recall of a list of items that are well within the working memory span (i.e., the number of items that can reliably be recalled) when intervening stimuli prevented rehearsal (Brown, 1958).

Regarding Sidaras et al. (2009), the delay between presentation and transcription might allow the more prevalent *accent-general* information to flood working memory, thus reducing the influences of talker information. An alignment task might show an influence of talker-specific speech because listeners reproduce the words immediately upon their presentation (by producing them).

Finally, speech alignment, like perceived accentedness, is moderated by word frequency. For example, Goldinger (1998) found greater speech alignment to low frequency (LF) words than to high frequency (HF) words. This influence of word frequency is said to occur because HF words are likely to have more memory traces and thus produce a more 'generic' echo. LF words have fewer traces in memory allowing the echo to contain more of the surface characteristics (e.g., talker-specific information) of the original stimuli.

Recent evidence shows that ratings of accentedness are also influenced by word frequency. Levi et al. (2007) showed that naïve listeners rated HF words as significantly less accented than LF words. Both sets of findings provide evidence for episodic models. Based on reports of Goldinger (1998), the more times a word is encountered (i.e., HF words), the more traces are stored in memory of that particular word. LF words will have fewer traces stored in memory. Possibly then, for speech alignment, this means that shadowers will be less likely to align to HF words and more likely to align to LF words. For ratings of accentedness, there are more exemplars of HF words, so they sound less accented. Whereas, less exemplars exist of LF words, so they sound more accented.

Investigating alignment to accented talkers at varying levels of frequency may reveal results that are quite similar or surprisingly different from those found in the alignment literature (Goldinger, 1998). For example, if alignment to talker-specific information during the perception of accented speech is influenced by accentedness, then there may be an interaction between accentedness and word frequency (i.e., LF

words may influence less alignment than HF words because they are perceived as more accented). Using a speech alignment methodology with a word frequency manipulation allows for comparisons within alignment and accent literature.

Chapter 5

Current Study

The individual and combined influence of talker-specific and accent-general information on the listener remains unresolved in the literature. The speech alignment methodology has been successful in showing evidence for the encoding of talker-specific information (e.g., Goldinger, 1998). The goal of this dissertation is to present experiments using an alignment methodology that address how accent bears on the influences of talker-specific information and to discover whether episodic theories need to be modified to account for the effects of accent.

The main questions of the dissertation are: *Does the perception of accented speech involve a process of talker normalization or is talker-specific information encoded? If talker-specific information is stored during the perception of accented speech, is it somehow 'masked' by accent-general information? Will using a more immediate and productive encoding task reveal the influence of talker-specific information in the perception of accented speech?*

Chapter 6

Experiment Series 1

6.1 Introduction

The first series of experiments investigates whether alignment will occur to talker-specific information in the context of accented speech. The presence of *talker* alignment to accented models would indicate the encoding of talker-specific information even in the presence of accented speech. This could mean that the task used by Sidaras et al. (2009) may not have been sensitive enough to expose the influences of talker-specific information on the perception of accented speech or that this task did not induce as much encoding of talker information as the alignment method. Regardless, finding evidence for talker alignment to accented speech would be supportive of the episodic accounts of speech perception (e.g., Goldinger, 1998).

In lieu of talker alignment, subjects may show alignment to accent, but not to talker. This finding could indicate that partial normalization of talker information is occurring or that accent information and talker information are processed in different ways. Finally, there is the possibility that alignment will occur to both talker and accent. This latter result would indicate encoding of both sources of information (or an overlap of these types of information) and again support an episodic interpretation of accented speech perception.

6.2 Experiment 1a

Subjects were asked to shadow the speech of either native English, Spanish- or Chinese-accented talkers. The degree of alignment was evaluated using a modified perceptual matching task (Miller et al., submitted). In this task, naïve raters were asked to judge whether an utterance shadowed after a given accented model sounds more similar to that model than does an utterance shadowed after another model with the *same accent*. Talker alignment would be indicated if raters judge the model's utterances as more similar to the subject who shadowed that model than to a subject who shadowed a different model with the same accent at greater than chance levels.

Keeping in mind previous research regarding the influences of word frequency on alignment (e.g., Goldinger, 1998), word frequency should also have an influence on alignment to accented talkers. Recall that Goldinger (1998) found that LF words induced greater alignment than HF words, potentially as a result of differences in the number of traces for these types of words. The present study of accented speech perception and shadowing may reveal similar word frequency results. Generally, listeners may align more to LF words from accented talkers than to HF words because LF words have less traces in memory and are more influential on speech productions (e.g., Goldinger, 1998).

6.2.1 Method

6.2.1.1 Participants

Three types of participants took part in the present study: Models, subjects, and raters. All participants had self-reported corrected-to-normal hearing and vision. All shadowers and raters were native American-English (native English) talkers.

Models. Four non-native and two native English talkers acted as models in the experiment and produced the original word list to be shadowed. All models were female. Two non-natives talkers were from a Mexican-Spanish language background. The remaining two non-native talkers were from a Mandarin-Chinese language background. All non-native talkers were recruited through the university extension center and local community college ESL programs. A description of non-native models' English language exposure is presented in Table 1. The native English talkers were born and raised in the United States and were recruited from the University of California, Riverside. Models were financially compensated for their participation in the study.

Subjects. Twenty-four female undergraduates aged 18 to 22 acted as subjects who were asked to shadow the models' words. Female subjects were used in order to improve the chances of finding alignment to accented speech. Prior research suggests that female subjects often align more than male subjects, potentially due to increased perceptual sensitivity (Namy et al., 2002; but see Pardo, 2006). These subjects were native English talkers with no speech impediments. The subjects were recruited from

the University of California, Riverside and participated in order to partially fulfill a course requirement.

Raters. Twenty-four undergraduates (21 female, 3 male) aged 18 to 22 acted as raters in an AXB matching task. These raters were native English speakers recruited from the University of California, Riverside. They participated in order to partially fulfill a course requirement.

Table 1. Language Background

Language background information for nonnative models

	Language Background				
Model	Native Language	Age	Time in U.S.	Age of English Exposure	Time Speaking English
Sp1	Spanish	21	11	10	11
Sp2	Spanish	41	3	38	3
Ch1	Chinese	24	0	10	14
Ch2	Chinese	27	3	12	15

Note. All values are in years

6.2.1.2 Materials and apparatus

A list composed of 120 English words was derived from Goldinger (1998) and used as stimuli (see Appendix A). The list consisted of an even representation of bisyllable and monosyllable words from four frequency classes (Kučera & Francis, 1967): High frequency (HF; >300 occurrences per million), medium-high frequency

(MHF; 150-200 occurrences per million), medium-low frequency (MLF; 50-100 occurrences per million), and low frequency (LF; <5 occurrences per million). A SONY DSR-11 camcorder was used to videotape the models.

All stimuli were presented to participants using PsyScope software. Text (baseline) words and listening block matrices were presented on a 20-in. video monitor positioned 3 ft in front of the participants. Auditory stimuli were presented through SONY MDR-V6 headphones. The models and shadowers responded verbally into a Shure Beta 58a microphone and were audio-recorded at a 44000 Hz, 16 bit rate using Amadeus software. Listening block responses were made using a Targus numeric keypad with the numbers 1-12 labeled on the face in a 3 x 4 matrix. Transcription responses and AXB ratings were collected using a standard keyboard.

6.2.1.3 Procedure

The experiment took place in three phases. For all three phases, individuals sat in a sound-attenuating booth.

Phase 1. In Phase 1, the six female models were filmed producing the 120 word list. Prior to filming, all models were giving the word list and asked to produce each word in order to familiarize themselves with the words and to reduce the chances for major production errors not associated with accent. During filming, the word list was presented to the models as text on a video monitor. The words were presented randomly at an interval of 1 word per second. Models were asked to speak the words “quickly, but clearly” into the microphone. These utterances were filmed using the camcorder and these recordings were edited on a computer to produce

tokens for later presentation to the subjects. The audiovisual recordings were digitized and the audio-only portion was edited using Amadeus software to create 120 audio tokens. All tokens were adjusted to an average RMS amplitude of -45.00 dB prior to presentation to shadowers and raters.

Phase 2. Phase 2 consisted of four tasks: (1) baseline (reading text) task, (2) listening task, (3) shadowing task, and (4) transcription task.

For the baseline (text) task, the 24 subjects (all female) were audio recorded producing the original word list, which they read from a video monitor with the words presented individually at one second intervals. Although these utterances were collected, they were not used in the present study.

After baseline recordings, the 24 subjects were then randomly assigned to one of the four experimental (e.g., Chinese- or Spanish-accented model) or two control (e.g., native English model) conditions, with four subjects assigned to each of the six models. Subjects were required to complete 10 blocks that each contained three tasks always in the following order: Listening, Shadowing, Transcription. All words were presented to the subjects over headphones.

For each trial of the Listening Task, subjects were asked to listen to each word and then to indicate the location of this word on a 3 x 4 matrix seen on the monitor as illustrated in Figure 1. Each text word in the matrix (e.g., Flannel, Social) was associated with a numbered location (e.g., “9”, “2”). Subjects were asked to respond by pressing the appropriate number on a keypad device (e.g., Goldinger & Azuma, 2004). For example, a subject might hear the model say the word “Flannel”, and they

would respond by pressing “9” on the keypad device. Each word was presented two times during the listening task. Two versions of the word matrix were produced for each listening task, so each word was in a different position on subsequent presentation (see Figure 1). Subjects were presented 12 words per listening task x 2 repetitions x 10 blocks for a total of 240 word trials during the listening tasks. All words and matrices were presented randomly using computer software.

A similar listening task has been used in previous alignment studies as a way to potentially increase alignment by presenting multiple repetitions of given words (e.g., Goldinger, 1998; Shockley et al., 2004) and to present words without requiring a spoken response (e.g., Goldinger & Azuma, 2004). In the former studies, it was shown that increasing the number of word repetitions a shadower hears increases the level of alignment. In the present study, this task was selected as a way to potentially increase encoding of surface characteristics (e.g., idiolect, accent) by presenting two repetitions of each word. It was also used as a method of allowing subjects to associate the heard version of the accented word with the correct English word. From Goldinger’s (1998) view, talker-specific information and lexical content are stored together in memory, which he suggests is evidenced by the influences of word frequency on alignment. Thus, it is important to assure that subjects are perceiving the accented speech as English words.

For each trial of the Shadowing Task, subjects were asked to say each word they heard “quickly, but clearly” into the microphone (e.g., Shockley et al., 2004; Miller et al., 2010). Subjects were never asked to imitate or repeat the model. Words

were presented randomly using computer software. Subjects were presented 12 words per shadowing task x 10 blocks for a total of 120 word trials during the shadowing tasks. All shadowed utterances were recorded and later edited to create the shadowed tokens for comparison purposes in Phase 3.

For each trial of the Transcription Task, subjects were asked to type (transcribe) the words they heard the models say. Words were presented randomly using computer software. Subjects were presented 12 words per transcription task x 10 blocks for a total of 120 word trials during the transcription tasks. Transcription responses were analyzed for accuracy (i.e., correct or incorrect responses).
The transcription task was implemented because **i**naccurate transcription responses may have indicated that subjects perceived the accented words as something other than the intended word, thus limiting further alignment comparisons. If this were the case, then speech alignment was not truly being tested in these trials. Due to the nature of the episodic encoding being tested, it was essential that subjects perceived the speech they heard as English words (e.g., Goldinger, 1998). Words that were transcribed incorrectly were removed from later comparisons in Phase 3, though transcription accuracy was generally high.

Figure 1

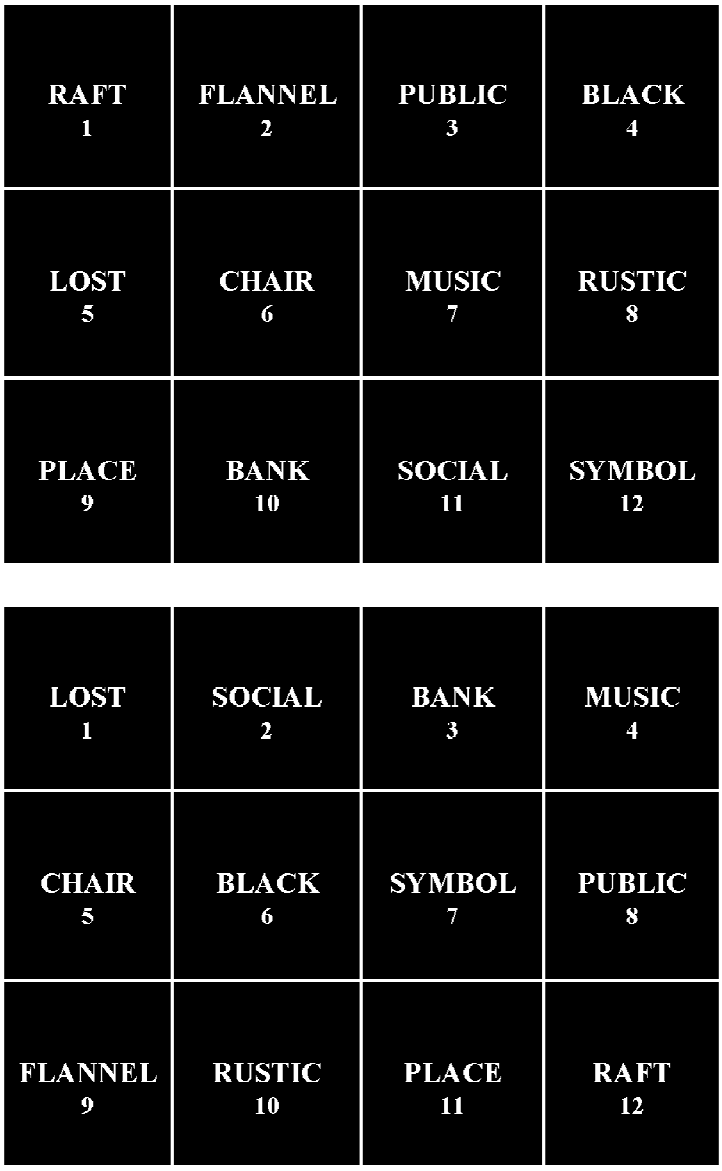

Figure 1: Sample of two versions of matrices presented to subjects during listening blocks. Subjects were asked to identify which word they heard by pressing the corresponding number for the word on a keypad device.

The words presented were different between blocks, but remained consistent within each block across tasks (12 words x 10 blocks = 120 words). Words were assigned to each block so that they equally represented syllable (e.g., monosyllable, bisyllable) and frequency class (e.g., HF, MHF, MLF, LF). Thus, any given participant completed a total of 480 word trials: 240 in the listening tasks, 120 in the shadowing tasks, and 120 in the transcription tasks. For reasons of transcription accuracy and pronunciation issues, a total of 116 words (29 per frequency class) were available for use in Phase 3 of the experiment.

