5-2018

Assessing the relationship between talker normalization and spectral contrast effects in speech perception.

Ashley Atri Assgari
University of Louisville

ASSESSING THE RELATIONSHIP BETWEEN TALKER NORMALIZATION AND SPECTRAL CONTRAST EFFECTS IN SPEECH PERCEPTION

By

Ashley Atri Assgari
B.A., James Madison University, 2011
M.A., James Madison University, 2014
M.S., University of Louisville, 2016

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Experimental Psychology

Department of Psychological and Brain Sciences
University of Louisville
Louisville, Kentucky

May 2018

ASSESSING THE RELATIONSHIP BETWEEN TALKER NORMALIZATION AND SPECTRAL CONTRAST EFFECTS IN SPEECH PERCEPTION

By

Ashley Atri Assgari
B.A., James Madison University, 2011
M.A., James Madison University, 2014
M.S., University of Louisville, 2016

A Dissertation Approved on

April 9, 2018

by the following Dissertation Committee:

_____

Dissertation Director
Christian Stilp

_____

Cara Cashon

_____

Paul DeMarco

_____

Sharon Miller

_____

Pavel Zahorik

DEDICATION

This dissertation is dedicated to my niece, Sophie Assgari.

Let this document be a reminder that you can achieve even the highest goals with a little dedication and perserverance.

ACKNOWLEDGEMENTS

First, I would like to express my appreciation to my committee, Dr. Cara Cashon, Dr. Paul DeMarco, Dr. Sharon Miller, and Dr. Pavel Zahorik, for dedicating their time to reviewing this dissertation and for providing excellent feedback.

Second, I would like to sincerely thank my research advisor, Dr. Christian Stilp, for his continued support and guidance over the last four years. Your willingness to listen to and assist me in achieving my goals in all of the ventures I pursued at U of L are what made me the researcher and the scholar that I am today. I will never forget the enthusiasm that was instilled in me when I finally had someone that was as excited as I was about speech acoustics and data analyses.

Third, I would like to thank my family for their constant support. I would like to thank my mother, Lori Assgari, for instilling in me a sense of dedication and perserverance; my father, Abdie Assgari, for teaching me that it is okay to voice my opinion and fostering my curiosity at a young age; my step-mom, Farzaneh Assgari, for her unwavering support and encouragement. Most of all, I would like to thank my sister, Tila Assgari, whose minor comment about her emphasis on learning rather than grades changed the way that I approached education and learning for the better. You have been, and continue to be, an inspiration to me.

Finally, I would like to thank my boyfriend, Robert Fratino, who was willing to uproot his life to support me in my adventure. I know without your support, I would not be writing this today.

ABSTRACT

ASSESSING THE RELATIONSHIP BETWEEN TALKER NORMALIZATION AND SPECTRAL CONTRAST EFFECTS IN SPEECH PERCEPTION

Ashley A. Assgari

April 9, 2018

Speech perception is influenced by context. This influence can help to alleviate issues that arise from the extreme acoustic variability of speech. Two examples of contextual influences are talker normalization and spectral contrast effects (SCEs). Talker normalization occurs when listeners hear different talkers causing speech perception to be slower and less accurate. SCEs occur when spectral characteristics change from context sentences to target vowels and speech perception is biased by that change. It has been demonstrated that SCEs are restrained when contexts are spoken by different talkers (Assgari & Stilp, 2015). However, what about hearing different talkers restrains these effects was not entirely clear. In addition, while these are both considered contextual influences on speech perception, they have never been formally related to each other. The series of studies reported here served two purposes. First, these studies sought to establish why hearing different talkers restrained SCEs. Results indicate that variability in pitch (as measured by fundamental frequency), a primary acoustic cue to talker changes, restricts the influence of spectral changes on speech perception. Second, these studies attempted to relate talker normalization and SCEs by measuring them concurrently. Talker normalization (as measured by response times) and SCEs were evident in the same task suggesting that they act on speech perception at the same time. Further, these measures of

talker normalization were shown to be influenced by f0 variability suggesting that SCEs and talker normalization are both related to f0 variability. However, no relationship between individual's SCEs and response times was found. Possible reasons why f0 variability may restrain context effects are discussed.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

When perceiving speech, there are many cues that can lead to one percept and one cue that can contribute to many different percepts. Despite this fact, speech perception is remarkably accurate. Speech perception research seeks to explain this incongruity by determining what cues listeners rely upon to accurately perceive speech. Through these lines of research, many cues that influence speech perception have been found. The proposed studies will seek to understand how two of these cues may be related. The first cue is talker acoustics. In speech, talker acoustics are highly variable. When hearing different people speak, the acoustics of the speech change dramatically. The other cue is the frequency composition of recent sounds. Each of these cues has its own influence on speech perception. While these cues have never been directly linked, there is evidence that they may be related.

This dissertation will start by explaining the different types of cues that help listeners cope with the variability of speech. Next, how hearing different people speak influences speech perception is presented. Third, how the frequency composition of surrounding sounds influences perception is discussed. Finally, the possible relationship between these two influences on speech perception is assessed.

How Do We Deal With Variability?

The perceptual world is full of variability. In every modality, we encounter objects that vary dramatically yet all belong to a single category. For example, consider an apple. Apples come in many different sizes, colors, and shapes. Despite this variability, we know an apple when we see one. Say we encountered two granny smith apples. Granny smith apples are characteristically green. Any two apples will surely vary in their exact shade of green, but if asked what kind of apples ours are, we will have no difficulty telling someone that they are both granny smith apples. Now, consider a gala apple and a honey crisp apple. Even though both of these apples are a similar shade of pink, we can still differentiate the two types of apples. Two objects that are ostensibly different on many key characteristics can still be labeled as the same category reliably. Further, two objects that are similar on key characteristics can accurately been differentiated. This problem often plagues perception researchers. If there is no direct relationship between a key characteristic and the identification of an object, how can it be argued that a characteristic is vital for accurate perception?

Acoustic Variability in Speech

Speech is no exception to this problem. Speech production is highly variable, so just like the granny smith apples, the acoustic characteristics of two instances of the same speech sound will vary substantially from each other. If two talkers produce the same sound, their productions will have very different acoustics. These differences arise from the physiology of the talker, which will be discussed later. Even within the same talker, when the same sound is produced, the acoustics will vary. This is due to differences in the articulation of speech sounds at any given time. When a single talker produces two of the same speech sounds, how they articulate the sounds is unlikely to be exactly the same

each time. Furthermore, like the gala and the honey crisp apples, two sounds that have similar acoustics can be perceived as two different speech sounds, depending on the context in which they are perceived.

Despite this variability, listeners are extremely accurate at identifying speech. This creates an interesting problem for speech researchers: despite substantial variability both within and between phoneme categories, speech perception is highly accurate. This problem is expressed through the concept of the lack of invariance (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The lack of invariance states that there is no direct correspondence between acoustic cues and speech perception. Put another way, no acoustic cue is both necessary and sufficient for recognizing a given speech sound. Researchers have attempted to address this issue by understanding what other acoustic information can influence speech perception.

Intrinsic vs Extrinsic Cues to Speech Perception

In the literature, there is a distinction made between intrinsic and extrinsic cues to speech perception. An intrinsic cue is an aspect of the stimulus that is self-contained. In our apple example, an intrinsic cue would be the apple's shape. While the exact shape of any given apple may vary from an ideal, they have a generic shape that helps you determine its identity. Extrinsic cues are other cues that help identify a stimulus but are not a part of the stimulus itself. As such, extrinsic cues are often characteristics of the context in which the stimulus is perceived. For our apple, if it was placed in a fruit basket, the surrounding fruit cues the perceiver that our stimulus is likely also a fruit.

In speech, many acoustic cues are considered intrinsic cues. For example, properties related to amplitude (e.g., Fairbanks, House, & Stevens, 1950), duration (e.g.,

Peterson & Lehiste, 1960), and frequency composition of that sound (e.g., Peterson & Barney, 1952). Speech perception is also influenced by extrinsic cues (e.g., Ladefoged & Broadbent, 1957; Ainsworth, 1975; Nearey, 1989; Stilp, Anderson & Winn, 2015). Extrinsic cues in speech are characteristics of the surrounding sounds. Both sounds that come before and after a stimulus can influence perception of a given (target) sound. In general, the literature shows that sounds that come before the target sound are much more influential than sounds that come after. There are many different characteristics of the surrounding sounds that influence the perception of the target sound. In this paper, the low-level acoustic characteristics of the preceding sounds and their extrinsic influence on the perception of the target are discussed.