Phase 3. In order for raters to make perceptual judgments of alignment to *talker*, rather than to *accent*, a modified AXB matching task was used (e.g., Miller et al., submitted). Twenty-four naïve raters (21 female, 3 male) were asked to judge the relative similarity between a models' words and the subject who had shadowed that model (shadowing subject) as compared to a second subject who had shadowed a different model with the same accent (comparison subject) (i.e., shadowing subject shadowed Chinese-accented Model 1; comparison subject shadowed Chinese-accented Model 2). Words were presented to raters in the form of triads, where the models' utterances were always in the middle, X position. The shadowing subjects' utterances appeared either in the A (first) or B (third) position and the comparison subjects' utterances appeared in the remaining A or B position. Position was counterbalanced across an experimental session. Alignment to talker was said to occur if the model's words were judged as more similar to the words shadowed after that model vs. the words shadowed after a different model with the same accent.

In this version of the AXB task, raters were assigned to rate the 116-word list produced by shadowers of the Chinese-accented, the Spanish-accented, or the English models. A rater was only assigned to make judgments for one accent group. A given rater would hear a total of six voices: two models (e.g., Model 1, Model 2), two subjects who shadowed Model 1, and two subjects who shadowed Model 2. However, because of the large number of trials necessary to properly counterbalance across shadowers, the word list was split into two lists (List 1, List 2). Each list contained 58 words (e.g., List 1 contained Typhoon, List 2 contained Flannel). Each subject who shadowed a given model was represented by one half of the word list. Each script represented half the words as presented by the shadowers, hence each shadower's words were judged by a total of four raters (two for List 1 words, two for List 2 words). This procedure was used to maintain the number of triads a rater had to judge at 464 (58 words x 4 subjects x 2 A-B positions).

Raters listened to the sets through SONY MDV-600 headphones at a comfortable listening level and were asked to choose which of the words, the first or third, sounded more similar to the second. Raters were instructed to press the key labeled "1" on the keyboard, if the first word sounded more similar to the second; or to press the key labeled "3" on the keyboard if the third word sounded more similar to the second.

6.2.2 Results and Discussion

The purpose of Experiment 1a was to assess subjects' alignment to talker-specific information in the context of accented speech. If subjects are aligning to a given talker, then raters should judge a model's utterances as more similar to the subject who shadowed that model than to a subject who shadowed a different model with the same accent.

Mean talker alignment was calculated for each shadowing subject measured as the number of model utterances chosen by raters as sounding more similar in pronunciation to those of the shadowing subject. The mean percentage of shadowing subjects' shadowed tokens chosen as being pronounced more like the models' tokens was 54.6%. This percentage was compared to chance (50%) using a t-test, which revealed that the shadowing subjects' shadowed tokens were judged as pronounced more like the models' tokens than were the comparison subjects' tokens [$t(23) = 2.181$, $p = .040$, Cohen's $d = .909$]. Although these effects are subtle, they are comparable to other alignment results (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Namy et al., 2002; Pardo, 2006; Shockley et al., 2004).

A repeated measures ANOVA was conducted to compare the separate between-subjects factors of model (Sp1, Sp2, Ch1, Ch2, En1, En2) and accent (Chinese, English, Spanish) and the within-subjects factor of word frequency (HF, MHF, MLF, LF). There was no significant effect of model [$F(5, 18) = 1.71$, $p = .184$, $\eta_p^2 = .322$] or accent [$F(2, 21) = .699$, $p = .508$, $\eta_p^2 = .973$]. These findings suggest

that overall talker alignment is occurring and is not driven by a specific model or accent.

There was a significant effect of word frequency [$F(3, 54) = 3.88$, $p = .014$, $\eta_p^2 = .177$].[2] Pairwise comparisons show significant differences in alignment judgments based on word frequency that do not follow the pattern of previous results regarding word frequency and alignment to native (unaccented) speech (e.g., Goldinger, 1998). For example, HF and MHF words were judged as more similar to the model significantly more often than MLF. There was no significant difference between LF word judgment and other levels of frequency. Although Goldinger reported significantly greater alignment for words at lower frequency levels, the present results suggest that alignment was greater for high frequency (HF) words as depicted in Figure 2. This effect could be occurring as a byproduct of perceptual differences between accented high and low frequency words (Levi et al., 2007). Recall that Levi et al. (2007) found that raters judged HF words as significantly less accented than LF words even though there were not consistently matching acoustical differences between these types of words. Perhaps, subjects in the present experiment perceived HF words as less accented making talker-specific information more available in this context.

[2] Because the within-subjects factor of word frequency has more than two levels, it is necessary to test for violations of sphericity. Throughout the present dissertation, Mauchly's test of sphericity was performed on these factors in order to assure that they did not violate the assumption of sphericity. Where the assumption of sphericity is violated in tests with significant effects, the results of Mauchly's test and the appropriate degrees of freedom corrections will be reported.

A significant word frequency x model interaction [$F(15, 54) = 1.895$, $p = .045$, $\eta_p^2 = .345$] was also observed. Further investigation into this interaction revealed that there was a significant frequency effect for model En2 [$F(3, 9) = 4.590$, $p = .033$, $\eta_p^2 = .605$]. Pairwise comparisons revealed that ratings of talker alignment for HF (57.6%) and LF (47.3%) words differed significantly for model En2 at the $p = .025$ level. No other model x frequency effects or interactions were significant.

Figure 2

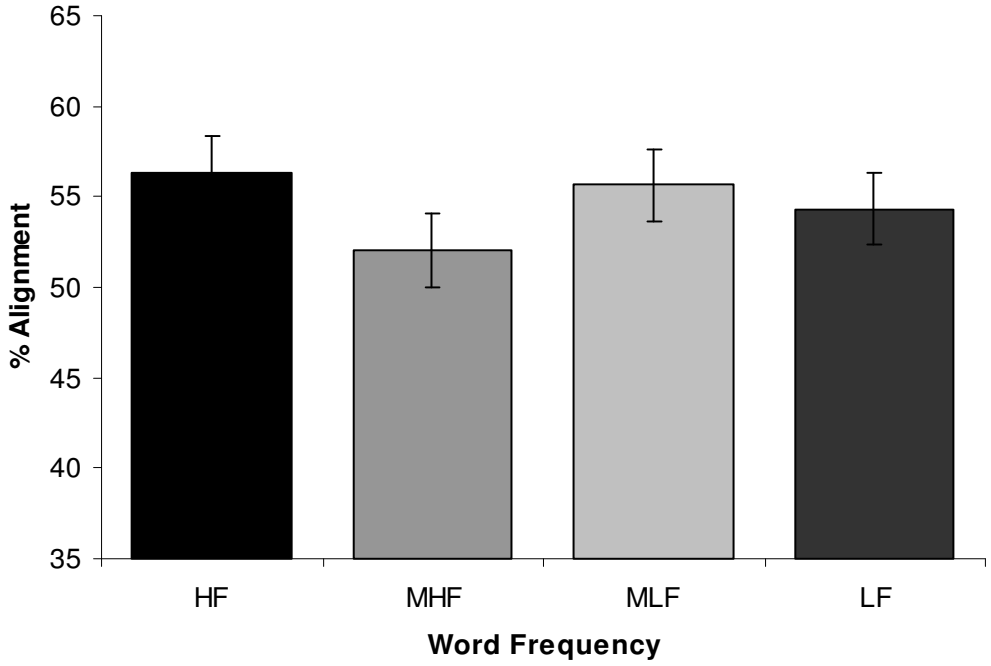

Figure 2. Graph of frequency effect found during talker alignment. HF and MHF words show significantly higher alignment than MHF. LF words were not significantly different.

The results demonstrate significant talker alignment in the context of accented speech. These findings could suggest that talker-specific information is encoded and

is available to influence speech productions during the perception of accented speech. As mentioned, the subtlety of these effects must be considered in line with other research on alignment, which presents similar means using AXB perceptual rating tasks (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Namy et al., 2002; Pardo, 2006; Shockley et al., 2004).

Contrary to the results of Sidaras et al. (2009), these results reveal a significant effect of talker-specific information in the context of accented speech. This finding suggests that talker-specific information is not removed from the speech signal due to a process of normalization (e.g., Halle, 1985; Joos, 1948; Neary, 1989; Summerfield & Haggard, 1975). In this sense, the results are supportive of episodic accounts, where talker information is stored in episodes and retrieved to influence later perception and production of speech with similar characteristics.

Although the present study shows talker-specific information can influence speech productions, it does not indicate whether accent-general information has this same type of influence. The role of accent-general information in speech alignment is addressed in Experiment 1b.

6.3 Experiment 1b

Experiment 1b investigated the encoding of accent-general information by testing speech alignment to accented models. Utterances shadowed after Spanish- and Chinese-accented models were used in a second modified perceptual matching task. In Experiment 1b, naïve raters were asked to judge whether an utterance shadowed

after a model with a given accent (e.g., Spanish) sounded more similar in accent to a different model with the *same* accent (e.g., Spanish) or a different model with a *different* accent (e.g., Chinese). Neither of the actual shadowed models was heard by the raters. Accent alignment would be indicated if raters judged the utterances shadowed after a model with a given accent as more similar in accent to the model with the same accent at greater than chance levels.

Evidence for alignment to a specific *accent* has not been established using a shadowing paradigm. Recall, however, that research has shown that talkers shift their articulations in the direction of an ambient language community (e.g., Flege, 1987; Major, 1992, Sancier & Fowler, 1997) or dialect (e.g., Delvaux & Soquet, 2007; Kraljic et al., 2008). Indeed, much like alignment, this *gestural drift* may be due to unconscious imitation (Sancier & Fowler, 1997). Additional evidence suggests that alignment will occur to systematic variation in speech that could be considered similar to accented speech (e.g., Nye & Fowler, 2003).

The present experiment will also investigate the influence of word frequency on alignment to accent. In Experiment 1a, frequency effects were in a different direction from previous alignment studies (Goldinger, 1998), perhaps, due to the perceived accentedness of the words (e.g., Levi et al., 2007). It may be the case that alignment to accented *talkers* occurs more for higher frequencies words because these are perceived as less accented (see 6.2.2). However, when looking at alignment to accent, subjects may align more to LF words than HF words because these words

deviate more from standard English (e.g., Nye & Fowler, 2003) and subjects have less exemplars for LF accented pronunciations (e.g., Goldinger, 1998). If this is the case, LF words should induce greater *accent* alignment than HF words, a pattern consistent with the previous results of Goldinger (1998).

6.3.1 Method

6.3.1.1 Participants

The models and shadowers were the same as those used in Experiment 1a. 16 new undergraduates (4 male, 12 female) aged 18 to 25 acted as raters in an XAB matching task. All raters were native speakers of American English with normal hearing. Participants were recruited from the University of California, Riverside and received credit in order to partially fulfill a course requirement. None had participated in Experiment 1a.

6.3.1.2 Materials and apparati

All materials and apparati were the same as those used in Experiment 1a. In this experiment, the audio-recorded utterances from Phase 1 and Phase 2 of Experiment 1a were used as comparison stimuli (see Procedure 5.1.1.3 for more information).

6.3.1.3 Procedure

The naïve raters judged whether a subject's token shadowed after an accented model was more similar in accent to a model with the same accent (e.g., Spanish) or a model with a different accent (e.g., Chinese). Alignment to accent was said to occur if

the utterances shadowed after a model with one accent were judged as more similar in accent to another model with the same accent (who was never heard) vs. a model with a different accent at greater than chance levels.

In order for raters to make perceptual judgments of alignment to *accent*, rather than to *talker*, an XAB matching task was used. The XAB format was chosen over an AXB format due to reported concerns over the difficulty of presenting a non-accented word surrounded by accented words. This format was tested in a pilot experiment using native English shadowers and these ratings were not significantly different from the original AXB judgments on which the test was based [$t(14) = .058$, $p = .954$, Cohen's $d = .03$]. This XAB test allowed for perceptual judgments of alignment to *accent* by having raters compare a subjects' utterances shadowed after a model with a given accent to another model with the same accent vs. a model with a different accent (i.e., Subject 1's utterances shadowed after Spanish-accented Model 1 [X] were compared to Spanish-accented Model 2 and Chinese-accented Model 3).

In this XAB task, raters were assigned to rate the 116-word list produced by shadowers of the Chinese-accented and of the Spanish-accented model. All raters were assigned to make judgments across accent groups. A given rater would hear a total of six voices: two models (e.g., Spanish-accented Model 2, Chinese-accented Model 3), two subjects who shadowed the other Spanish-accented model (Sp Shadowers), and two subjects who shadowed the other Chinese-accented model (Ch Shadowers). Because of the large number of trials necessary to properly counterbalance across shadowers, the word list was again split into two lists (List 1,

List 2). Each list contained 58 words (e.g., List 1 contained Typhoon, List 2 contained Flannel). Each subject who shadowed a given model represented one-half of the word lists (e.g., Sp Shadower 1, List 1; Sp Shadower 2, List 2; Ch Shadower 1, List 1; Ch Shadower 2, List 2). Each of the eight scripts represented half the words as presented by the shadowers. Each shadower's words were judged by four raters (two for List 1 words, two for List 2 words). This procedure was used to maintain the number of triads a rater had to judge at 464 (58 words x 4 subjects x 2 A-B positions).

Raters listened to the sets through SONY MDV-600 headphones and were asked to choose which of the words, the second or third, sounded more similar in accent to the first. Raters were instructed to press the key labeled "2" on the keyboard, if the second word sounded more similar in accent to the first; or to press the key labeled "3" on the keyboard if the third word sounded more similar in accent to the first. These instructions were used in order to direct raters' attention to qualities of the accent, rather to other qualities of the recording (e.g., background noise) on which they may have based their judgments.

6.3.2 Results and Discussion

The purpose of Experiment 1b was to evaluate subjects' alignment to accent-general information during accented speech perception. If subjects are aligning to an accent, then raters should judge an utterance shadowed after a model with one accent as more similar to model with the same accent than to a model with a different accent.

Mean *accent* alignment was calculated based on rater and same accent model as determined by the number of shadowing subjects' utterances chosen as sounding more similar in accent to those of the same accent model. The mean percentage of shadowing subjects' tokens considered to be more similar in accent to the same accent models' tokens was 49.7%. This percentage was compared to chance (50%) using a t-test, which revealed that the shadowing subjects' tokens were not judged to more similar in accent to the same accent models' tokens than were the different accent models' tokens [$t(15) = -.051$. $p = .960$, Cohen's $d = -.026$].

A repeated-measures ANOVA was conducted to test the between-subjects factor of comparison model (Sp1, Sp2, Ch1, Ch2) and the within-subjects factor of word frequency (HF, MHF, MLF, LF) on accent alignment ratings. There was a significant effect of comparison model [$F(3, 12) = 107.74$, $p = .000$, $\eta_p^2 = .964$]. Pairwise comparisons showed that raters were significantly more likely to make matches to comparison model Sp1 (M = 77.7%) than comparison model Ch1 (M = 21.7%) at the $p < .000$ level. There was no significant difference between matches made to comparison model Sp2 (M = 51.3%) or comparison model Ch2 (M = 48.3%). The repeated measures revealed no significant effect of frequency [$F(3, 36) = 1.01$, $p = .400$, $\eta_p^2 = .078$] and no significant frequency x comparison model interaction.

A one-way ANOVA of accent (Spanish, Chinese) revealed a significant difference in matching between accent groups [$F(1, 15) = 16.01$, $p = .001$]. When separately compared against chance, Spanish-accent models (M = 64.5%) were

matched to the shadowed token at greater than chance levels [$t(7) = 2.77$, $p = .027$, Cohen's $d = 2.09$], while Chinese-accented models (M = 35%) were also significantly different from chance [$t(7) = -2.882$, $p = .023$, Cohen's $d = -2.18$], but in the opposite direction.

As a whole, the results of Experiment 1b are inconclusive regarding accent alignment. Raters did not judge overall alignment to accent when comparing subjects' shadowed utterance to a same accent vs. a different accent comparison model. Further, the significance of ratings for Spanish-accent shadowers occurs in a direction appropriate to be called alignment, while this significance for Chinese-accent shadowers occurs in the completely opposite direction. At first glance, these results seem to indicate that shadowers were aligning to Spanish-accented models and diverging (i.e., adjusting their speech to make themselves distinct) from the Chinese models (Giles & Ogay, 2007). However, these findings are more likely representing a bias in the direction of choosing Spanish-accented models as the match. Post-hoc analyses addressing a potential response bias are addressed next.

Response Bias Analysis. To assess response bias in the present experiment, d' and λ were calculated in a manner appropriate for a *two-alternative forced choice* (2-AFC) task. When a study is performed using a forced-choice procedure (i.e., where each trial contains the signal or stimuli), then a bias in detecting the stimuli is no longer an issue (Wickens, 2002). However, Wickens (2002) suggests that subjects in these types of task can show a bias to choose one response over another for "idiosyncratic" reasons (e.g., preference for one stimuli above another).

In the case of Experiment 1b, raters were presented with a 2-AFC task, where they were always asked to make a forced-choice response selecting either a Spanish-accented model or a Chinese-accented model as correct. Here, a response bias would be occurring if raters were consistently selecting the Spanish-accented (or Chinese-accented) model in all trials. Raters would be "correct" 100% of the time for trials where the shadowed utterance was based on shadowing a Spanish-accented model and "correct" 0% of the time if that utterance had been based on shadowing a Chinese-accented model.

A one-sample t-test revealed that the d' (M = -0.02) in the present study was not significantly different from zero [$t(15) = -0.91$, $p = .38$] suggesting that λ is a more appropriate measure of response bias[3] (Wickens, 2002).

In 2-AFC tasks, λ is equal to c, which has a zero-point halfway between the means for the noise and signal-to-noise distributions. To assess the direction of the bias, Spanish-accented models were chosen as the "signal" distribution and Chinese-accented models were chosen as the "noise" distribution. In this case, a positive result for λ would represent a bias towards choosing Spanish-accented models, while a negative result would be a bias to choose Chinese-accented models. The presence of a response bias was assessed using a one-sample t-test to establish whether λ was equal to zero. The results revealed that λ (M = .42) was significantly different from zero [$t(15) = 3.55$, $p > .01$]. The positive nature of this result suggested that raters were

[3] Although $\log\beta$ is also used as a measure of response bias, λ is a more sensitive measure than $\log\beta$ in instances where d' values are close to zero (i.e., where the noise and signal-to-noise curves to sit on top of each other), as in Experiment 1b.

more likely to respond that the Spanish-accented model was correct, regardless on what accent the shadowed utterance was based.

There are a couple of reasons why this response bias in the direction of Spanish-accented models might be occurring. First, raters may have found the Spanish accent more familiar than a Chinese accent, making it a more salient choice in the matching task. In fact, Scales, Wennerstrom, Richard, & Wu (2006) found that American, undergraduate listeners asked to identify the country of origin for native and accented talkers, were better able to make this identification for Spanish- than Chinese-accented talkers. These authors also observed that these undergraduates *preferred* a Spanish accent over a Chinese accent, even though neither was considered more intelligible (i.e., "easy to understand").

Likewise, it is possible that raters made matches because Spanish-accented talkers sound "less accented" than Chinese-accented talkers. For example, Flege (1988) found that native Chinese talkers who learned English at an early age had a perceptible accent, while Flege & Fletcher (1992) did not find a comparable result for native Spanish talkers with similar learning backgrounds. In this vein, raters in Experiment 1b may have matched shadowed words to the Spanish-accented models because these models sounded less accented than the Chinese-accented models. This possibility will be addressed in Experiment 1c.

6.4 Experiment 1c

Experiment 1a suggests that subjects are aligning to talker-specific information when perceiving accented speech. However, the results of Experiment 1b are unclear. It could be that subjects are aligning to the Spanish accent, while others are diverging from the Chinese accent. On the other hand, the analysis of λ establishes that a response bias exists where raters are more likely to match shadowed tokens to the Spanish-accented models, regardless of what accent was originally shadowed. As addressed in Experiment 1b (see 6.2.2), raters might simply be matching more to the Spanish models because Spanish-accented models sound "less accented" than Chinese-accented models.

In order to examine whether level (or amount) of accentedness of a given model is influential in these judgments, a third experiment asked judges to rate the accentedness of the models (see Sidaras et al., 2009). Judges were presented with accented words and asked to indicate how accented a given word sounds on a scale of 1 ("not at all accented") to 7 ("very accented"). Accentedness ratings were examined across models to determine if there were significant differences in accentedness between models. These accentedness ratings were then compared to the XAB responses (i.e., matches) made by raters in Experiment 1b to see if there was a connection between rated accentedness and response bias.

6.4.1 Method

6.4.1.1 Participants

The Spanish- and Chinese-accented models used in Experiment 1a were the same for this experiment. Ten new undergraduates (1 male, 9 female) aged 18 to 20 acted as judges in order to rate the level of accentedness of the models' words. All raters were native speakers of American English with normal hearing. Participants were recruited from the University of California, Riverside and received credit in order to partially fulfill a course requirement. None had participated in Experiment 1a or 1b.

6.4.1.2 Materials and apparati

All materials and apparati were the same as those used in Experiment 1a. In this experiment, only the audio-recorded utterances from *non-native* models (i.e., Spanish-accented; Chinese-accented) were used as stimuli (see Procedure 5.1.1.3 for more information).

6.4.1.3 Procedure

The judges rated the level of accentedness for each of the non-native models (2 Spanish-accented, 2 Chinese-accented) using a seven-point, Likert-type scale from 1 = "not at all accented" to 7 = "heavily accented" (e.g., Sidaras et al., 2009).

All judges were randomly presented 116 words (29 HF, 29 MHF, 29 MLF, 29 LF) from each model. Words were repeated across models, so each judge was asked to rate the level of accentedness for 116 words x 4 models (Sp1, Sp2, Ch1, Ch2) for a total of 464 word trials.

Judges listened to the trials through SONY MDV-600 headphones and were asked to make their responses on a standard keyboard. Judges were instructed to use the number row on the keyboard and press the numbers "1" through "7", which corresponded to the perceived level of accentedness for each word. Judges were asked to use the full scale when making their judgments and to make their ratings based both on their own knowledge of the standard, American "accent" and in comparison to other words in the list (Sidaras et al., 2009).

6.4.2 Results and Discussion

Mean *level of accentedness* (LoA) was calculated based on pooled judges' ratings per talker. The overall mean LoA pooled across talkers was 4.05 (based on the 7 point scale).

A one-way ANOVA was performed to determine the effects of model (Sp1, Sp2, Ch1, Ch2) on ratings of accentedness (LoA). Results revealed a significant model effect [$F(3, 27) = 18.376$, p = .000], suggesting that some models were judged as being more (or less) accented than others. Pairwise comparisons showed that Chinese-accented Model 1 (Ch1) was judged as being significantly more accented than any of the other three models: Sp1 – Ch1 [$t(9) = -6.120$, p = .000, Cohen's $d = -4.08$]; Sp2 – Ch1 [$t(9) = 5.839$, p = .000, Cohen's $d = 3.89$]; Ch1 – Ch2 [$t(9) = 9.507$, p = .000, Cohen's $d = 6.34$].

An additional repeated measures ANOVA was conducted looking at the effects of accent group (Spanish, Chinese) and word frequency (HF, MHF, MLF, LF)

on level of accentedness. There was a significant effect of accent group on ratings of LoA [$F(1, 38) = 6.55$, $p = .015$, $\eta_p^2 = .147$] with Spanish-accented models being considered less accented than Chinese-accented models. There was also a significant main effect of word frequency [$F(3, 114) = 79.93$, $p < .001$, $\eta_p^2 = .678$]. Pairwise comparisons revealed that across accent type the HF words were judged as significantly less accented than MLF and LF; while LF words were significantly more accented than all other levels of word frequency (at the $p < .01$ level).

Additionally, there was a significant word frequency x accent interaction [$F(3, 114) = 7.62$, $p < .001$, $\eta_p^2 = .167$]. Separate repeated measures ANOVAs show a significant effect of frequency on LoA for both Spanish-accented [$F(2.61, 49.66) = 60.01$, $p < .001$, $\eta_p 2= .760$] and for Chinese-accented [$F(3, 57) = 28.36$, $p < .001$, $\eta_p^2 = .599$] models. Mauchly's test indicated that the assumption of sphericity was violated ($\chi 2(5) = .13.18$, $p = .022$) for Spanish-accented models, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .87$) [4].

Post-hoc results suggested that for Spanish-accented models, there was no difference in accentedness between HF-MHF and MHF-MLF, however for Chinese-accented models all word frequency levels differed from each other. The effects of accent and frequency on judged LoA are depicted in Figure 3**.** Similar to the findings of Levi et al. (2007), the effect of frequency on LoA (i.e., LF words being considered

[4] In order to correct for the violation of sphericity, Huynh-Feldt correction was selected because the epsilon value was greater than .75. Where this value is < .75, Greenhouse-Geisser corrections will be used.

more accented than HF words) suggests that talker-independent factors can influence the perceived magnitude of an accent.

Figure 3

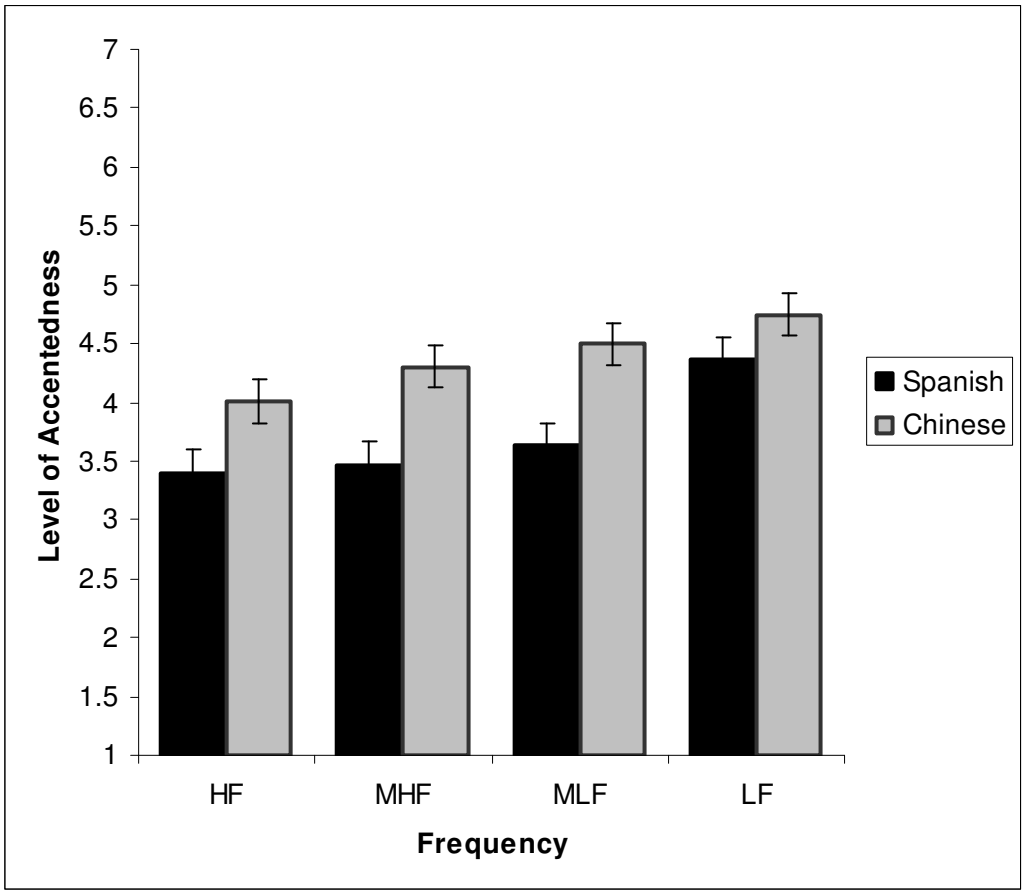

Figure 3. Graph displays results showing a frequency x accent interaction. Spanish-accented models were rated as less accented than Chinese-accented models of word frequency. This figure also shows the this increase in ratings of level of accentedness from HF to LF words, as found in Levi et al. (2007).

These results indicate that there are differences between LoA of individual models and the overall LoA for different accent groups. Spanish models were rated as

having less accent (mean LoA = 3.72) than Chinese accented models (mean LoA = 4.39). Spanish-accented models might be considered perceptually less accented by judges in the present study because the geographical location of this study (e.g., Southern California) allows for more familiarity with Spanish-accented than Chinese-accented speech. However recall that Scales et al. (2006) found no perceived difference in subjective "ease of understanding" between Spanish-accented and Chinese-accented speech, although listeners had a preference for Spanish-accented speech.

Similarity in language structure may also be a factor in Spanish-accented models sounding less accented and being picked more often. There are several differences in the phonemic structure of Spanish and Chinese as they relate to English (Finegan, 2004). The Spanish language, for example, is more similar phonetically to English than the Chinese language. For instance, English and Spanish share seven phonemes (e.g., /b/, /g/) not found in Chinese, while English and Chinese share one phoneme (e.g., /x/) not found in Spanish (American Speech-Language-Hearing Association, n.d.). These phonemic differences could be a result of English and Spanish sharing a language family (e.g., Indo-European), while English and Chinese are from different language families (e.g., Sino-Tibetan)(Crystal, 1987).

Finally, there are differences in how much time accented models were in the U.S., as well as how long they had been speaking English. For example, though Spanish-accented models had more time in the U.S. (M = 7 years) than Chinese-accented models (M= 1.5 years), these models had less time spent speaking English

(M = 7 years) than Chinese-accented models (M = 14.5 years). The question of whether language exposure (i.e., being in the U.S.) is more or less important for accent reduction than language knowledge (i.e., learning English) is not in the scope of this dissertation. However, it is clear that being in a language community for a period of time will have an affect on an individuals' speaking habits and, thus, may reduce an accent (e.g., Sancier & Fowler, 1997; Trudgill, 1986; Wolfram, Carter, & Moriello, 2004).