The following two sections will discuss two extrinsic cues to speech perception in more detail: (1) who is producing the speech and (2) the frequency content of the surrounding (i.e., earlier) sounds. First, the consequences of hearing different talkers are discussed. As previously mentioned, when hearing different talkers, their speech acoustics vary dramatically**,** and this can impair speech perception. Next, the frequency content of earlier sounds has a considerable influence on identification of later sounds; this influence will be discussed. Finally, a discussion of the relationship between these talker acoustics and the frequency composition of earlier sounds is presented. Gaps in the understanding of this relationship are used to motivate the proposed studies.

Talker Variability

One of the sources of acoustic variability in speech is the acoustic changes that arise from hearing different talkers. These changes can impair and/or slow the listener's speech perception. This is of particular interest because, throughout any given day,

listeners will hear speech from many talkers. Furthermore, it is common to hear speech from different talkers in close succession. Since the ultimate goal for researchers is to understand the perception of speech in natural environments, understanding the perceptual consequences of hearing different talkers is essential.

The Acoustic Underpinnings of Talker Normalization

Talker information does not influence speech only at the time of perceiving a particular speech sound, it can also give the listener some sense of what to expect next. When hearing the same talker, the acoustics of the surrounding sounds remain fairly consistent and speech perception is uninterrupted. However, when hearing different talkers, the surrounding sounds differ dramatically and the listener must perceptually adjust for those differences. A large body of research has investigated the consequences of hearing different talkers when perceiving speech. The general finding of these studies is that perception is slower and less accurate when hearing different talkers relative to hearing the same talker. This effect is referred to as talker normalization (e.g., Creelman, 1957; Fourcin, 1968;Assmann, Nearey, & Hogan, 1982; Geiselman & Bellezza, 1976; Mullenix, Pisoni, & Martin, 1989; Mullenix & Pisoni, 1990; Logan & Pisoni, 1987). This general pattern has held up in decades of research and for a variety of speech related tasks (for a review, see Pisoni, 1993). These tasks include recall of word lists (Goldinger, Pisoni, & Logan, 1991), word identification (Ryalls & Pisoni, 1997), vowel monitoring (Magnuson & Nusbaum, 2007), word monitoring (Magnuson & Nusbaum, 2007), consonant perception (e.g., Rand, 1971), and vowel perception (e.g., Assmann, Nearey, & Hogan, 1982). Since this general finding is well established, it is important to

understand exactly how and when speech perception is slower and less accurate with different talkers.

The acoustics of any sound depend on how it was produced; speech is no exception. Speech is produced through two separable processes (Fant, 1960). The first process is phonation, or the vibration of the vocal folds as air is pushed through them. When the vocal folds vibrate, the rate at which they vibrate corresponds to the pitch of the talker's voice. The perception of pitch is most strongly influenced by the measurement of fundamental frequency (f0). Thus, when researchers refer to pitch changes, they are actually measuring changes in f0. Men generally have longer and thicker vocal folds due to testosterone. Men's vocal folds vibrate relatively slowly producing a low f0. Women, on the other hand, generally have shorter vocal folds that vibrate more quickly, corresponding to a relatively high f0. Individuals are relatively consistent in the f0 of their speech. As a result, hearing different talkers leads to larger changes in f0 than hearing the same talker.

Pitch is a primary if not the primary cue listeners use to distinguish different talkers' voices. Thus, talker normalization effects are likely produced in part by changes in the talkers' f0s. If the vocal folds are held open during speech production (i.e., are not vibrating), the result is whispered speech that has no pitch. This makes different talkers' voices less distinct from each other, which might mitigate the perceptual costs associated with hearing different talkers. Indeed, the effect of hearing different talkers was smaller when the speech was whispered (Fourcin, 1968). This result suggests that f0 is necessary to observe talker normalization effects. Later investigations built upon this finding. Goldinger (1996) asked listeners to report whether words in word pairs were the same or

different. Importantly, on every trial, the words were spoken by different talkers. The author observed a wide range of response latencies for accurate responses. Through non-metric multi-dimensional scaling, the author found that the differences in response latencies were best explained by two dimensions. The first dimension was the gender of the talker: same-talker-gender responses were faster than different-talker-gender responses. The second dimension was the relative f0 within each gender category: response times were faster when talkers were acoustically similar (smaller range of f0s) than when talkers were more acoustically different (wider range of f0s). These studies suggest that f0 variability has a strong influence on talker normalization effects in speech perception.

In follow up experiments, the effects of talker similarity were explicitly manipulated (Goldinger, 1996). Listeners were exposed to a list of words and asked to identify each word. After a delay, listeners were presented with a different list of words and asked whether they had heard the word during exposure (i.e., respond "old") or they had not (i.e., respond "new"). Importantly, listeners heard half of the old words in a different voice than they heard during exposure. A range of hit rates was observed, but performance was predicted by an index of similarity between talkers' voices (higher hit rates for acoustically similar voices). Thus, the degree to which the talkers were considered similar influenced the accuracy of recall (Goldinger, 1996). These results suggest that talker normalization is dependent on the perceived similarity of the talkers. When talkers are similar, listeners have difficulty telling them apart (Magnuson & Nusbaum, 2007). If listeners cannot tell talkers apart, then it is possible the talkers are treated as the same and there is no normalization across different voices. When talkers

are less similar, the costs of hearing different talkers increase and talker normalization processes are required.

As previously stated, speech is produced by two separable processes (Fant, 1960). The first process is phonation. The second process of speech production is articulation, or filtering done by the vocal tract. The vocal tract can be conceptualized as a complex tube with many different compartments. Tubes have resonances and so does the vocal tract. The shape and size of the vocal tract creates different resonances that amplify certain frequencies. These resonances are called formants. These formants that are labeled by number based on ascending frequency (i.e., the first major resonance is labeled F1, the second resonance is F2, etc.). The first three formants are known to have a strong influence on what speech sound in perceived. Anatomical differences produce predictable differences in the acoustics of speech for men and women (Peterson & Barney, 1952; Hillenbrand, Getty, Clark, & Wheeler, 1995). Men, who generally have longer vocal tracts, produce formants that are overall lower in frequency. Women, whose vocal tracts are relatively shorter, produce formants that are overall higher in frequency. Further, each individual has a fairly unique combination of f0 and formants. As such, there are often clear acoustic consequence in speech when hearing different talkers. Researchers attempt to tie these acoustic differences between talkers to the behavioral consequences of listeners hearing different talkers.

Initial explanations of talker normalization suggested that listeners were 'learning' the talker. When repeatedly hearing a single talker, listeners 'learned' the characteristics of the talker's speech (i.e., f0 and formants). This knowledge could then inform the perception of subsequent speech produced by that talker. When hearing

different talkers, that 'learning' is interrupted, giving the listener less information to inform their perception of subsequent speech. One of the characteristics that is supposedly 'learned' is talkers' vowel spaces, or their specific frequency distribution of vowels (Joos, 1948; Ladefoged & Broadbent, 1957). As previously mentioned, the formants of talkers' vowels vary (Peterson & Barney, 1952; Hillenbrand et al., 1995). Listeners were thought to 'learn' each talker's vowel formants through experience with their point vowels (i.e., the vowels at the extremes of the vowel space; Joos, 1948). However, hearing a talker's point vowels before a stimulus does not improve the perception of that stimulus (Verbrugge, Strange, Shankweiler, & Edman, 1976). Thus, the specific content of the surrounding speech sounds had no influence on talker normalization. However, when these same stimuli were blocked by talker rather than presented in random order, there was an improvement in perception of the target sounds (Verbrugge, Strange, Shankweiler & Edman, 1976). By blocking talker, the acoustic variability from trial to trail was decreased. This decrease in variability produced an increase in accuracy. Once again, the variability of the talkers (or lack of) influences the accuracy of speech perception.

Non- Acoustic Influences on Talker Normalization

There have been multiple demonstrations of talker variability influencing talker normalization effects in speech perception. When talkers are different, listeners are slower and less accurate recognizing speech than when talker are similar. Not only does actual acoustic variability influence talker normalization, but listeners' expectation of variability also has an influence. Magnuson & Nusbaum (2007) took two similar-sounding talkers (two male talkers with a 10 Hz difference in average f0) and changed

listeners' expectations of what they were going to hear. In one group, listeners were told that they would only hear one talker. In another group, listeners were told they would hear two talkers. The group that expected to hear two talkers were significantly slower when talkers were mixed relative to when talkers were blocked. The group that expected to hear a single talker showed no increase in response time from the blocked-talker to the mixed-talker condition. These results show that listeners' expectations can also influence talker normalization. However, it is important to mention that a 10 Hz difference in f0 is well within the range of variability that an individual talker can produce. It is unlikely that listeners' expectations could negate a large difference in f0 between talkers, but the limit of this effect has not been tested.

In general, talker normalization effects are reported with talkers that are novel to the listener. But, what if the listener is already familiar with the talker? Nygaard and colleagues have run a series of studies assessing the role of talker familiarity in speech perception (Nygaard, Sommers, & Pisoni, 1994; Nygaard & Pisoni, 1998). When listeners were trained to identify words spoken by different talkers, word identification was better for those talkers (Nygaard, Sommers, & Pisoni, 1994; Nygaard & Pisoni, 1998). In addition, when sentences were used for training, identifying words that were presented in sentences was better for familiar voices (Nygaard & Pisoni, 1998). In other words, when using familiar talkers, there is benefit to speech perception. However, when trained with sentences, no improvement was observed when tested with words presented in isolation. This suggests that the benefit observed with familiar talkers is dependent on the context in which familiarization occurs (Nygaard & Pisoni, 1998). It is clear that there

is a benefit to speech perception by being familiar with a talker. However, whether or not that benefit can overcome the cost of switching talkers has yet to be assessed.

Higher processing demands have been offered as an explanation for talker normalization. As previously mentioned, when hearing different talkers, the listener must adjust to the characteristics of each new talker's voice (e.g., Joos, 1948). This adjustment is thought to result in processing demands higher than when the talker stays the same, and this leads to longer response times than in same-talker tasks. Evidence from fMRI studies support that there are higher processing demands when hearing multiple talkers. While in a scanner, participants were asked to monitor word lists for target words (Wong, Nusbaum, & Small, 2004). The lists were spoken either by a single talker or by multiple talkers. Regardless of the number of talkers, the middle/superior temporal and superior parietal regions of the brain were activated bilaterally. However, lists spoken by multiple talkers led to higher levels of activation in these areas (Wong, Nusbaum, & Small, 2004). Higher levels of activation suggest a harder task, supporting the theory that hearing multiple talkers requires more resources than hearing a single talker. The authors attributed this to the listener having to 'learn' the characteristics of a new talker (Wong, Nusbaum, & Small, 2004).

If talker normalization arises from higher processing demands when hearing multiple talkers, then other ways of increasing processing demands should result in slower and less accurate speech perception. One way to increase processing demands is to increase the difficulty of the task. When listeners hear 'hard' words (i.e., low lexical frequency), they are less accurate than when listening to 'easy' words (i.e., high lexical frequency; Goldinger, Pisoni & Logan, 1991). This result replicates well-known findings

that ‘hard’ words are less likely to be recalled accurately (Hall, 1954). Lower accuracy was also found for words spoken at a fast speaking rate relative to a slower speaking rate (Goldinger, Pisoni & Logan, 1991). In this study, higher processing demands lead to similar decreases in accuracy observed in talker normalization.

However, when listeners were distracted by varying the amplitude of stimuli, leading to higher processing demands, no decreases in speech perception accuracy or increases in reaction times were observed (Magnuson & Nusbaum, 2007). This suggests that higher processing demands alone cannot fully account for talker normalization.

In addition to higher processing demands, how words are initially encoded in memory has been offered as an explanation for talker normalization. As mentioned previously, the recall of words in word lists is a common task in talker normalization research (e.g., Goldinger, Pisoni, & Logan, 1991). The recall of word lists exhibits serial position effects (Murdock, 1962). Words at the beginning and end of a list are recalled more accurately than words in the middle. These effects are known as the primacy and recency effect, respectively (Murdock, 1962). The primacy effect occurs because words at the beginning of the list are rehearsed more than words at the middle or end. The recency effect is attributed to smaller delays between presentation and recall, making recall easier. Talker normalization effects have been shown to be more prevalent at the beginning and end of lists. In other words, the difference between recall accuracy with multiple talkers versus a single talker is greater at the beginning and end of the list (Goldinger, Pisoni, & Logan, 1991). These results suggest that talker normalization effects arise from influences on the encoding of words. When hearing different talkers, encoding and/or rehearsal of the words is disrupted. This suggest that talker

normalization may not only arise due to higher processing demands, but also from poorer rehearsal of the words spoken by multiple talkers.

Conclusion

The preceding section argued that when hearing speech from different talkers, there is a cost to speech perception (i.e., slower and less accurate). How large of a cost is observed is related to how similar the talkers are (especially in regard to f0). When talkers are the same or considered very similar, speech perception is faster and more accurate. However, listeners' expectations of the talkers can also influence how large of a cost is observed. These costs are suggested to be a result of higher processing demands from hearing different talkers and in turn affects the ability to encode words in memory. However, higher processing demands cannot fully account for these costs. It is clear that talker can create a type of context that influences speech perception. The next section discusses another type of context, the frequency composition of surrounding sounds, and its influence on speech perception.

Spectral Contrast Effects

Contrast effects occur when objects or events differ from surrounding objects/events, and that difference is perceptually magnified. For example, imagine if you were to open your laptop in a normally lit room. You would perceive the screen to be of medium brightness. However, when you open your laptop in a dark room, the screen is going to seem extremely bright compared to the dark room. The change in brightness from dark to a medium level brightness is perceptually magnified, and you perceive your laptop as very bright. That same level of brightness of your screen would seem relatively dim if opened in a very bright room such as a hospital room with bright fluorescent

lights, white floors and white walls. Now, the change from a high level of brightness to a medium level of brightness has been magnified and your laptop may seem dimmer than it actually is. Contrast effects happen in every modality and for many perceptual cues (Kluender, Coady, & Kiefte, 2003).

Of particular interest to this line of research are spectral contrast effects. Spectral contrast effects (SCEs) occur when there is a change in the frequency compositions of earlier sounds compared to a later target sound. That change will be perceptual magnified such that the perceived change is larger than the physical change. In very simplistic terms, if earlier sounds contain more low frequencies and are followed by a mid-frequency sound, that mid-frequency sound will be perceived as higher in frequency. Similarly, if earlier sounds contain more high frequencies and are followed by a mid-frequency sound, that same sound will be perceived as lower in frequency. SCEs can occur on short-term and long-term time courses. In this example, short-term SCEs occur when the earlier sounds contain more low frequencies right at the end of the sound (e.g., Lotto & Kluender, 1998). The sound immediately following will be perceived higher in frequency. Long-term contrast effects occur when earlier sounds contain more low frequencies overall or as part of its long-term average (e.g., Ladefoged & Broadbent, 1957). This bias becomes part of the overall quality of the sound. A change from that quality to the sound quality of a subsequent sound will be perceptually magnified.

As previously mentioned, many speech sounds can be differentiated based on their formant frequencies. Contrast effects occur for changes in these formant frequencies. One example is the vowels /ɪ/ (‘ih’ like in the word “bit”) and /ɛ/ (‘eh’ like in the word “bet”), which are primarily differentiated on the first formant (F1) with /ɛ/

having a higher F1. If earlier sounds contain a low frequency F1 and the target sound has an ambiguous F1 between that found in /ɪ/ and /ɛ/, that ambiguous formant will be perceived as a higher, leading to more /ɛ/ responses. Thus, the frequency composition of earlier sounds can influence identification of the speech target.