LoA and Rating Judgments. Models' mean LoA (Exp 1c) and XAB rating judgments (Exp 1b) are depicted in Table 2**.** Combined with the response bias in Experiment 1b, these findings seem to suggest that when raters are asked to judge alignment to accent (see 6.2.1.3), they are more likely choosing the least accented models as more similar in accent to the native, English speech. Recall that listeners learn to recognize accented speech more quickly and accurately based on familiarity with the accent (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009). Perhaps, judges in the present study are highly familiar with the Spanish accent making it seem "less accented" in general. Beyond geographical familiarity with Spanish accents, the Spanish language also shares more phonetic similarity to English than does the Chinese language (ASHA, n.d.). Taken together, these results suggest that in order to resolve the response bias and uncover alignment to accent, it will be necessary to perform future experiments that control for LoA.

Table 2.

Level of Accentedness and rating judgments for each accent.

Accent	LoA[a]	Rating[b]
Spanish	3.72	.645
Chinese	4.39	.350

Note[a]: LoA (Exp 1c) was measured on a 1 ("not at all accented") to 7 ("heavily accented") scale. Note[b]: Rating is the average number of times a rater (Exp 1b) judged the Spanish- or Chinese-accented model as the correct response in an XAB matching task.

6.5 Discussion of Experiment Series 1

The results of the present series of experiments suggest that complete talker normalization is not occurring in the context of accented speech. In particular, alignment to an accented *talker* (Experiment 1a) suggests that talker-specific information is encoded and capable of influencing speech productions. This finding differs from the results of Sidaras et al. (2009) who showed no significant influence of talker in the context of a transcription task. As mentioned previously, the alignment methodology has been proffered as revealing a direct and immediate influence of speech perception on speech production (e.g., Goldinger, 1998). Thus, the use of an alignment task and measuring influences of talker on speech productions may have been more revealing of talker information. This evidence could also suggest that the alignment (shadowing) task provides more encoding of talker-

specific information than does a transcription task. This possibility, as well as others, will be addressed in Experiment Series 2.

In addition, evaluations of alignment based on word frequency suggest that alignment to accented *talkers* is sensitive to the effects of a talker-independent characteristics. Thus, talker alignment in the context of accented speech may be further mediated by an underlying interaction between word frequency and level of accentedness (i.e., HF words are perceived as less accented allowing greater encoding of talker-specific information). This contention requires additional investigation.

What is not clear from this series of experiments is whether accent-general information influences speech productions. While speech alignment is understood to be a cognitive process that compels individuals to sound more similar in pronunciation, rate and intonation either to isolated speech stimuli or to a conversational partner (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Namy et al., 2002; Pardo, 2006; Shockley et al., 2004), it may not guide this tendency in the direction of a specific accent. The response bias of Experiment 1b makes it difficult to say whether or not accent alignment is occurring. Future studies where LoA of models is controlled (e.g., comparison models matched on level of accentedness in the XAB task) might reveal alignment to a given accent.

Chapter 7

Experiment Series 2

7.1 Introduction

Experiment Series 1 provided evidence that talker-specific information *can be* encoded and influences speech productions (i.e., through talker alignment) during the perception of accented speech. This raises the question as to why the results of Sidaras et al. (2009) did not show significant talker influences for perceptual identification of novel, accented speech.

Recall that the general design of Sidaras et al. (2009) used a high variability training and test paradigm. In this paradigm, listeners were presented multiple, accented talkers at *both* training and test to assess the influences of accent and talker familiarity on perceptual identification accuracy. The authors trained listeners on multiple, Spanish-accented talkers by having them rate the accentedness and transcribe speech from these talkers. At test, listeners transcribed novel speech from either the same or a different group of Spanish-accented talkers. If learning occurred for accent-general information, those trained on Spanish-accented speech should have better perceptual identification accuracy at test than the control groups (i.e., listeners with no training or who were trained on native English talkers). If talker-familiarity was also influencing perception of Spanish-accented speech, then listeners trained

and tested on the same group of talkers should have shown significantly higher perceptual identification accuracy than those tested on the different group of talkers.

Although, Sidaras et al. (2009) found that listeners in the Spanish-training groups performed significantly better than did the control groups; there was no significant differences found between groups tested on the same vs. different talkers. In other words, being familiar with a Spanish accent helped listeners perceive Spanish-accented speech better, regardless of whether or not they were familiar with the talkers. The following sections discuss differences that may account for why talker-specific influences occurred in Experiment 1a, while they did not occur in the research of Sidaras et al.

7.1.1 Single- vs. Multiple-Accented Talkers

There are several important differences to note between the design of Experiment 1a and the Sidaras et al. (2009) study. First, although both studies used accented talkers as their models, Experiment 1a and 1b presented each subject with a single talker instead of the six talkers presented by Sidaras et al. (2009). Bradlow & Bent (2008) suggest that high variability training (i.e., training on multiple, accented talkers) may be important for learning non-native phonemic contrasts. They further suggest that exposure to a wide range of stimuli produced by a single talker or limited range produced by multiple talkers may offer alternative means to becoming better at perceiving foreign accented speech. Perhaps, the large amount of accent-general information did reduce the influence of any one talker in the research of Sidaras et al.,

while a wide range of stimuli from one talker allowed more efficient encoding of talker-specific information in Experiment 1a.

7.1.2 Alignment vs. Perceptual Identification Measures

The *measure* of talker familiarity effects also differs between these studies. Experiment 1a considered perceptual changes in a listener's speech productions (i.e., alignment) as indicative of the influences of talker-specific information. However, Sidaras et al. (2009) looked to improvements in perceptual identification accuracy as representative of talker effects.

The alignment vs. perceptual identification tasks may reveal differences in how episodic traces (that include talker-specific information) are accessed across tasks. For example, Goldinger and Azuma (2004) suggest that alignment (or imitation) may be driven by the degree of activation of stored traces (echo *content*), which is a unique grouping of the weighted averages of relevant traces. Recognition, on the other hand, might reflect the sum of the total activation of memory probes (echo *intensity*), which reaches a threshold and signals familiarity. In fact, Goldinger and Azuma revealed a dissociation between alignment results and the results of a recognition memory task. These authors found alignment to words presented in a listening task, even though subjects were never asked to shadow the talkers. The level of perceived alignment, as measured in an AXB-matching task, was affected by word frequency and number of repetitions, which is in line with previous alignment findings (Goldinger, 1998).

Goldinger and Azuma (2004) also found significant recognition for words previously heard, which gradually improved with repetitions. Yet, the authors discovered that many of the words with high recognition accuracy had failed to induce significant alignment. They propose that this result may be due to the differences in how episodic traces are accessed across tasks.

This interpretation could potentially explain some of the differences in results found between Sidaras et al. (2009) and Experiment 1a. In the research of Sidaras et al., if memory influences are based on the sum of all activation, the intensity of the echo could have been most influenced by similarity of the original *accented* stimuli to other *accented* traces. This would make later perception of accented speech better, regardless of talker. In Experiment 1a, the content of the echo may have included traces with both *talker* and accent information, but the weighting was more in the direction of the *talker*. This could be why subjects showed talker alignment in Experiment 1a.

It is important to note that there are differences in how explicit (i.e., recognition) and implicit (i.e., perceptual identification) memory tasks are influenced by talker-specific characteristics (e.g., surface characteristics tend to affect implicit tasks for longer periods of time; Goldinger, 1996). So, the explanation regarding recognition memory as reflecting echo *intensity* may not be completely applicable to perceptual identification. However, Lachs, McMichael, & Pisoni (2003) suggest that parallel effects found between the explicit (recognition) and implicit (perceptual

identification) memory paradigms indicate a single memory system that stores highly-detailed traces of speech events.

7.1.3 Shadowing vs. Transcription: Encoding Differences

One last reason that talker-specific influences occurred during Experiment 1a could stem from the immediate and productive nature of the *alignment (shadowing) task.* This type of task may allow for greater encoding of talker-specific information. Recall that Sidaras et al. (2009) trained listeners using a transcription task where they were presented with words over headphones and asked to type out the words they heard. Although this type of transcription training quite commonly reveals that talker familiarity improves later perception of native (unaccented) speech (e.g., Nygaard & Pisoni, 1998; Nygaard et al., 1994), it does not do so for accented speech. It could be that the encoding of talker-specific information is reduced when a transcription response is made to *speech* stimuli, at least for accented speech. Whereas the immediacy of a speech (shadowed) response might allow for (or improve) the encoding of talker-specific information during accented speech production.

In fact, *speech* responses are exceedingly fast (e.g., Fowler et al., 2003; Porter & Castellanos, 1980; Porter & Lubker, 1980) and response times exceedingly small (Fowler et al., 2003; Porter & Castellanos, 1980; Porter & Lubker, 1980) when these responses are made to *speech*, as opposed to non-speech stimuli. This suggests that shadowing accented speech provides a more immediate response, potentially allowing for the encoding of more talker-specific information. In addition, shadowed response

to accented speech are of a productive nature much like subject-performed events (e.g., Bäckman et al., 1986). This may suggest that motor features (e.g., gestures) and surface characteristics (e.g., idiolect, accent) of speech are automatically encoded in memory much like they are during enactment (e.g., Bäckman et al., 1986; Cohen, 1983).

A shadowing task where subjects respond immediately by producing speech may improve encoding of talker-specific information enough to reveal talker familiarity effects on accented speech perception. This final point, along with other questions, will be addressed in Experiment Series 2.

7.2 Experiment 2a

Experiment 2a replicates and extends the study of Sidaras et al. (2009), by having groups of listeners perceptually identify words produced by either the same or a different group of Spanish-accented talkers from training to test. However, during training, groups of listeners are asked to either transcribe or shadow Spanish-accented words. At test, all listeners are then asked to identify (through transcription) novel words produced by either the same or a different group of Spanish-accented talkers. If the immediate and productive nature of the shadowing task helps with the encoding of talker-specific information, listeners should show greater perceptual identification accuracy for accented speech in the *same* vs. *different* talker conditions even though they are transcribing at test. In particular, the *perceptual identification accuracy* for the group asked to shadow words from the *same* talkers should be greater than for the

group asked to shadow *different* talkers, and these groups should be greater than those simply asked to transcribe the words during training.

7.2.1 Method

7.2.1.1 Participants

Models. Eight non-native (4 male, 4 female) and four native, monolingual English (2 male, 2 female) talkers acted as models in the experiment and produced the original stimuli word list. The eight non-natives talkers were from a Mexican-Spanish language background. Spanish-accented talkers were selected in order to keep the current study consistent in language background with Sidaras et al. (2009). A description of non-native models' English language exposure is presented in Table 3. Non-native talkers were recruited through the university extension center, through local community college, ESL programs, or through an online community. The native English talkers were born and raised in the United States and were recruited from the University of California, Riverside. Models were financially compensated for their participation in the study.

Listeners. Seventy undergraduate subjects (29 male, 41 female) aged 18 to 32 acted as listeners who were trained and then tested on the models' words. Recall that Namy et al. (2002) showed that female shadowers align more than male shadowers (but see, Pardo, 2006). Although the present experiment uses a shadowing task, the measure being performed was not speech alignment. For that reason, male and female subjects both participated in the present study. These subjects were native, American-

English talkers from monolingual households with no speech impediments. The subjects were recruited from the University of California, Riverside and participated in order to partially fulfill a course requirement. All participants had corrected-to-normal hearing and vision.

Table 3.

Language background information for accented models (Experiment Series 2)

	Language Background				
Model	Gender	Age	Time in U.S.	Age of English Exposure	Time Speaking English
F1	Female	21	11	10	11
F2	Female	41	3	38	3
F3	Female	41	18	23	18
F4	Female	22	0	4	18
M1	Male	20	1	19	1
M2	Male	28	12	16	12
M3	Male	32	28	4	28
M4	Male	23	18	6	17

Note. All values are in years

7.2.1.2 Materials and apparatus

A list composed of 200 English words was compiled and recorded as stimuli (see Appendix B). The list consisted of monosyllable and bisyllable words from four frequency classes (Kučera & Francis, 1967): High frequency (HF; >300 occurrences per million), medium-high frequency (MHF; 150-200 occurrences per million), medium-low frequency (MLF; 50-100 occurrences per million), and low frequency (LF; <5 occurrences per million) (e.g., Goldinger 1998). A SONY DSR-11 camcorder and digital recorder were used to videotape the models. Only the audio-component of the recordings was used for the present study. All words were edited into separate audio files and amplitude equalized so each word had an average RMS amplitude of -45.00 dB.

In order to group accented models for the experiment, a pilot test was conducted where native English listeners were asked to rate the accentedness and the intelligibility of the eight Spanish-accented models. Ten listeners rated the level of accentedness of 75 words from each of the accented models on a 1 (“not at all accented”) to 7 (“heavily accented”) Likert-type scale. An additional group of ten listeners were asked to transcribe 50 words from each of the eight accented models with the mean accuracy being considered a measure of baseline intelligibility. In order to properly counterbalance, models were divided into two groups based on their mean accentedness with each group containing two male and two female talkers (see Table 4). In comparison to Sidaras et al. (2009) whose models had a mean intelligibility 48.3% across the groups, the intelligibility of the present models was

quite high (82.6% across groups). For this reason, test words were mixed with white noise at a +0 signal-to-noise (SNR) ratio to reduce the possibility of ceiling effects in the speech identification task. Although Sidaras et al. did not add white noise to word stimuli, they did so for their sentence stimuli for similar reasons.

Table 4.

Accentedness and intelligibility for Spanish-accented talkers.

Talker Group	Gender	Mean Accentedness Ratings[a]	Mean word intelligibility (%)
Spanish Group 1	Female	4.12	82.0
	Female	4.36	82.0
	Male	3.60	81.0
	Male	5.52	79.6
Spanish Group 2	Female	4.28	86.2
	Female	5.33	76.6
	Male	4.06	85.4
	Male	4.98	79.8

Note[a]: LoA (Exp 1c) was measured on a 1 ("not at all accented") to 7 ("heavily accented") scale.

From the original recorded word list, 104 words were selected as training and test stimuli for the experiment to represent an equal distribution of frequency classes. All stimuli were presented to participants using PsyScope software. All text words

and feedback were presented on a 20-in. video monitor positioned 3 ft in front of the participants. Auditory stimuli were presented through SONY MDR-V6 headphones. The models and listeners responded verbally into a Shure Beta 58a microphone and were audio-recorded at a 44000 Hz, 16 bit rate using Amadeus software. Typed responses and accentedness ratings were collected using a standard keyboard.

7.2.1.3 Procedure

The experiment consisted of a training phase, which differed across conditions, and a test phase. During training, listeners were presented with words either spoken by one of two groups of four Spanish-accented talkers (group 1, group 2) or four native English talkers (group 3), or as text words presented on a monitor (group 0). For the Spanish-training and English-training groups, listeners were further divided into transcription or shadowing conditions during training. The English training and text-reading groups served as controls. This provided a total of seven training conditions. During the test phase, conditions were counterbalanced so that half the listeners heard Spanish-accented group 1 and half heard Spanish-accented group 2. All listeners received an 8 word practice block according to their condition.