One of the first demonstrations of SCEs in speech was Ladefoged and Broadbent (1957). In this seminal paper, context sentences ('please say what word this is') were followed by target words. Listeners reported what word out of the four options they heard: 'bit', 'bet', 'bat', or 'but'. Importantly, the authors shifted the F1 and F2 frequencies in the context sentence. The results presented here concentrate on the responses 'bit' and 'bet' because these options include the vowels /ɪ/ and /ɛ/. When the first formant of the context sentence was not shifted, participants reported hearing a target word as 'bit' more often. However, when the first formant of the context sentence was shifted down, that same target word was reported as the high-F1 'bet' most often. Similarly, when the first formant of the context sentence was shifted up to a high frequency F1, listeners labeled the target word as low-F1 'bit' more often. Thus, the change from the context to the target word was perceptually magnified and listeners reported hearing the option opposite of the context more often

SCEs in Speech vs Non-Speech

Originally, Ladefoged and Broadbent (1957) interpreted these results as a means of compensating for differences between different talkers' voices. However, subsequent evidence has shown that a talker is not necessary to produce SCEs. In fact, later lines of research demonstrated SCEs are both consistent and predictable based solely on the acoustics of the surrounding sounds (e.g., Holt, 2005; 2006). Holt and colleagues have

repeatedly shown that non-speech contexts (i.e., a series of sine tones) were successful at producing SCEs. When these sine tones contexts contained more low frequency tones, they produce high frequency responses more often. When the sine tones contexts contain more high frequency tones, there are more low frequency responses. Non-speech SCEs have been observed with long delays between context and target (Holt, 2005), with different levels of variation in the frequencies of the context (Holt, 2006) and with different durations of contexts (Holt, 2006). In addition to sine tones, SCEs have been shown with contexts that were musical instruments (Stilp, Alexander, Kiefte, & Kluender, 2010) or signal correlated noise (Watkins, 1991). Further, it has been demonstrated that SCEs can be observed when listeners are categorizing musical instruments (varying from tenor saxophone to French Horn; Stilp, Alexander, Kiefte & Kluender, 2010; Assgari, Frazier, & Stilp, 2018). So, SCEs are observed when both the context and targets are non-speech. These findings show that SCEs are an acoustic phenomenon and not speech specific.

Non-speech SCEs show that adjusting for talker differences cannot be the only purpose of these effects. However, just because SCEs occur with non-speech does not mean that they do not serve a purpose in speech perception. Speech is a unique stimulus for many reasons that non-speech stimuli often fail to capture. Primarily, the spectra of non-speech stimuli are extremely simple relative to speech. The majority of studies establishing non-speech influences on SCEs use sine-tones as context (Holt, 2005; 2006). These sine tones differ from speech in important ways. First, the speech signal spans a wide range of frequencies whereas sine tones sampled very limited ranges of frequencies. In addition, the amplitude fluctuations of sine tones are extremely simplistic (on versus

off) relative to speech. In speech, the amplitudes of different frequencies vary producing a complex stimulus. Another non-speech stimulus used in SCE experiments is signal correlated noise (e.g., Watkins, 1991). While signal correlated noise captures the overall amplitudes of frequencies in speech, it still fails to capture the minute fluctuations of frequencies over time. Thus, in order to assess the role of the acoustic complexity of speech may play in SCEs, natural speech should be used.

Listeners have extensive experience with speech as our primary means of communication. As such, speech is encountered in many different contexts, while certain non-speech stimuli are only encountered in a laboratory environment. There is mixed evidence on whether SCEs using non-speech contexts are equivalent to SCEs with speech contexts. In a series of studies, Watkins and colleagues (Watkins, 1991; Watkins & Makin, 1994) compared speech and non-speech contexts and their ability to elicit contrast effects. The authors found that while both types of context elicited contrast effects, speech appeared to be more effective, leading to larger SCEs. They also found that synthetic speech produced smaller effects than unmodified natural speech, but these comparisons were not directly quantified. In addition, Sjerps, Mitterer and McQueen (2011) found that when stimuli lacked speech-like acoustic variation, particularly in f0, contrast effects appeared smaller. In other words, natural stimuli were more effective at producing SCEs than synthesized speech or other laboratory stimuli, but these comparisons were only qualitative. On the other hand, Laing et al. (2012) reported no statistical difference between the magnitudes of SCEs that were produced with sine tone contexts compared to a speech context. Thus, whether non-speech SCEs are comparable

to SCEs with speech is not clear. As such, attempts to generalize findings regarding SCEs produced by non-speech contexts versus speech contexts need to be done cautiously.

In addition to observing SCEs in a variety of conditions, a variety of methodologies have been successful in producing these effects. Each method is slightly different but accomplishes a similar goal: to create a spectral change from earlier sounds to the target sound. Originally, Ladefoged and Broadbent (1957) shifted the entire contour of F1 across a sentence up or down using speech synthesis. In a later study by Ladefoged (1989), instead of manipulating F1 contours on a computer, the author simply adjusted the shape of his mouth when producing sentences. F1 is related tongue height: when the tongue is raised, F1 is lower in frequency; when the tongue is lower, F1 is higher in frequency. This results in the same manipulation as Ladefoged & Broadbent (1957), albeit less controlled. In order to have more acoustic control, researchers digitally filter speech in order to produce the desired characteristics. Filtering speech also enables experimenters to use a single sentence filtered in two different ways as the context (e.g., low F1 emphasized, high F1 emphasized). This way, any observed effects are due to the filtering and not any other acoustic differences between sentences. Stilp, Anderson and Winn (2015) tested three different methods of filtering: narrowband, broadband, and spectral envelope difference filters (Watkins, 1991). All methods were successful in producing SCEs.

Level of Processing of SCEs

It can be argued that SCEs make speech perception easier by disambiguating otherwise ambiguous sounds. If the frequency composition of a vowel sound does not promote clear identification, the frequency content of earlier sounds can bias

identification, through SCEs, to make the vowel sound unambiguous. Thus, SCEs aid phoneme perception with in turn aids word perception. However, it was not clear when in speech perception SCEs occur. Sjerps and Reinisch (2015) sought to address this question using a lexically guided learning task. Lexically guided learning occurs when listeners identify an ambiguous phoneme based on the context in which is presented (Norris, McQueen, and Cutler, 2003). For example, if a sound that is called /f/ or /s/ equally often in isolation is presented to listeners in the context of a word that can only end in f (e.g., giraffe), listeners learn to identify that sound as /f/. Later, when that same sound is presented in the context of words that could end in /s/ or /f/ (e.g., leaf vs lease), listeners are more likely to continue to label that sound as /f/ (e.g., indicate they heard leaf). In other words, listeners will adjust their category boundaries based on the training they receive. Sjerps and Reinish (2015) had listeners perform a lexically guided learning task, but the context was also filtered in order to produce a SCE. Using spectral envelope difference filters, the authors made context words more /f/-like or /s/-like to promote more /s/ and /f/ responses, respectively. When tested, listeners failed to show a lexically guided learning effect. This suggests that the ambiguous sound was disambiguated by the SCE prior to lexical decision making. At test, with no spectral manipulation to disambiguate the sound, the sound was just as ambiguous as before training. This finding suggests that SCEs act on speech prior to listeners deciding which word they heard. Further evidences for the prelexical nature of SCEs comes from an ERP study. Sjerps, Mitterer, & McQueen (2011) found evidence of SCEs occurring at N1. N1 has been argued to be pre-decision making (Roberts, Flagg, & Gage, 2004, as cited by Sjerps,

Mitterer, & McQueen, 2011). This study supports that SCEs influence speech perception prior to listeners deciding what word they heard.

Specific inventories of speech sounds vary dramatically across languages. With the immense variation in languages, it is possible that SCEs are specific to a select few languages. However, evidence exists that suggest otherwise. First, SCEs have been demonstrated in a variety of languages including Dutch, English and Spanish (e.g., Sjerps & Smiljanic, 2013). Furthermore, SCEs have been shown for listener's non-native language (Sjerps & Smiljanic, 2013). Importantly, all the languages tested have the /o/ - /u/ distinction which was the phoneme distinction listeners identified. These results suggest that contrast effects are not language specific: as long as the phoneme distinction is present in the listener's language, an SCE can be observed. However, if the listener cannot distinguish between the two response options because their language only possesses one of those phonemes, all target sounds will be categorized into one category and no response shifts can be observed (Kang, Johnson, & Finley, 2016). In other words, the context will influence speech perception no matter the language. However, the ability to measure that influence may depend on the language experience of the listener.