Training Phase. The training phase was comprised of four comparison and three variability blocks presented in alternating order. In Sidaras et al. (2009), comparison blocks seem to be used in order to familiarize the listeners to the talkers, while variability blocks provide the opportunity to assess improvements in accuracy over the course of training. In the present study, improvements over training were not assessed due to the different responses collected from a shadowing vs. transcription

task. During a comparison block, listeners heard a total of 4 words each produced by four Spanish-accented (or English-control) talkers for a total of 16 comparison trials. The listeners were asked to rate the level of accentedness for each word using a seven-point, Likert-type scale, from 1 = "not at all accented" to 7 = "heavily accented" (e.g., Sidaras et al., 2009). In order to keep the tasks consistent, listeners in the English-training conditions were asked to rate the level of "dialect" for the English talkers using a 7-point scale. Listeners were asked to use the full scale to make their responses and to compare the words to their knowledge of a standard-American accent and other words in the list.

During a variability block, listeners heard two sequential repetitions of 16 words presented randomly and matched in such a way that listeners never heard the same word paired with the same talker more than once. During these blocks, listeners in the transcription group were asked to identify the words they heard by typing them. Transcription was self-paced, with the listener entering their response to move the trials forward. Listeners in the shadowing group were asked to identify the English word they heard "quickly, but clearly" into the microphone. Because no typed response was needed, the experimenter pressed a key after the listeners made a vocal response to move the trials forward. After a response in either group, the intended word was displayed on the screen and played over the headphones.

Sidaras et al. (2009) includes a No-Training control condition, where listeners were simply tested on their ability to perceive the accented speech (i.e., they completed the test phase). The text-reading control group in the present study did

complete a training task, where they were presented identical words to other training groups in seven blocks. During the blocks, these control subjects were asked to identify how many syllables the word contained by pressing "1" on the keyboard for monosyllable words and "2" on the keyboard for bisyllable words. There was a nearly equal distribution of mono- and bisyllable words as depicted in Appendix B.

Test Phase. The test phase consisted of a single block, for which all listeners were presented with Spanish-accented words. Listeners heard a total of thirty-two novel words produced by one of two sets of four Spanish-accented talkers (2 male, 2 female). Using software, eight words per talker were presented randomly in a background of white noise. No feedback was given. Dependent on the training group, listeners would encounter a familiar accent and familiar talkers (same condition), a familiar accent and unfamiliar talkers (different condition), or an unfamiliar accent and unfamiliar talkers (control conditions). Listeners were asked to identify the words they heard by typing them on a keyboard. Transcription was again self-paced.

7.2.2 Results and Discussion

In the present experiment, a shadowing task was implemented within the framework of a high variability training and test paradigm (Sidaras et al., 2009) to assess whether or not this task would improve the encoding of talker-specific information during training on Spanish-accented speech. Perceptual identification accuracy at test was averaged across words for each listener. A word was considered accurately identified (i.e., correct) if listeners provided the correct spelling or a

homophone equivalent. All scores are based solely on transcription accuracy during the *test phase*.

A oneway ANOVA was performed to compare perceptual identification accuracy for the different training-test conditions: (1) Shadowing-Same, (2) Shadowing-Different, (3) Transcription-Same, (4) Transcription-Different, (5) English, Shadowing-Different, (6) English, Transcription-Different, and (7) Text Reading-Different, where conditions 1-4 were training conditions and 5-7 were control conditions. There was no significant difference between conditions [$F(6, 279) = .492$, $p = .814$]. Figure 4 shows overall perceptual identification accuracy for each condition.

Additionally, a repeated-measures ANOVA was performed to assess the between-subjects effects of training (Spanish-training vs. English-training and Text Reading) and the within-subjects effect of word frequency (HF, MHF, MLF, LF) on perceptual identification accuracy. There was no effect of training [$F(1, 48) = .040$, $p = .842$, $\eta_p^2 = .915$], suggesting that listeners who had Spanish-accented training did not perform better than control listeners in their subsequent perception of Spanish-accented speech. There was, however, a significant effect of word frequency [$F(3,144) = 28.81$, $p = .000$, $\eta_p^2 = .375$]. There was a significant difference in accuracy at all levels of word frequency, as depicted in Figure 5. Listeners were the best at identifying HF words (M = .50) and the worst at identifying LF words (M = .153). These results could be due to the fact that HF words are perceived as less accented than LF words (e.g., Levi et al., 2007) or because more exemplars exist of

high than low frequency words making them more easily recognized (e.g., Goldinger, 1998). There was no significant training x frequency interaction [$F(3,144) = .199$, $p = .897$, $\eta_p^2 = .004$].

Figure 4

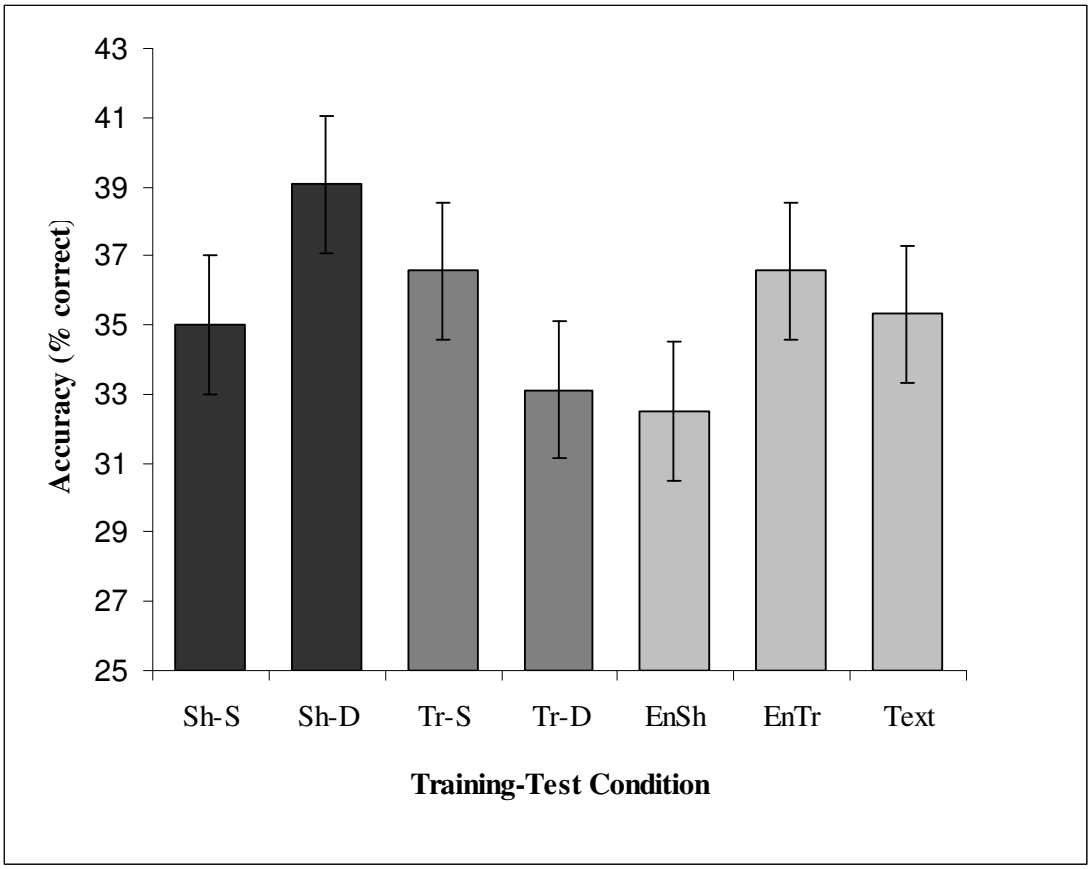

Figure 4. Shows perceptual identification accuracy (% correct) for each training-test condition. Training-test conditions include: Shadowing-Same, Shadowing-Different, Transcription-Same, Transcription-Different, English-Shadowing, English Transcription, and Text-Reading.

The results of the present study revealed no effect of Spanish-accented training on later perceptual identification of Spanish-accented speech when compared

to control conditions. This is in contrast to the results of Sidaras et al. (2009) who found that trained listeners improved in their perception of Spanish-accented speech compared to control listeners, regardless of talker familiarity.

Further, the two critical training conditions (Shadowing vs. Transcription) and test conditions (Same vs. Different) were also not significantly different from one-another. Recall that the predictions for Experiment 2a were that listeners in the Shadowing condition would perform better overall at test than listeners in the Transcription conditions, and that these shadowing subjects would show talker familiarity effects at test (i.e., would do significantly better when tested on the same vs. different talkers).

The present study did not show that Spanish-accented training improved later perception of Spanish-accented speech, as it did for Sidaras et al. (2009). The lack of hypothesized results may have occurred for several reasons. First, the present experiment used speech-in-noise stimuli at test, however, training was done on clear speech (i.e., words presented without noise). As mentioned, the average word intelligibility for models in Sidaras et al. (2009) was 48.3% across the groups. In the present study, the average intelligibility of the accented talkers (in the clear) was 82.6% across groups. This large difference made it necessary to reduce the intelligibility of these groups with background noise, in order to avoid the possibility of ceiling effects at test. However, this could have limited effects talker familiarity (e.g., Schacter & Church, 1992; see also Goldinger, 1996).

Figure 5

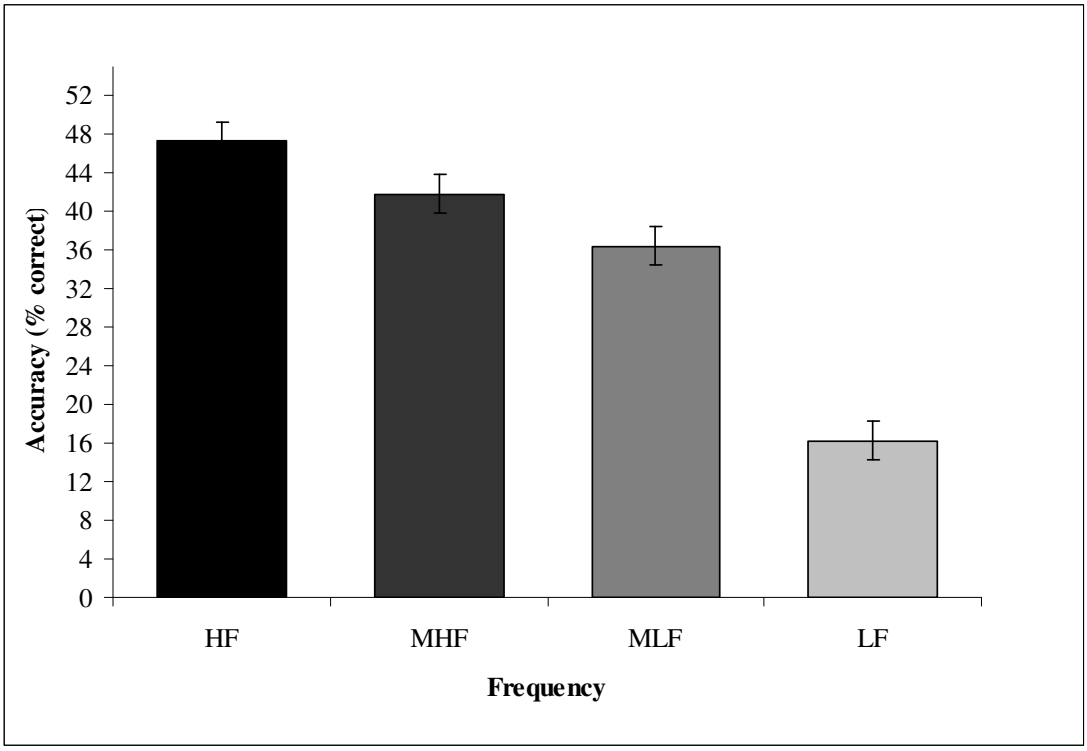

Figure 5. Graph of perceptual identification accuracy (% correct) at each level of word frequency. There was a significant difference between all levels of word frequency.

Next, the geographical location (e.g., Southern California) of this study versus the study of Sidaras et al. (2009) could be a factor in the lack of accent training results. According to the U.S. Census Bureau (2007-09), population estimates for this region boast a Hispanic or Latino population of 48.7% (43.2% Mexican), while the region of Sidaras et al.'s study has a Hispanic or Latino population of 10.5% (6.5% Mexican)(http://www.census.gov). The choice to use Spanish-accented models was made in an attempt to replicate the findings of Sidaras et al. (2009) who elected to use Spanish-accented speech. However, these population numbers clearly show large

differences in demographics, which might have influenced how familiar listeners are with Spanish-accented speech. In fact, Bent and Bradlow (2003) showed that a shared native language background increases perceived intelligibility of non-native speech (see also, Bowers, Mattys, & Gage, 2009). Perhaps the reason the present experiment does not show an improvement from accent training is because listeners were already familiar with this language information. Thus, this choice may have been flawed and training listeners on an accent with which they were less familiar may have been of benefit in this set of studies.

Finally, the reason for the lack of differences between shadowers and transcribers is not clear. It may be due to the general design of Experiment 2a (e.g., stimulus properties, accent familiarity issues) or because the productive nature of shadowing does not increase encoding of talker-specific information. Perhaps, for example, although shadowing influences speech productions that are more similar to a shadowed talker, it may not improve the perception of that talker's speech. In order to clarify the potential encoding influences of shadowing, it is necessary to investigate shadowing vs. transcription using native English speech.

7.3 Experiment 2b

Based on the lack of training results in Experiment 2a, a follow-up experiment was designed to test whether shadowing would improve perceptual identification accuracy over transcription with *native* (unaccented) speech. As mentioned, the immediacy and productive aspects of shadowing might increase the encoding of

talker-specific information. However, shadowing may be similar to transcription in not showing talker familiarity effects on identification accuracy when speech is accented. While previous studies have shown that perceptual identification accuracy improves when listeners are familiar with native talkers (e.g., Nygaard & Pisoni, 1998; Nygaard et al., 1994), these studies used transcription tasks from training to test. Asking listeners to shadow during training may reveal an increase in the encoding of talker information during a later word identification test.

During training, groups of listeners were asked to either transcribe or shadow words spoken by native English talkers. At test, the listeners were then asked to identify (through transcription) novel words produced by either the same or a different group of native English talkers. Listeners should show improvement in perceiving speech when presented with the *same* talkers from training to test (e.g., Nygaard & Pisoni, 1998), regardless of whether they shadow or transcribe. However, if a greater amount of talker-specific information is encoded during shadowing due to the immediacy of the task, then shadowers in either test condition should show significantly higher accuracy than transcribers in either test condition. This finding would suggest that shadowing may increase the encoding of talker-specific information during native speech perception. While, much like with transcription, this encoding may be reduced for accented speech.

7.3.1 Method

7.3.1.1 Participants

The four native English talkers (2 male, 2 female) from Experiment 2a and a new set of four native English talker (2 male, 2 female) acted as models in the experiment and produced the original stimuli word list. All native English talkers were born and raised in the United States and were recruited from the University of California, Riverside and surrounding areas. Models were financially compensated for their participation in the study.

Fifty undergraduate subjects (25 male, 25 female) aged 18 to 23 acted as listeners who were trained and then tested on the models' words. These subjects were native American-English talkers with no speech impediments. The subjects were recruited from the University of California, Riverside and participated in order to partially fulfill a course requirement. All participants had normal-to-corrected hearing and vision.

7.3.1.2 Materials and apparati

The English talker stimuli list and apparati were the same as those used for Experiment 2a, with the exception that additional native English models were recorded producing the list and non-native stimuli were not used in this study (see 7.2.1.2 Materials and apparati for more information).

7.3.1.3 Procedure

The experiment again consisted of a training phase, which differed across conditions, and a test phase. During training, listeners were presented with words

either spoken by one of two groups of native English talkers (group 1, group 2) or as text on a monitor (group 0). For the English-training groups, listeners were further divided into transcription or shadowing conditions during training. The text-reading group served as a control. This provided a total of five training conditions. During the test phase, conditions were counterbalanced so that half the listeners heard native English group 1 and half heard native English group 2.

Training Phase. The training phase was again comprised of four comparison and three variability blocks presented in alternating order. This experimental design only differed from Experiment 2a in terms of instructions during comparison blocks. Instead of rating accentedness, listeners were asked to rate the level of dialect for each word using a seven-point, Likert-type scale, from 1 = "no dialect" to 7 = "heavy dialect". They were asked to use the full scale to make their responses and to compare the words to their knowledge of a standard-American dialect and other words in the list. Variability blocks were designed in the same way as those used in Experiment 2a, as was training in the text-reading condition (see 7.2.1.3 Procedure for more information).

Test Phase. The test phase consisted of a single block, where all listeners were presented with English words. Listeners heard a total of thirty-two, novel words produced by one of two sets of four native English talkers (2 male, 2 female), with eight words being produced by each model. Using software, words were presented randomly in a background of white noise with an SNR of -5. This level of white noise was selected based on pilot measures of intelligibility performed by research

assistants. No feedback was given to listeners. Based on training group, listeners would encounter either a group of familiar talkers (same condition) or a group of unfamiliar talkers (different and control conditions). Listeners were asked to identify the words they heard by typing them on a keyboard. Transcription was self-paced.

7.3.2 Results and Discussion

Similar to Experiment 2a, this follow-up study also involved a shadowing task, which was applied into a high variability training and test paradigm (Sidaras et al., 2009). The goal was to assess whether or not shadowing would induce talker familiarity effects like those found using transcription tasks in previous studies on native (unaccented) speech perception (e.g., Nygaard & Pisoni, 1998; Nygaard et al., 1994). As in Experiment 2a, perceptual identification accuracy was averaged across words for each listener. A word was considered accurately identified (i.e., correct) if listeners provided the correct spelling or a homophone equivalent. All scores are based solely on transcription accuracy during the *test phase*.

A oneway ANOVA was performed to compare perceptual identification accuracy for the different training-test conditions: (1) Transcription-Same, (2) Transcription-Different, (3) Shadowing-Same, (4) Shadowing-Different, and (5) Text-Reading-Different, where conditions 1-4 were training conditions and 5 was the control condition. Figure 6 shows perceptual identification accuracy at test for each condition. There was a marginally significant effect of condition [$F(4,45) = 2.01$, $p = .096$, $\eta_p^2 = .158$], which seems to be accounted for by significant differences between

the Text-Reading-Different (M = .450) and both the Shadowing-Same (M = .52) and Transcription-Different (M = .516) conditions at $p < .05$ level. Though marginal, this effect suggests that training might have an influence on perceptual identification accuracy.

Figure 6

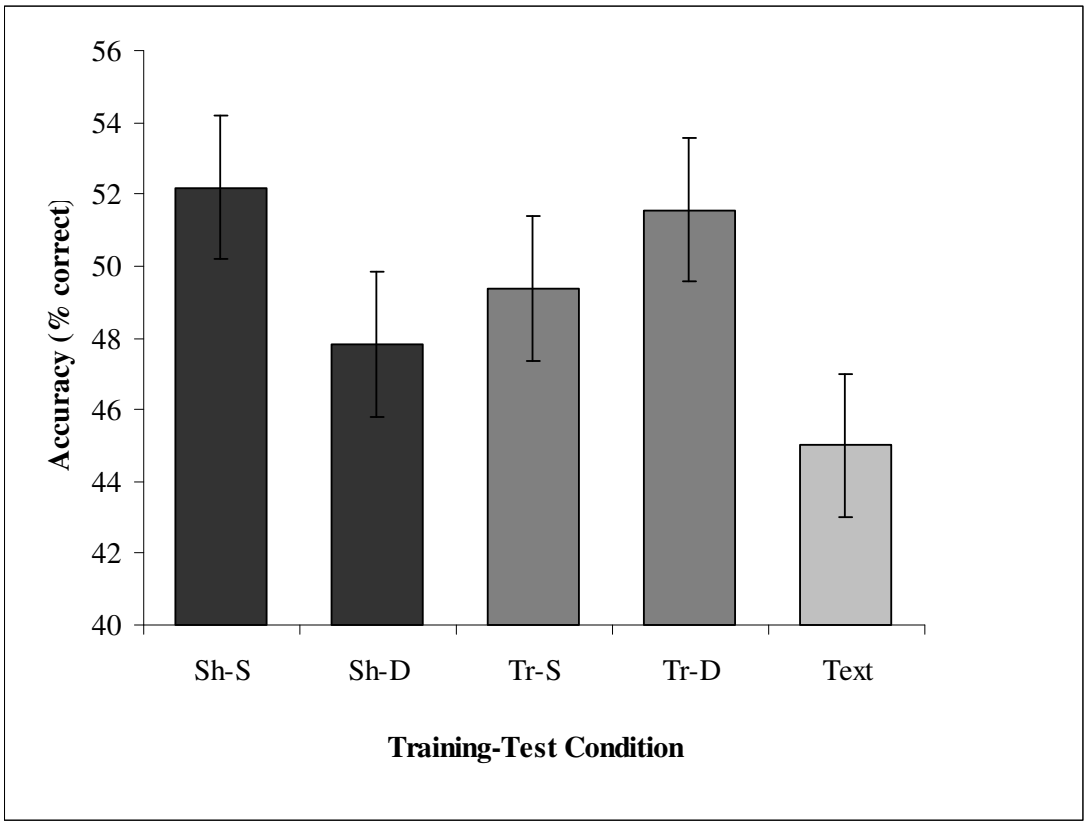

Figure 6. Shows perceptual identification accuracy (% correct) for each training-test condition. Training-test conditions include: Shadowing-Same, Shadowing-Different, Transcription-Same, Transcription-Different, and Text-Reading Control.

A repeated measures ANOVA was conducted to further uncover the effects of the between-subjects factor of training (English-training vs. Control) and the within-

subjects factor of word frequency (HF, MHF, MLF, LF) on perceptual identification accuracy. There was a significant effect of training [$F(1, 48) = .5.41$, $p = .024$, $\eta_p^2 = .101$], suggesting that training helps improve later speech perception. There was also a significant effect of word frequency [$F(3,144) = 30.92$, $p = .000$, $\eta_p^2 = .392$]. Pairwise comparisons revealed that LF words (M = .255) were identified significantly less than words at all other levels of word frequency, at the $p < .001$ level. Identification accuracy for different levels of word frequency is shown in Figure 7. There was no significant training x frequency interaction [$F(3,144) = .476$, $p = .700$, $\eta_p^2 = .010$].

Following the significant differences found between training and control groups (and the marginal effect of condition), a final one-way ANOVA was performed to assess differences between English-training groups (Transcription-Same, Transcription-Different, Shadowing-Same, Shadowing-Different). There was no significant difference in accuracy among these groups [$F(3, 39) = 1.49$, $p = .233$].

Unlike Experiment 2a, the present results showed a significant effect of English-training on later perception of native speech. This suggests that when listeners are trained on native English talkers, they are more accurate at test than listeners without this training. However, much like Experiment 2a, there was no significant difference between the two critical training conditions (Shadowing vs. Transcription) or the test conditions (Same vs. Different Talker). In particular, the finding that familiarity with talkers is not aiding later perception of the same vs.

different talkers runs contrary to previous research on talker familiarity effects (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993).

Figure 7

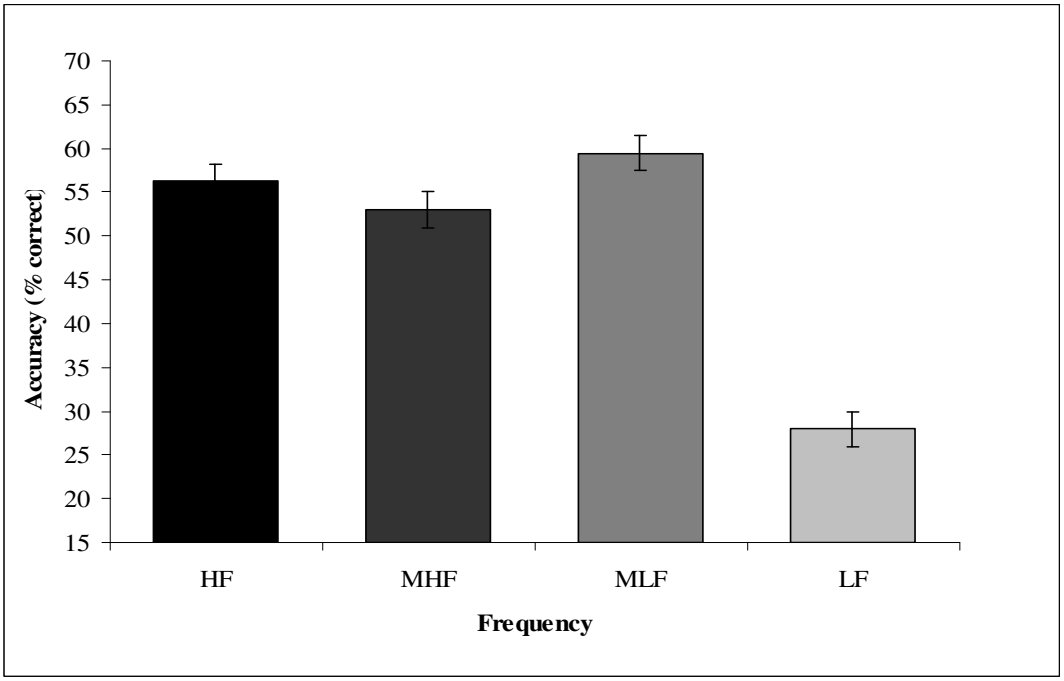

Figure 7. Graph of perceptual identification accuracy (% correct) at each level of word frequency. LF words were identified significantly less than any other frequency. No other differences were significant.

This lack of talker familiarity effects may be driven by the fact that word stimuli were embedded in white noise at test and not during training. Indeed, Goldinger (1996) suggests that a change in stimulus properties (e.g., presence or absence of noise) from training to test may reduce voice effects because this information is encoded alongside other details of the speech event (see also, Schacter

& Church, 1992). If this is the case, then the change in stimulus properties might account for a lack of talker familiarity effects.

A second possible reason for the lack of talker effects in Experiment 2b might be the difference between training (e.g., task and time) in the present study versus training in studies using novel voice training paradigms (e.g., Nygaard & Pisoni, 1998; Nygaard et al., 1994). In the comparison blocks of the study by Nygaard and Pisoni (1998), for example, listeners were trained to identify voices by name (e.g., *Erica* said "Flannel", *Angelica* said "Social") over the course of ten days (see also, Nygaard et al., 1994). However, in comparison blocks during the present experiment, listeners were asked to rate the level of dialect (see 7.2.1.3 for more details) and training lasted less than an hour. The difference in task (i.e., voice identification vs. dialect rating) and time (i.e., ten days vs. ~ 1 hour) may have played a role in the lack of talker familiarity effects occurring in Experiment 2b, and possibly Experiment 2a.

7.4 Discussion of Experiment Series 2

Results of Experiment 2a are not consistent with the results of Sidaras et al. (2009) in showing an effect of Spanish-accented training on the later perception of Spanish-accented speech. For native (unaccented) speech, Experiment 2b showed an influence of training, but there was no difference between groups who had familiar (same) vs. unfamiliar (different) talkers at test. This does not follow from the literature on talker familiarity effects (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993).

These results in both experiments could stem from the difference in stimulus properties between training words in the clear and test words in noise (e.g., Goldinger, 1996; Schacter & Church, 1992). The number of subjects in these two experiments is also rather low (i.e., ten subjects per condition). Previous studies have typically used twice as many subjects (i.e., listeners) per condition (Nygaard & Pisoni, 1998; Nygaard et al., 1994, Sidaras et al., 2009).

One interesting note is that neither Experiment 2a nor 2b revealed differences between listeners in the Shadowing conditions vs. the Transcription conditions. These results are hard to interpret for Experiment 2a because of a lack of training effects. Beyond this, the lack of difference between Shadowing and Transcription could be showing that the immediate and productive nature of the shadowing task does not increase encoding of talker-specific information above the transcription task. Alternatively, the lack of improvement in the Shadowing conditions could be the result of *transfer appropriate processing* masking actual differences. Transfer appropriate processing is a general memory framework that instantiates the importance of overlap between training and testing conditions (e.g., Graf & Ryan, 1990; Morris, Bransford, Franks, 1977). For example, Loebach, Pisoni, & Svirsky (2010) showed that listeners' were better able to transcribe spectrally degraded sentences at test when training was done on degraded, as opposed to undegraded, stimuli. However, the results of the present experiment are again difficult to interpret because they do not replicate classic talker familiarity effects (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993).

In regards to Experiment 2a and 2b, the overlap between the training and test phases for transcribing listeners may limit the interpretation of a lack of difference between Shadowing and Transcription because shadowing subjects did not shadow at test. One way to avoid the problem of transfer appropriate processing would be to include a test phase in the present design where listeners are asked to *shadow* their responses. This and other future directions will be discussed in more details in the General Discussion (8.2).

Chapter 8

General Discussion

The goal of the present dissertation was to examine talker-specific influences on the perception of accented speech by answering the following questions: *Does the perception of accented speech involve a process of normalization or is talker-specific information encoded in memory? If talker-specific information is stored during the perception of accented speech, is it somehow 'masked' by accent-general information? Will using a more immediate and productive encoding task reveal the influence of talker-specific information in the perception of accented speech?*

These questions were addressed in two series of experiments using a speech alignment methodology. The results of the first series of experiments revealed that talker-specific information is, in fact, encoded during the perception of accented speech. This was evidenced in the speech productions of subjects who shadowed accented models. Talker-specific alignment is consistent with prior studies of this phenomenon (Goldinger, 1998; Goldinger & Azuma, 2004; Miller et al., 2010; Namy et al., 2002; Nye & Fowler, 2003; Pardo, 2006; Sanchez et al., 2010; Shockley et al., 2004). In addition, at least in the case of single-talker shadowing (Experiment 1a), accent-general information does not seem to mask talker-specific information; while

it might be masked in speech identification tasks (Sidaras et al., 2009). A study involving an alignment methodology where subjects shadow multiple, accented models may reveal whether or not talker-specific information is masked by accent-general information. Although Experiment 2a had subjects shadow multiple, accented models, actual talker alignment to those models was not measured. Assessment of the shadowed stimuli from Experiment 2a using an AXB perceptual matching task might provide evidence of talker alignment. But, it would also be prudent to test talker alignment to multiple, accented talkers specifically using the a general speech alignment methodology (e.g., Goldinger, 1998).