Talker Normalization and SCEs

Arguments for No Influence of Talker in SCEs

Originally, Ladefoged and Broadbent (1957) interpreted their findings as a means for compensating for differences between talkers. With later research demonstrating that non-speech context could elicit SCEs, some argued that talker information did not play a role in SCEs. An attempt to directly measure the influence of talker information on SCEs used acoustic manipulations of sentences to induce the perception of different talkers

(Laing, Liu, Lotto & Holt, 2012). Laing, Liu, Lotto and Holt (2012) amplified different regions of F1 or F3 in sentences in order to induce the perception of different talkers. The authors reported that these filtered sentences were equally discriminable. In addition, non-speech contexts were presented with similar manipulations (i.e., sine tones sampling the same F1 or F3 regions). These contexts were presented before consonant targets varying from /da/ to /ga/, which are primarily differentiated based on F3. Participants had to indicate whether they heard the consonant /da/ or /ga/ following the context. SCEs were observed with F3 manipulations but not F1 manipulations. Since /da/ and /ga/ are differentiated based on F3, this result is expected. In addition, the same pattern of results (i.e., SCEs with F3 tones, no SCE with F1 tones) were replicated using non-speech contexts. Since tones were equally effective at producing SCEs as speech, and SCEs were not observed simply due to different-sounding talkers (i.e., in all conditions), the authors suggested that talker information has no influence on SCEs.

Evidence for Talker Influence in SCEs

However, manipulation of single formants fails to capture the complexity of acoustic differences between talkers. It is unclear whether these manipulations actually resulted in the percept of different talkers. The authors reported that their manipulations were discriminable (Laing et al., 2012)but did not explicitly state that listeners were asked if the manipulations sounded like different talkers. Just because two sentences are discriminable does not mean that they are perceived as different talkers. As previously mentioned, talkers differ on a variety of acoustic cues, not just F1 or F3. To fully capture the acoustic differences between talkers, natural speech from different talkers should be used. Assgari & Stilp (2015) used speech from 200 different talkers, each speaking a

different sentence, in an SCE experiment. In addition, two other conditions were tested: a single talker producing one sentence (presented 200 times), and a single talker producing 200 different sentences. The one talker 200 sentence condition allowed the assessment of whether SCEs vary due to the acoustic variability stemming from different talkers or simply the variability from different sentences. In all cases, on each trial, listeners heard a sentence followed by a target vowel (perceptually varying from "ih" as in "bit" to "eh" as in "bet"). Results showed that the degree to which talker information influenced SCE magnitudes was dependent on the magnitude of the spectral peaks added to the sentences. When spectral peaks were large (i.e., +20 dB amplification), as was typical of contrast effect research, all conditions showed equivalent SCE magnitudes. In this case, talker information had no influence on the magnitude of contrast effects. However, when spectral peaks were modest (i.e., +5 dB amplification), there were clear difference between conditions. The 200-talker condition showed smaller SCEs than both single-talker conditions. Thus, talker variability did lead to a clear decrease in SCE magnitude (Assgari & Stilp, 2015). In other words, talker variability decreased the influence of earlier sounds on categorization of the subsequent vowel sound.

These findings likely explain some of the disparity in previous research investigating whether or not talker information influences SCEs. Studies that reported no effect of talker used dramatic changes from the frequency compositions of earlier sounds to the target (e.g., Laing et al., 2012; Assgari & Stilp 2015 Experiment 1). When smaller changes from the frequency composition of the earlier sounds to the target were used, talker information does influence SCEs (Assgari & Stilp, 2015 Experiment 2). What is not clear is exactly what it is about hearing different talkers that influences SCEs. The

sentences in the 200-talker condition were chosen randomly. Post hoc acoustic analysis of these sentences revealed that these stimuli were highly variable based on f0. Closer inspections of this f0 variability suggested that SCEs were smallest when the talkers' mean f0 calculated across the entire sentence were most variable. However, this f0 variability was confounded with talker gender variability, as talker gender was also not controlled.

Talker normalization and SCEs are both influences of context on speech perception. The extent of these influences is both mitigated by low-level acoustic variability. But, there is an important distinction between these influences. Talker normalization effects are a cost to the listener. When hearing different talkers, the listener is slower and less accurate at speech perception compared to hearing just one talker. When the talkers are very different, this cost is larger. On the other hand, SCEs act as a benefit to the listener. When a sound is otherwise ambiguous, the context of earlier sounds can help to disambiguate that sound. When hearing different talkers, this benefit is reduced but still present. This is an important distinction when making connections between talker normalization and SCEs.

Study Motivation

Talker normalization effects are well established. Speech perception is slower and/or less accurate when different talkers are heard relative to hearing a single talker. These effects are due in part to differences in the pitch (f0) of each talker's voice (Goldinger, 1996). Recent studies found that talker information influences the extent to which SCEs bias categorization of subsequent speech sounds (Assgari & Stilp, 2015). Post-hoc acoustic analyses of the stimuli used to measure this effect hinted at an

influence of talker f0**,** but these stimuli confounded f0 variability with gender variability. In addition, these sentences were randomly selected, limiting the ability to quantify potential links between these measures and SCEs. In order to understand why and how talker information modulates SCEs in speech perception, direct investigations of possible explanations are still needed. The proposed studies will test a variety of possible ways that talker information might influence SCEs in speech perception.

The first study will seek to control for two different sources of talker variability that initially varied in Assgari and Stilp (2015): f0 variability and talker gender. Experiment 1 of the first study will isolate talker gender and f0 variability in order to test their influence on SCEs separately. Results suggest f0 variability in preceding sentences has a stronger influence on SCEs than talker gender. Experiment 2 will confirm this influence of f0 variability in preceding sentences on SCEs regardless of talker gender.

If f0 variability has an influence on SCEs, it is possible that other sources of low-level acoustic variability will also affect SCEs in speech categorization. In Study 1, stimuli were chosen based on f0, but f0 is just one low-level acoustic cue to talker changes. Perhaps, the same approach can be applied to F1 since our target vowels (/ɪ/ and /ɛ/) differ primarily on F1. Study 2 will investigate this possibility and look specifically at the influence of F1 variability on SCEs.

In order to establish a stronger relationship between low-level variability and SCEs, low-level variability should be directly manipulated. Study 3 uses the same stimuli in each condition, except for the manipulations of f0 variability. This way, any differences in observed SCEs must be due to the manipulation and not any differences between stimulus sets. Study 3 will manipulate low-level acoustic variability in natural

speech in order to establish a more direct relationship between low-level variability and SCEs.

Research has supported the idea that local variability (i.e., trial-to-trial variability) has a stronger influence on talker normalization than global variability (i.e., total variability in a condition) (Johnson, 1991). This has been demonstrated through smaller talker normalization effects when stimuli are blocked by talker compared to when they are randomly presented (i.e., no increase in response time or decrease in accuracy due to talker variability). In addition, blocking by talker essentially eliminates all talker variability in that condition. While studies controlling trial-to-trial variability are common in talker normalization research, no SCE experiments have manipulated trial-to-trial variability in talker acoustics. Study 4 tested conditions where talker changes from trial to trial were relatively small but overall variability in each the condition remained high. In essence, Study 1 measured how high local (trial-to-trial) and global (across the entire block) acoustic variability influences SCEs in speech perception; this experiment minimized local variability but maintained global variability.

Finally, measures of talker normalization were collected in each of the above studies. The ways in which talker normalization and SCEs are measured are very different. Talker normalization is measured through decreased accuracy and slower response times when comparing responses following multiple talkers to responses following single talker. SCEs are measured through the magnitudes of shifts in categorization of the target speech sounds. Response times and accuracy were measured in SCE experiments. This way, we probed whether or not talker normalization can be measured in an SCE task. This study served to formally link SCEs and talker

normalization by showing that low-level acoustic variability influences both processes in the same task.

CHAPTER II

GENERAL METHODS AND ANALYSIS

In the current studies, many methodological details were consistent across studies. They are briefly introduced here.

Acoustic Measurements

Acoustic measurements were obtained through Praat (Boersma & Weenink, 2017). Briefly, Praat is a commonly used tool used by speech researchers that measures a variety of parameters in speech. All reported measurements were hand checked and edited where necessary to ensure the best degree of accuracy.

Sentences

To ensure the current studies were evaluating the effect of speech on contrast effects, natural speech was used whenever possible. Context sentences were drawn from the same corpus of sentences as Assgari and Stilp (2015): Texas Instrument and Massachusetts Institute of Technology speech corpus (TIMIT; Garofolo et al., 1993). This corpus consists of 6300 sentences spoken by 630 different talkers from 8 different dialect regions in the United States. There are 438 male and 192 female speakers. In all conditions of all studies, gender was balanced except where explicitly manipulated.

Chosen sentences were randomly assigned to low F1 (100-400 Hz) or high F1 (550-850 Hz) conditions. Sentences were processed with a band pass filter to add +5 dB spectral peaks in either the low F1 or the high F1 region based on condition assignment.

In order to ensure that our filters had the desired effect, only sentences with relatively equal energy in the high vs low f1 regions (within ±5dB of each other) were considered. As such, adding spectral peaks resulted in the desired magnitude of spectral difference.

In all studies except Study 1a, a single talker condition was included. This condition served two purposes. First, it served as a replication of previous results (e.g., the Single talker one sentence condition from Assgari & Stilp, 2015). Second, it allowed for within-subject comparisons for a condition with no long-term acoustic variability. Since participants heard the same talker produce the same sentence on every trial, there is no long-term f0 variability, which should produce an "upper limit" in terms of SCE magnitude.

Vowels

Target vowels were a 10-step continuum of vowels ranging for /ɪ/ to /ɛ/. These vowels have been used in previous studies and are reliably biased by SCEs (e.g., Stilp, Anderson & Winn, 2015; Assgari & Stilp, 2015; Stilp & Assgari, 2018). Vowels were synthesized based on natural recordings from a male talker. These speech samples were resynthesized using Linear Predictive Coding (LPC) in Praat (Boersma & Weenink, 2017). The /ɪ/ endpoint has an F1 that linearly increases from 400 to 430 Hz while F2 linearly decreases from 2000 to 1800 Hz. The /ɛ/ endpoint has an F1 that linearly decreases from 580 to 550 Hz while F2 linearly decreases from 1800 to 1700 Hz. The vowel continuum was created by taking these endpoint vowels and morphing their formants through a script in Praat. Final vowel stimuli were 246 ms in duration with fundamental frequency set to 100 Hz throughout the vowel.

Trials

All filtered sentences and target vowels were equated for root mean square (RMS) amplitude. Experimental trials consisted of a filtered sentence followed by a 50-ms silent interstimulus interval and then a target vowel. All stimuli were upsampled to 44100 Hz.

Participants

All listeners in the reported studies participated in exchange for course credit. No listener participated in more than one study and all self-reported normal hearing.

Procedure

All experiments began by obtaining informed consent. Listeners were then seated in sound attenuated chambers (Acoustic Systems, Inc., Austin, TX). Stimuli were presented at 70 dB SPL over circumaural headphones (Beyerdynamic DT-150, Beyerdynamic Inc. USA, Farmingdale, NY). The experiments were executed by custom scripts in Matlab and were self-paced.

In all studies, participants first completed a set of practice trials before continuing to test. These trials were composed of 20 sentences from the AzBio corpus (Spahr et al., 2012) paired with the endpoint vowels, as categorizing endpoints of the vowel continuum is objectively correct or incorrect. If the participant failed to reach 80% accuracy on endpoint vowels after one block, they could repeat the practice trials up to two more times to reach 80% accuracy. If after three blocks of practice trials they did not reach criterion, they did not continue to test. This ensured that participants could categorize the target vowels into their intended categories and made the shifts of category boundaries (SCEs) interpretable. If participants passed practice trials, they continued to test. However, they had to maintain 80% correct on vowel continuum endpoints across all

conditions in order to be included in the data analysis. If a participant elected to withdraw from the experiment at any point, his/her data were not included in the analyses.

In all experiments, participants were asked to respond to the target vowel, indicating that they heard "'ih' as in 'bit'" or "'eh' as in 'bet'". They were told to respond as quickly and accurately as possible. Participants were allowed to take breaks in between blocks. Blocks were 160 trials long (4 repetitions of each unique sentence/vowel pairing) and each took about 12 minutes to complete. Each experiment tested up to four blocks, making the maximum duration of an experimental session roughly one hour.

Data Analysis

SCEs

A schematic of the measurement process is found in Figure 1. In these experiments, SCEs were measured as shifts in categorization of vowels following contexts filtered to have different reliable spectral properties. To quantify these shifts, logistic regressions predicting /ɛ/ responses were fit to each individual's responses to vowels in the low-F1-amplified context sentences and high-F1-amplified context sentences individually. The 50% points of these regressions curves, where listeners are equally likely to respond /ɪ/ and /ɛ/ to a given target vowel, were identified. The 50% point was then translated to the corresponding stimulus number along the continuum (from 1 to 10); this number was interpolated if needed. The SCE was operationalized as the difference between the 50% points, measured in stimulus steps (black arrow in Figure 1). SCEs were calculated for each participant in a given condition, then averaged over all participants. Mean SCEs were then compared across experimental conditions using repeated-measures ANOVAs.

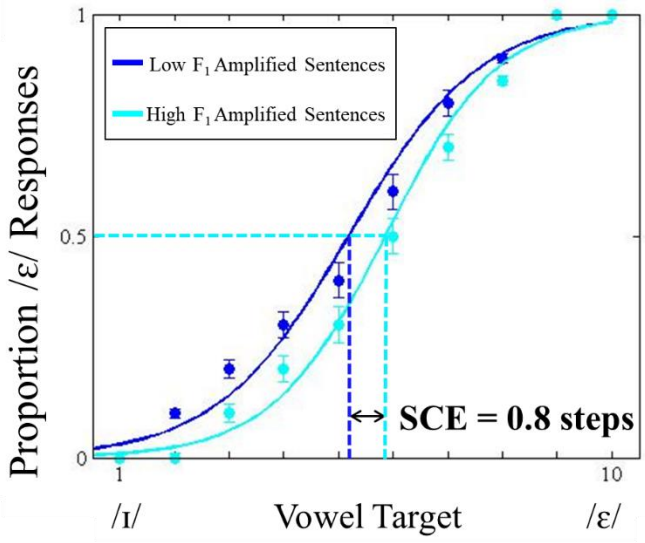

Figure 1. Measurement of SCEs on example data. Mid-point shifts were assessed for each individual as illustrated here on example data (not from any study reported here). These mid-point shifts were averaged to obtain a mean midpoint shift in each experimental condition. The mean mid-point shifts were compared.

Deviance Measures

To assess how well the logistic functions used to obtain mid-points fit each participant's data, measures of deviance were obtained through the glmfit command in Matlab. Deviance measures were obtained for the logistic fits to each individual's responses to vowels following the low-F1 and high-F1 context sentences separately. Plots of deviance measures can be found in the Appendix.

Confidence Intervals around midpoints

In order to determine a measure of spread around mid-points, 95 percent confidence intervals were calculated around the midpoint of each listener's logistic functions using the psignifit package in Matlab. In addition, a measure of overlap of confidence intervals was obtained by subtracting the upper bound of the low-F1 confidence interval from the lower bound of the high-F1 confidence interval. Plots of confidence intervals and measures of overlap can be found in the Appendix.

Response Times

Response times were collected in Matlab during Study 1b through Study 4. Response times were collected using a button box. The button box records a spike in amplitude when a button is depressed, completing a circuit and indicating a response. Response times are specifically measured as the duration between the onset of the vowel and the spike in amplitude described above.