Accent Effects. A response bias during Experiment 1b leaves it unclear as to whether subjects align to a specific accent. The response bias analysis revealed that raters tended to match shadowed tokens to the Spanish-accented models, regardless of which accent was shadowed. When combined with the results of Experiment 1c, it seems most likely that raters were making these matches on the basis of Spanish-accented models sounding less accented than Chinese-accented models.

As mentioned, this response bias may have arisen because of the demographic make-up of our subjects: It is likely that the subjects in the current experiments had much more experience hearing Spanish accented, than Chinese accented speech. It is also likely that the Spanish-accent experience of the current subjects was much greater than subjects tested by Sidaras et al. (2009) based on demographic differences in the subject pools. Still, previous research has shown shifts in an individual's speech in the direction of an ambient language community (e.g., Flege, 1987; Major,

1992; Sancier & Fowler, 1997) and dialect (Delvaux & Soquet, 2007; Kraljic et al., 2008) suggesting that accent alignment to some extent possible. To reduce the potential confounds of familiarity and demographics, a follow-up study is necessary that involves a language background that is less familiar to the subject pool. Less familiar accented speech should also reduce the intelligibility of this speech, making it unnecessary to add noise at test as was necessary in Experiment 2a.

The second series of experiments implemented a shadowing task in the course of a high variability training and test paradigm (e.g., Sidaras et al., 2009) to investigate whether or not the immediate and productive nature of shadowing will increase the encoding of talker information during the perception of accented speech. Experiment 2a was not able to uncover either effects of Spanish-accented training or talker familiarity on later perception of Spanish-accented speech. The former finding is in contrast to Sidaras et al. (2009) who found that training on accented speech later improves perception of accented speech. In this sense, the findings of the experiment were uninformative about talker familiarity effects due to the lack of training effects.

A follow-up experiment (2b) was performed to test whether shadowing would improve encoding of talker-specific information enough to show talker familiarity effects during the perception of native (unaccented) speech. Experiment 2b revealed that listeners trained to shadow or transcribe native speech were better able to perceive this speech than listeners in a text reading control condition. This suggests that, unlike Experiment 2a, training did have an influence on later perception. However, once again, there were no talker familiarity effects for either the shadowing

or the transcription groups. In other words, training on familiar talkers did not improve later perception of those same vs. different talkers as it has in previous research (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993).

Frequency Effects. Another consideration is the effect of word frequency on alignment to and perception of accented speech. Recall that word frequency is considered a talker-independent property that has been shown to influence speech alignment (e.g., Goldinger, 1998) and ratings of accentedness (Levi et al., 2007). The lexical frequency of a word also influences reaction time of responding to that word (Dahan, Magnuson, & Tannehaus, 2001), where high frequency (HF) words are responded to more quickly than low frequency (LF) words. A similar trend occurs for the word recognition accuracy of accented speech, where HF, Spanish-accented words are recognized more accurately than LF, Spanish-accented words (Imai, Walley, & Flege, 2004).

In the present dissertation, there were significant word frequency effects across all experiments (though they will not be considered for Experiment 1b). For example, shadowers in Experiment 1a were more likely to align to the talker-specific characteristics of an accented model speaking HF and MLF words than MHF words, however, LF words did not induce significant alignment. Though inconsistent with alignment literature (Goldinger, 1998), this could be due to HF accented words being easier to perceive and less accented (Imai et al., 2007; Levi et al., 2007), allowing talker-specific information to be encoded more efficiently for these types of words. In

Experiment 1c, ratings of accentedness were influenced by word frequency with HF words being considered the least accented and LF words being considered the most accented. These results are consistent with those found by Levi et al. (2007). Word frequency also had an effect on the perceptual identification of speech in Experiment 2a and 2b. For Experiment 2a, there was a significant difference in accuracy at all levels of word frequency (i.e., HF>MHF>MLF>LF) when listeners were identifying Spanish-accented speech. Finally, Experiment 2b revealed that LF words were recognized significantly less well than all other levels of word frequency for native, English words.

Taken together, these findings for accented speech suggest that the influences of word frequency on perceived accentedness might have play an important role in talker alignment and perceptual identification. Further inquiries into the interaction between word frequency and level of accentedness are necessary.

It is important to mention that the selection of word frequency stimuli were based on the word corpus Kučera and Francis (1967) used in previous related research (e.g., Goldinger, 1998; Miller et al., 2010; Shockley et al., 2004). The word frequencies in this corpus may be dated and not reflective of present word usage. In fact, Burgess and Livesay (1998) suggest that word frequency predictions of listeners' reaction time are more accurate when using an up-to-date corpus (i.e., HAL; CELEX).

8.1 Theoretical Implications

Previous research regarding the influence of talker-specific information on the perception of accented speech was unclear as to whether this information is removed during a process of normalization or encoded in memory and somehow masked (Sidaras et al., 2009). The present findings of alignment to accented talkers (Experiment 1a) provide some support for the encoding of talker-specific information in the context of accented speech perception.

This evidence is consistent with episodic accounts of speech perception that proffer the encoding of highly-detailed traces of speech events in memory (Goldinger, 1996, 1998; Johnson, 2005, 2008). It is also supportive of a link between speech perception and production that may occur due to a common currency being shared between these functions (Liberman & Mattingly, 1985; Fowler, 1986; 2003).

It is not clear why talker-specific effects can be observed during alignment, but not in speech identification responses. On the one hand, it could be that the alignment method is better at revealing talker-specific influences than perceptual identification tasks. In other words, talker information may have been encoded during the experiments of Sidaras, et al (2009), but the speech identification measure was not as sensitive as alignment at revealing talker influences. On the other hand, alignment might induce greater amounts of encoding of talker information than transcription tasks. If alignment induced more encoding, then the listeners in Experiment 2a, trained by shadowing accented talkers should have improved in later perceptual identification for these same talkers at test when compared to listeners who were

trained through transcription. The results of Experiment 2a and 2b did not indicate that shadowing improved later perception of accented or native speech above a transcription task.

However, for Experiment 2a, these results cannot be interpreted because there was no evidence of accent learning (i.e., improvement in perceptual identification accuracy when trained on the accent). While, for Experiment 2b, the stimulus property differences (i.e., noise at test, training in the clear) may have reduced these effects. Regardless, this suggests that shadowing may not improve encoding, though other explanations (e.g., problems with retrieval; differing stimulus properties from training to test) could account for this lack of findings.

8.1.1 Dialect Change

The finding of alignment to accented talkers (Experiment 1a) may have implications for our understanding of the early stages of dialect formation (e.g., Babel, 2009; Munro, Derwing, & Flege, 1999; Wolfram et al., 2004) and language change (Kerswill, 2000). Recall that shifts in articulatory gestures (e.g., gestural drift) occur as a result of contact with an ambient language community (e.g., Flege, 1987; Major, 1992, Sancier & Fowler, 1997) or dialect (Delvaux & Soquet, 2007). Dialect change might simply be a form of long-term accommodation (i.e., alignment) to new speech patterns (e.g., Babel, 2009; Trudgill, 1986). Trudgill (1986) suggests that new dialects form when a talker comes into and remains in contact with individuals who have differing speaking habits. Changes in speech occur as a result of this contact and

continue to later generations. For example, Wolfram et al., (2004) showed that native, Spanish talkers who moved to the U.S. adjusted to different aspects of a "Southern" dialect, which seemed to begin with acquiring lexical items common to that dialect (e.g., fixin' vs. fixing)(see also, Chambers, 1992). Both short- and long-term alignment might occur to accented talkers and be in part responsible for dialect change.

8.1.2 The Nature of Accent Information

Although the present dissertation uncovers the encoding of talker-specific information during the perception of accented speech, it is not able to clarify the nature of talker-specific and accent-general information. In the current dissertation and prior research (Sidaras et al., 2009), accent-general information has been investigated as an indexical characteristic distinct from talker-specific information. In fact, the very definition of accent-general information is that it is systematic variation shared by native talkers of the same language (Sidaras et al., 2009). However, accent information could be part of the idiolect used to identify a given talker. This seems reasonable from an episodic account, where both sources of information could be tied to lexical items (e.g., Goldinger, 1998); or from a gestural perspective, where talker and accent information are potentially evidenced through the same articulatory gestures (e.g., Fowler, 1986; 2003). In either case, accent-general information may simply be an additional layer of voice information. The present results were unable to

establish whether or not this was the case. It would be useful for future research to be performed in order to untangle these sources of information.

8.2 Future Directions

There are many interesting ideas that can be guided by the knowledge that some talker-specific information is available during the perception of accented speech. For one, it seems critical to clearly establish whether or not talker familiarity can influence accented speech *perception*. While evidence was gained that talker information is encoded and can influence immediate speech *production*, it is unclear from the present research whether actual perception is effected. Future studies that are carefully designed to assess how talker familiarity bears on the later perceptual identification of accented speech are crucial. For example, Sidaras et al. (2009) and Experiment 2b had listeners rate the accentedness of a model as a part of training. Thus, listeners were unintentionally directed to pay attention to accent information instead of talker information. Studies in which listeners are instead trained to identify the accented talkers by name, might increase talker familiarity effects by bringing attention to the talker (e.g., Nygaard & Pisoni, 1998; Nygaard et al., 1994). In this way, listeners in a revision of Sidaras et al. (2009) may not be focused on the accent, as much as the talker, and thus improve in later identification of speech from the same talkers.

But perhaps the lack of talker familiarity effects in the shadowing condition of Experiment 2a and 2b is due to limited encoding of talker-specific information in the

framework of the high variability training and test paradigm (Sidaras et al., 2009). It is possible to estimate the level of encoding of talker-specific information in these experiments by measuring how much listeners in the shadowing condition have *aligned* to their models. Recall that listeners in the shadowing condition were recorded producing shadowed responses after accented models. The degree of talker alignment can be measured by having naïve raters compare the listeners' shadowed responses produced during training and the accented models' utterances. Ratings of talker alignment could then be compared to the perceptual identification results. If listeners who are judged as aligning more, also show higher perceptual identification accuracy in the *same* vs. *different* condition, this would suggest that the shadowing task is capable of revealing talker familiarity effects on the *perception* of accented speech.

The lack of difference between *same* and *different* conditions for shadowing subjects in Experiment 2a and 2b could also be due to issues involving *transfer appropriate processing* (e.g., Graf & Ryan, 1990; Morris et al., 1977). As mentioned in the discussion of Experiment Series 2, a revision of these experiments where subjects shadow at both training and test could assess whether task changes from training to test (i.e., shadowing training to transcription test) are masking actual differences for shadowing subjects. Subjects would be trained on multiple-accented (or native English) talkers by either shadowing or transcribing. They would then be asked to perceptually identify speech from the same or a different group of talkers by producing (shadowing) the talkers' speech. If shadowing subjects in the *same*

condition of this follow-up study perform better at test than those in the *different* condition, it suggests that transfer appropriate processing was influencing results during Experiment Series 2. This test might also reveal differences between Transcription and Shadowing groups that were not evident in the previous experiments.

It could also be that presenting novel words from training to test reduces talker familiarity effects for accented speech. Recall that listeners were trained on a set of words, then presented a completely novel set of words at test. From the episodic perspective of Goldinger (1998), talker-specific and lexical information are tied together in memory. This suggests that although novel words are produced by the same talkers at test, talker-specific information would not be as influential as when the *same* words are presented at test. Numerous studies have revealed that talker-specific information aids memory for repeated words (Church & Schacter, 1994; Goldinger, 1996; Palmeri et al., 1993; Schacter & Church, 1992). Future studies using a continuous recognition memory task (CRMT; Craik & Kirsner, 1974; Palmeri et al., 1993) may allow for the effects of talker and accent on memory for *previously* heard words to be assessed.

Though talker alignment was found in Experiment Series 1, accent alignment was not found due to a response bias. In order to establish if alignment to accent occurs, it is necessary to perform additional experiments that control for problems inherent in comparing models with more vs. less of an accent. For example, making sure comparison models in an XAB task have similar levels of accentedness might

reduce this bias. It is also possible that having raters make comparisons between two shadowed tokens (i.e., a token shadowed after a Spanish-accented model, a token shadowed after a Chinese-accented model) and one accented model (e.g., Spanish) would remove the potential for bias.

The present study (Experiment 1a) showed that talker alignment occurs to accented speech. But, how important is it that shadowing occurs immediately after speech and would talker alignment occur if shadowing was not productive?

The importance of immediacy in shadowing can be addressed quite easily by a delayed-shadowing experiment, where subjects are presented accented speech stimuli and wait 3 – 4 s before making a shadowed response (Goldinger, 1998). This delay between presentation and production could decrease the encoding of talker-specific information, which might reduce talker alignment. An alternative explanation of these potential results would be that accent-heavy traces might flood working memory making talker alignment less likely, and accent alignment more likely, if this phenomenon occurs. Additional studies involving delayed shadowing of accented talkers may reveal differential influences of talker and accent information on speech alignment.

The productive nature of shadowing is more difficult to manipulate, but speech alignment has been shown in non-productive tasks (e.g., Goldinger & Azuma, 2004). For example, Goldinger & Azuma (2004) had subjects read a list of text words (pre-task), complete a listening task where they were presented repetitions of words produced by multiple talkers, then read the same list of text words again (post-task).

Even though, subjects never "shadowed" the model's speech, post-task utterances were rated as more similar to the models' words than were pre-task utterances. This finding suggests that alignment can be induced by a listening task. This experimental design could be used in order to test the relative importance of production in inducing talker alignment to accented speech.

Finally, fundamental to a full understanding of talker-specific and accent-general information are studies that look at multimodal perception of accented speech. Research on native (unaccented) speech perception shows that talker-specific information is available not only through auditory, but through visual speech (e.g., Lachs & Pisoni, 2004; Rosenblum, Niehus, & Smith, 2007; Rosenblum et al., 2002). In fact, visual speech information for talker characteristics can influence speech alignment (e.g., Gentilucci & Bernardis, 2007; Miller et al., 2010; Sanchez et al., 2010). This may be the result of the amodal nature of speech information (i.e., information available across modalities; Fowler, 2004). Making visual speech information available during the perception of accented speech or shadowing will provide an additional source of talker information, which may subsequently influence talker effects and increase talker, and potentially accent, alignment.

8.3 Practical Implications

The present research also has several practical implications that are spurred by our growing global economy. For example, a clearer understanding of the influences of talker-specific and accent-general information can aid in the development of

automatic speech recognition systems that take these sources of information into account (e.g., Doddington, 1985; Ikeno & Hansen, 2007). An awareness of talker and accent familiarity is also relevant in identifying criminal suspects (i.e., "earwitness" identification; Kerstholt et al., 2006; Thompson, 1987). This information may also help to generate programs aimed at facilitating individual's comprehension of and responses to accented speech, thus allowing non-native talkers to be better understood.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Aldine Pub. Co., Chicago.

Adank, P., Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging*, 25(3), 736-740.