Independent of manipulations of talker / acoustic variability, response times were expected to differ based on what stimulus along the vowel continuum was presented. Ambiguous stimuli (i.e., towards the middle of the continuum) will naturally elicit longer response times because listeners are slower to respond to stimuli that are harder to identify. Conversely, vowels at the end of the continuum will elicit shorter response times because they are less ambiguous and objectively easier to categorize. The changes in response times from single-talker to multi-talker conditions were most important. While the exact pattern of this change in response times is unknown, some of the more likely possibilities are illustrated in Figure 2. If the effect of multiple talkers influences the entire continuum equally, then there will be an overall and equal increase in reaction times (upper left panel in Figure 2). It is possible that response times are already at ceiling for ambiguous stimuli in single talker conditions. Thus, the influence of multiple talkers could only occur for non-ambiguous stimuli. If this is the case, then response times will increase only for non-ambiguous stimuli toward the ends of the continuum (upper right panel in Figure 2). If multiple talkers only influence categorization of ambiguous stimuli, then response times will increase for ambiguous stimuli near the middle of the continuum (lower left panel in Figure 2). If multiple talkers increase

reaction times for the whole continuum but influence ambiguous stimuli more than unambiguous stimuli, then a combination of these trends will appear with larger increases in response times for ambiguous stimuli (lower left panel in Figure 2). Given these possibilities, response times will be broken down and analyzed based on vowel target.

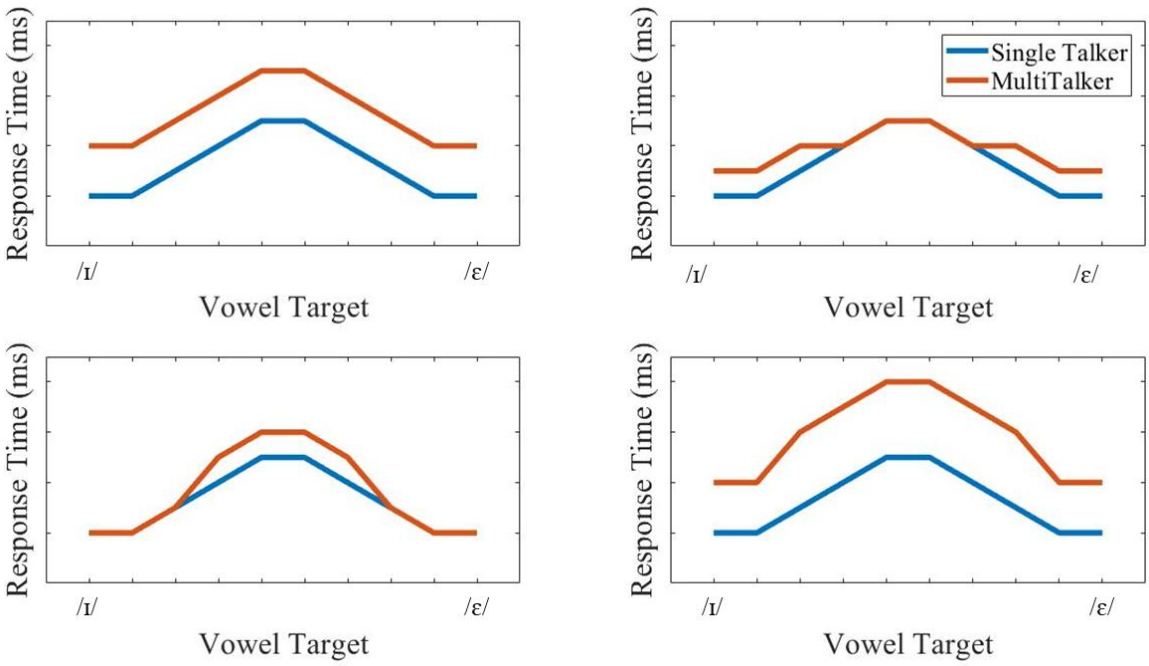

Figure 2. Possible changes in response times from single to multi-talker conditions

Any response times below 150 ms or 3 standard deviations above the participant's average response time were removed. It has been demonstrated that 150 ms is the lower limit of responding after taking into account recognition and motor movements (Luce, 1986). Three standard deviations above the individual response mean suggest that individual may not have responded 'as quickly and accurately as possible' for the removed trial. After removing these points, average response times were calculated for

each vowel collapsing across filter condition since filter condition should not theoretically influence response times.

Accuracy

Response accuracy was collected for Study 1b through Study 4. Throughout these studies, response accuracy was assessed at endpoints during practice and throughout the task. When vowels other than the endpoints are considered along a vowel continuum, accuracy can be difficult to characterize. Several of the stimuli are known to fall in an ambiguous region between the two vowel categories for listeners (i.e., vowels near the middle of the continuum). At the very least, accuracy at the endpoints of the continuum can be assessed (e.g., vowels 1 and 10 in the 10-step continuum). Here again, of most interest was the change in accuracy from a single-talker to a multi-talker condition. As such, proportion of correct responses were compared between single-talker and multi-talker conditions.

Accuracy was characterized by how often participants labeled the continuum endpoints as the intended categories and measured as proportion correct. Accuracy was then averaged over the /ɪ/ and /ɛ/ endpoints giving a single measure of accuracy for each condition. All accuracy measures were rationalized arcsine transformed due to ceiling performance in every condition (Studebaker, 1985). All analyses are performed on transformed accuracy while the plot in the Appendix P shows accuracy as proportion correct.

CHAPTER III

STUDY 1 (ISOLATING CONTRIBUTIONS OF GENDER VARIABILITY AND F0 VARIABILITY)

Aims

Following the suggestion of Assgari & Stilp (2015), this study investigated what aspects of talker changes influences SCEs. Study 1 sought to separate the contributions of talker gender and f0 variability to diminished SCEs reported in Assgari & Stilp (2015). In previous literature, there have been conflicting reports of what influence gender may have on SCEs. Watkins (1991) found that talker gender had no influence, with contexts spoken by both male and female talkers producing SCEs of similar magnitude. However, in this study, there was a single talker from each gender, and perhaps more importantly, talker gender was blocked. On the other hand, Johnson, Strand and D'Imperio (1999) reported that listeners' categorization boundaries shifted based on their expectation of the gender of the talker. Since f0 and gender were free to vary in Assgari & Stilp (2015), the degree to which either or both of these variables influenced SCEs is unclear. In addition, since talker gender and f0 are closely related (lower f0s for men, higher f0s for women), gender variability and f0 variability were also closely related, so the influences of these two characteristics need to be explicitly separated. Study 1a investigated the influence of f0 variability when talker gender was blocked. Study 1b mixed talker gender and measured the influence of f0 variability in the presence of talker gender variability.

Methods

In Study 1a, the average f0 across each sentence was measured in Praat (Boersma & Weenink, 2017). When necessary, f0 contours were hand edited to ensure accurate measures. These measures of f0 were converted to z-scores within gender. A distribution of z-scores was formed for each gender. High variability sentences were sampled from the tails of these distributions while low variability sentences were sampled from the centers. The resulting distribution of sentences are displayed in Figure 3. Thus, variability was established across trials using average f0 across sentence duration. Forty sentences were selected for each condition. This formed four groups: men low f0 variability (*Mean f0* = 121.29, *SD* = 9.16), men high f0 variability (*Mean f0* = 123.39, *SD* = 33.18), women low f0 variability (*Mean f0* = 203.89, *SD* = 9.17), women high f0 variability (*Mean f0* = 199.84, *SD* = 33.27). High and low variability conditions were matched as closely as possible based on the standard deviation to ensure the variability within each gender is equivalent.