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science, 21*(12), 1903-1909.

American Speech-Language-Hearing Association (n.d.). Phonemic inventories across languages. ASHA's Office of Multicultural Affairs. Retrieved from: http://www.asha.org/practice/multicultural/Phono.htm#resources

Anderson, J. R. M. (2006). An approximation to d' for n-alternative forced choice. Project Report. Psychology Department, Hobart, Tasmania. Retrieved from: http://eprints.utas.edu.au/475/1/nAFCrev207.pdf

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Docherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task corpus. *Language and Speech, 34*, 351–366.

Bäckman, L., Nilsson, L., & Chalom, D. (1986). New evidence on the nature of the encoding of action events. *Memory & Cognition, 14*(4), 339-346.

Babel, M. (2009). *Phonetic and social selectivity in speech accommodation*. Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Bent, T., & Bradlow, A. R., (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114*(3), 1600-1610.

Bond, Z. S., & Small, L. H. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Perception & Psychophysics, 34*(5), 470-474.

Bowers, J. S., Mattys, S. L., & Gage, S. H. (2009). Preserved implicit knowledge of a forgotten childhood language. *Psychological Science, 20*(9), 1064-1069.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707–729.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12-21.

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting RT in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers, 30,* 272–277.

Chambers, J. K.(1992). Dialect acquisition. *Language, 68*, 673-705.

Chang, Y, & Fu, Q. (2006). Effects of talker variability on vowel recognition in cochlear implants. *Journal of Speech, Language, and Hearing Research, 49,* 1331-1341.

Chartrand, T. L., Bargh, J. A., (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality & Social Psychology*, *76*, 893-910.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 521-533.

Clarke, C. M. (2000). *Perceptual adjustments to foreign-accented English*. In Indiana University's Research on Spoken Language Processing Progress Report, No. 24.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116*, 3647-3658.

Cohen, R. L. (1983). The effect of encoding variables on the free recall of words and action events. *Memory & Cognition,* 11, 575-582.

Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology, 26*, 274-284.

Creelman, C. D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America, 29,* 655.

Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. New York: Cambridge University Press.

Cunningham-Andersson, U., & Engstrand, O. (1989). Perceived strength and identity of foreign accent in Swedish. *Phonetica, 46*, 138-154.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*, 317-367.

Delvaux, V. & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*, 145-173.

Doddington, G. R. (1985). Speaker recognition: Identifying people by their voices. Proceedings of the IEEE, 11, 1651-1664.

Fadiga, L., Fogassi, L., Povesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*, 2608–2611.

Finegan, E. (2004). *Language: Its structure and use* (4th ed.). Boston: Thomas-Wadsworth.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J . E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America, 84*, 70-79.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on perceived degree of foreign accent. *Journal of the Acoustical Society of America, 91*(1), 370-389.

Flege, J., Bohn, O-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3-28.

Fowler, C. A. (2004). Speech as a supermodal or amodal phenomenon. In Calvert, G. A., Spence, C., & Stein, B. E. (Eds.) *The Handbook of Multisensory Processing* (pp.189-201). Cambridge, MA: MIT Press.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory & Language, 49*, 396-413.

Gentilucci, M. & Bernardis, P. (2007) Imitation during phoneme production. *Neropsychologia, 45*, 608-615.

Gerstman, H. (1968). Classification of self normalized vowels. IEEE Transactions on Audio and Electroacoustics, 16, 630-640.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton-Mifflin.

Giles, H., Coupland, J. & Coupland, N. (1991). Accommodation theory: Communication, context, and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1-68). Cambridge: Cambridge University Press.

Giles, H., & Ogay, T. (2007). Communication accommodation theory. In B. B. Whalen & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293-310). Mahwah, NJ: Lawrence Erlbaum.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166-1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Goldinger, S. D. & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin, 11*(4), 716-722.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*(1), 152-162.

Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 978-992.

Halle, M. (1985). *Speculation about the representation of words in memory*. In V. Fromkin (Ed.), Phonetic linguistics (pp. 101-114). New York: Academic Press.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411-428.

Ikeno, A., & Hansen, J. H. L. (2007). The effect of listener accent background on accent perception and comprehension. *EURASIP Journal on Audio, Speech, and Music Processing, 2007*, 1-8.

Imai, S., Walley, A. C., & Flege, J. E. (2004). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America, 117*(2), 896-907.

Irwin, A. Thomas, S. (2006) *Identification of language and accent through visual speech,* Poster presented at the 2006 Conference on Speech Prosody, Dresden, 2 - 5th May.

Jacoby, L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 1-47). New York: Academic Press.

Johnson, K. (1997) Speech perception without speaker normalization: an exemplar model. In K. Johnson and J.W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145-166). San Diego: Academic Press.

Johnson, K. (2005). Speaker normalization in speech perception. In D. Pisoni and R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp 363–389). Oxford: Blackwell.

Johnson, K. (2008). *Quantitative Methods in Linguistics.* Oxford: Blackwell.

Joos, M. A. (1948). Acoustic phonetics. *Language, 24*, Supplement 2, 1-136.

Kappes, J., Baumgaertner, A., Peschke, C., & Ziegler, W. (2009). Unintended imitation in nonword repetition. *Brain & Language, 111*, 140-151.

Kerstholt, J. H., Jansen, N. J. M., Van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology, 20*, 187–197.

Kerswill, P. E. (2002). Koineization and accommodation. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*, (pp. 669–702). Oxford: Blackwell.

Kim, M., Horton, W. S., & Bradlow, A. R.(2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology, 2*, 125-156.

Kolers, P. A. (1979). Remembering operations. *Memory & Cognition, 1*(3), 347-355.

Kolers, P.A., & Roediger, H.L. III. (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior, 23*, 425-449.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107,* 54-81.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American-English*. Providence, RI: Brown University Press.

Lachs, L., McMichael, K. & Pisoni, D.B. (2003). Speech perception and implicit memory: Evidence for detailed episodic encoding of phonetic events. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory.* Oxford: Oxford University Press. Pp. 215-235.

Lachs, L., & Pisoni, D. B., (2004). Crossmodal Source Identification in Speech Perception. *Ecological Psychology*, *16*(3), 159-187.

Ladefoged, P. (1980). What are linguistic sounds made of. *Language, 56,* 485-502.

Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America, 35*, 451-453.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America, 121*(4), 2327-2338.

Liberman, A. M. (1983). What a perception-production link does for language. *The Behavioral and Brain Sciences, 2*, 216.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 224–234

Magen, H. S. (1998). The perception of foreign-accented speech. *Journal of Phonetics, 26*, 381-400.

Major, R. C. (1987). Phonological similarity, markedness, and rate of L2 acquisition. *Studies Second Language Acquisition, 9*, 63-82

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal, 76,* 190-208.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 676-681.

Meltzoff, A. N. & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting, 6*, 179-192.

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, *72*, 1614-1625.

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (submitted). *Is speech alignment to talkers or task?* Manuscript submitted for publication.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.

Mullennix, J. M., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365-378.

Munro, M.J., Derwing, T.M., & Flege, J.E. (1999). Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics, 27*, 385–403.

Namy, L. L., Nygaard, L. C. & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*(4), 422-432.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology, 32*, 790–804.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85,* 2088-2113.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*, 132-142.

Nusbaum, H. C., and Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson and J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 109 – 132). San Diego: Academic Press.

Nye, P. W., & Fowler, C. A. (2003). Shadowing latency and imitation: The effect of familiarity with the phonetic patterning of English. *Journal of Phonetics, 31*(1), 63-79.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific perceptual learning in spoken word recognition. *Perception & Psychophysics, 60,* 355-376.

Nygaard, L. C., Sommers, M., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Oztop, E., Kawato, M., & Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. Neural Networks, 19, 254–271.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 309-328.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, *119*, 2382-2393.

Pickering, M. J., Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral & Brain Sciences*, *27*(2), 169-226.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology V11* (pp. 101-140). Berlin: Mouton de Gruyter.

Porter, R. J., Jr., & Castellanos, F. X. (1980). Speech production measures of speech perception: Rapid shadowing of VCV syllables. *Journal of the Acoustical Society of America*, *67*, 1349-1356.

Porter, R. J., Jr., & Lubker, J. F. (1980). Rapid reproduction of vowel–vowel sequences: Evidence for a fast and direct acoustic–motoric linkage in speech. *Journal of Speech & Hearing Research*, *23*, 593-602.

Roberts, L. G. (1965). Machine perception of three-dimensional solids. In J. T. Tippett (Ed.), *Optical and electro-optical information processing* (pp. 159-197). Cambridge, MA: MIT Press.

Rosenblum, L. D, Miller, R., & Sanchez, K. (2007). Lipread me now, hear me better later: Crossmodal transfer of talker familiarity effects. *Psychological Science, 18*, 392-396.

Rosenblum, L. D., Niehus, R. P., & Smith, N. M. (2007). Look who's talking: Recognizing friends from visible articulation. *Perception, 36*, 157-159.

Rosenblum, L. D., Yakel, D. A., Baseer, N., Panchal, A., Nodarse, B. C., & Niehus, R. P. (2002). Visual speech information for face recognition. *Perception & Psychophysics, 64*, 220-229.

Sanchez, K., Miller, R. M., & Rosenblum, L. D. (2010). Visual influences on alignment to voice onset time. *Journal of Speech, Language, and Hearing Research, 53*, 262-272.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421-436.

Scales, J., Wennerstrom, A., Richard, D., & Wu, S. (2006). Language learners' perceptions of accent. *TESOL Quarterly*, 40 (4), 715-738.

Schmid, P. M., & Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Language, and Hearing Research, 42*, 56-64.

Schacter, D. L., & Church, B. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915-930.

Shankweiler, D. P., Strange, W., & Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, and Knowing: Toward an Ecological Psychology* (pp. 315-345). Hillsdale, NJ: Erlbaum.

Shockley, K., Sabadini, L., & Fowler, C. A., (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422-429.

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 326-332.

Sidaras, S., Alexander, J. J., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America, 125*(5), 3306-3316.

Summerfield, Q., & Haggard, M. (1975). Vocal tract normalization as demonstrated by reaction times. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 115-141). London: Academic Press.

Tarone, E. E. (1987). The phonology of interlanguage. In G. Ioup, & S.H. Weinberger (Eds.), *Interlanguage Phonology: The Acquisition of a Second Language Sound System* (pp. 70-85). Cambridge, MA: Newbury House Publishers.

Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology, 1,* 121-131.

Trudgill, P. (1986). *Dialects in Contact*. Oxford, U.K.: Blackwell.

U.S. Census Bureau. (2007-09). Dekalb County, Georgia – ACS demographics and housing estimates. Retrieved from http://factfinder.census.gov/.

U.S. Census Bureau. (2007-09). Riverside County, California – ACS demographics and housing estimates. Retrieved from http://factfinder.census.gov/.

van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication, 35,* 103–113.

Warren, J. E., Wise, R. J. S., & Warren, J. D. (2005). Sound do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in neuroscience, 28*(12), 636-643.

Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.

Wolfram, W., Carter, P., & Moriello, B (2004). Emerging Hispanic English: New dialect formation in the American South. *Journal of Sociolinguistics,* 8(3), 339-358.

Appendix A

Stimulus Words (and Frequencies) for Experiment Series 1

Bisyllabic	Frequency	Monosyllabic	Frequency	Bisyllabic	Frequency	Monosyllabic	Frequency
High Frequency words (> 300)				Medium Low Frequency words (50-100)			
before	1016	case	362	active	88	bank	86
between	730	door	312	balance	90	chair	66
city	393	great	665	captain	85	crowd	53
country	324	group	390	careful	62	dust	70
matter	308	house	591	coffee	78	fresh	82
never	698	last	676	cousin	51	grass	53
number	472	light	333	dozen	52	knife	76
order	376	next	394	favor	78	lake	54
program	394	part	500	forget	54	moon	60
public	438	place	569	garden	60	phone	54
rather	373	point	395	handle	53	prove	53
social	380	school	492	listen	51	speed	83
system	416	side	380	master	72	throat	51
water	442	white	365	symbol	54	tree	59
		work	760	title	77		
		young	385	vision	56		
Medium High Frequency words (150-250)				Low Frequency words (<5)			
basis	184	black	203	cavern	1	chunk	2
beyond	175	book	193	deport	1	dire	1
common	223	bring	158	detest	1	fade	2
figure	209	care	162	flannel	4	germ	3
final	156	fire	187	garter	2	hoop	3
market	155	floor	158	hectic	3	kelp	2
music	216	girl	220	jelly	3	knack	4
nature	191	ground	186	maltreat	1	leash	3
party	216	hard	202	mingle	2	malt	1
police	155	late	179	nectar	3	raft	4
recent	179	lost	171	parcel	1	stale	4
river	165	plan	205	portal	3	vest	4
single	172	rest	163	rustic	3	weed	1
table	198	sound	204	stony	5		
value	200	wall	160	tricky	1		
				typhoon	1		
				wedlock	2		

Note. Adapted from "Echoes of Echoes? An Episodic Theory of Lexical Access" by S. D. Goldinger, 1998, *Psychological Review, 105*(2), p. 278. Copyright 1998 by the American Psychological Association, Inc. Word frequencies based on Kučera and Francis (1967).

Appendix B

Stimulus Words (and Frequencies) for Experiment Series 2

Bisyllabic	Frequency	Monosyllabic	Frequency	Bisyllabic	Frequency	Monosyllabic	Frequency
High Frequency words (> 300)				Medium Low Frequency words (50-100)			
become	361	back	967	balance	90	band	53
before	1016	case	362	captain	85	bank	86
better	414	door	312	career	67	chair	66
between	730	great	665	careful	62	crowd	53
later	397	group	390	coffee	78	dust	70
never	698	house	591	cousin	51	fresh	82
number	472	last	676	dozen	52	knife	76
people	847	light	333	favor	78	lake	54
power	342	part	500	garden	60	phone	54
rather	373	place	569	handle	53	safe	58
second	373	point	395	listen	51	speed	83
social	380	school	492	master	72	throat	51
		work	760	novel	59	tree	59
		young	385				
Medium High Frequency words (150-250)				Low Frequency words (<5)			
common	223	black	203	arid	2	dire	1
father	183	book	193	bicep	1	fade	2
figure	209	bring	158	conquer	4	germ	3
final	156	class	207	druid	1	hoop	3
nature	191	cold	171	eagle	5	kelp	2
party	216	floor	158	lacquer	2	leash	3
picture	161	hard	202	lasso	2	mule	4
police	155	north	206	leftist	1	stale	4
recent	179	plan	205	melon	1	vest	4
report	174	sound	204	navel	2	vine	4
river	165	stage	174	negate	2	weed	1
spirit	182	wall	160	rascal	1	wilt	3
table	198			rattle	5		
value	200			typhoon	1		

Note. Adapted from "Echoes of Echoes? An Episodic Theory of Lexical Access" by S. D. Goldinger, 1998, *Psychological Review, 105*(2), p. 278. Copyright 1998 by the American Psychological Association, Inc. Word frequencies based on Kučera and Francis (1967).