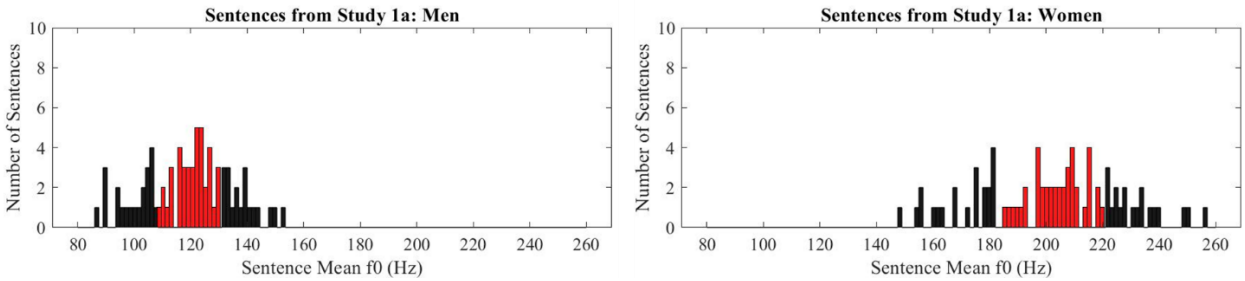

Figure 3. Distributions of Mean f0 of sentences for Study 1a. The left panel depicts sentences chosen for men and the left panel depicts sentences chosen for women. Average sentence f0 is along the x-axis and number of sentence is along the y-axis. Sentences in red were chosen for the low variability conditions. Sentence in black were chosen for the high variability conditions.

In Study 1b, f0 measures were obtained through Praat. These measures were converted to z-scores between genders. Similar to Study 1a, a distribution of candidate sentences was created based on these z-scores while intentionally mixing talker gender.

Low and high variability groups were formed by sampling specific sections of the distribution. The resulting distribution of sentences is displayed in Figure 4. Low variability sentences were sampled from the center of the distribution (*M* = 164.81, *SD* = 9.78). High variability sentences were sampled from the tails of the distribution (*M* = 161.78, *SD* = 45.64). In general, male speakers were pulled from the lower tail (lower f0) while female speakers were pulled from the upper tail (higher f0). Forty sentences were pulled for each condition and gender was balanced within each variability condition keeping gender variability the same across conditions.

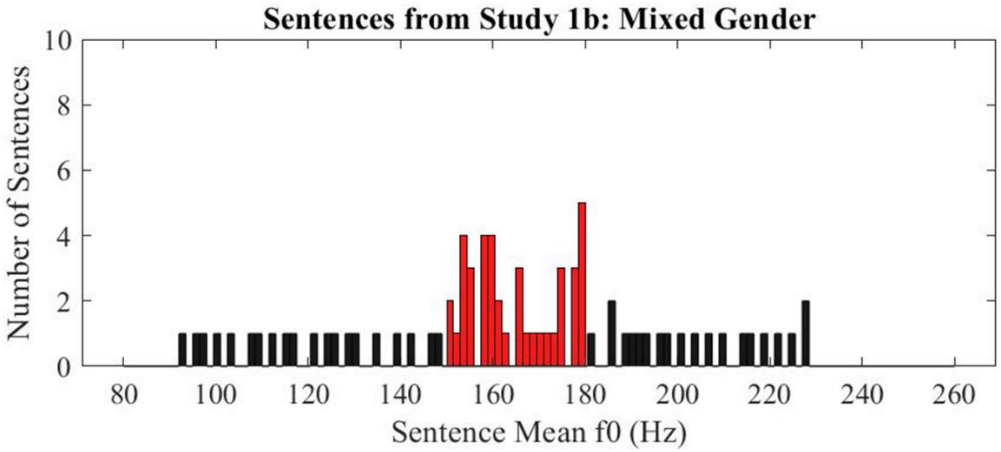

Figure 4. Distribution of Mean f0 of sentences for Study 1b. Average sentence f0 is along the x-axis and number of sentences is along the y-axis. Sentences in red were chosen for the low variability condition. Sentences in black were chosen for the high variability condition.

Hypotheses

In Study 1a, if f0 variability influences SCEs regardless of gender, a main effect of f0 variability would be evident such that SCEs are smaller in high-f0-variability conditions than low-f0-variability conditions (similar to Assgari & Stilp, 2015). If talker gender influences SCEs regardless of variability, differential effects of talker gender would be expected for men versus women. Since the target vowels were spoken by a male talker, smaller SCEs would be expected when context sentences were spoken by

women compared to men. If both gender and f0 variability influence contrast effects, we would expect both main effects to be significant. An interaction would suggest that the effect of f0 variability depends on talker gender.

Results

SCEs

In Study 1a, 20 listeners participated. No listener was excluded from the analysis. A 2 (talker gender: female, male) by 2 (f0 variability: high, low) repeated measures ANOVA was conducted (see Figure 5). A main effect of f0 variability was significant, with high variability conditions (M = *0.30*, SE = 0.09) producing smaller contrast effects than low variability conditions (M = 0.50, SE = 0.09) ($F(1,19) = 5.31$, $p = .03$, $\eta_p^2 = 0.22$). Neither the main effect of gender ($F(1,19) = 2.32$, $p = .14$, $\eta_p^2 = 0.11$) nor the interaction ($F(1,19) = 0.004$, $p = .95$, $\eta_p^2 = 0.00$) were statistically significant. This pattern of results suggests that f0 variability, not talker gender, influences SCEs in speech categorization. Specifically, when f0 variability is high, there are smaller shifts in categorization of the target vowels regardless of the gender of the talkers.

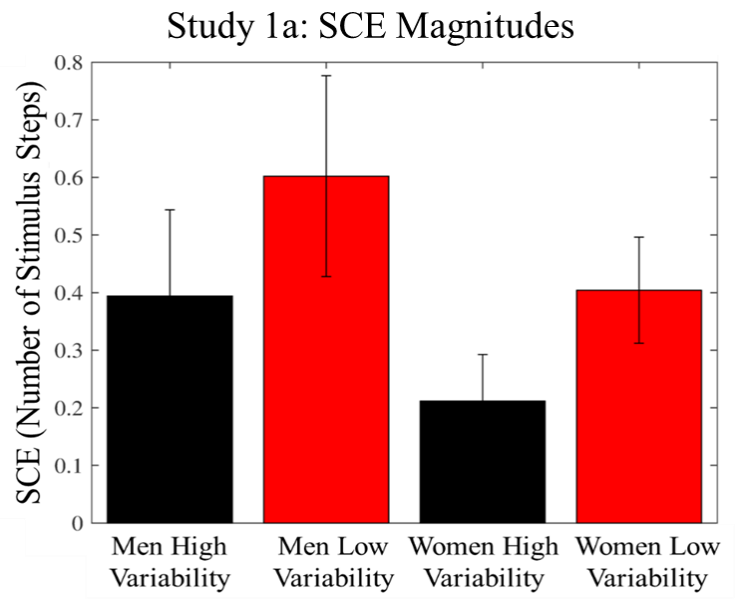

Figure 5. Contrast effect magnitudes from Study 1a. Gender and variability conditions are along the x-axis and contrast effect magnitude, measured in stimulus steps, is along the y-axis. Red bars represent low variability conditions and black bars represent high variability conditions. Error bars depict standard error of the mean.

In Study 1b, 25 listeners participated. Two listeners failed practice and one was removed for failing to maintain 80% accuracy on endpoints across the entire experiment, leaving 22 listeners in the analysis. In addition, upon inspection of the data, one outlier was identified in the High F0 variability condition. This participant exhibited a contrast effect that was 4.16 stimulus steps. This is well beyond what is typically expected from a +5 dB peak where context effects typically fall between 0.2 and 0.8 stimulus steps (see Assgari & Stilp, 2015 and Stilp, Anderson & Winn, 2015). In addition, a contrast effect of 4.16 steps was more than three standard deviations from the mean of the high f0 variability group (M = 0.4104, SD = 0.938 steps). This outlier was removed and the remaining 21 participants were included in the analyses. Deviance measures did not differ systematically across filter conditions or experimental conditions (see Appendix A). The confidence intervals and their overlap around midpoints did not differ

systematically based on filter conditions and experimental conditions (see Appendix B). The ANOVA was significant ($F(2,40) = 3.360$, $p = 0.045$, $\eta_p^2 = 0.114$; see Figure 6). Bonferroni corrected pairwise t-tests indicated that the SCE in the high f0 variability condition ($M = 0.21$, $SE = 0.10$) was significantly smaller than the SCE in the single talker condition ($M = 0.56$, $SE = 0.10$ $p = 0.03$). The low f0 variability ($M = 0.58$, $SE = 0.14$) condition was not significantly different from the high f0 variability condition ($p = 0.20$)

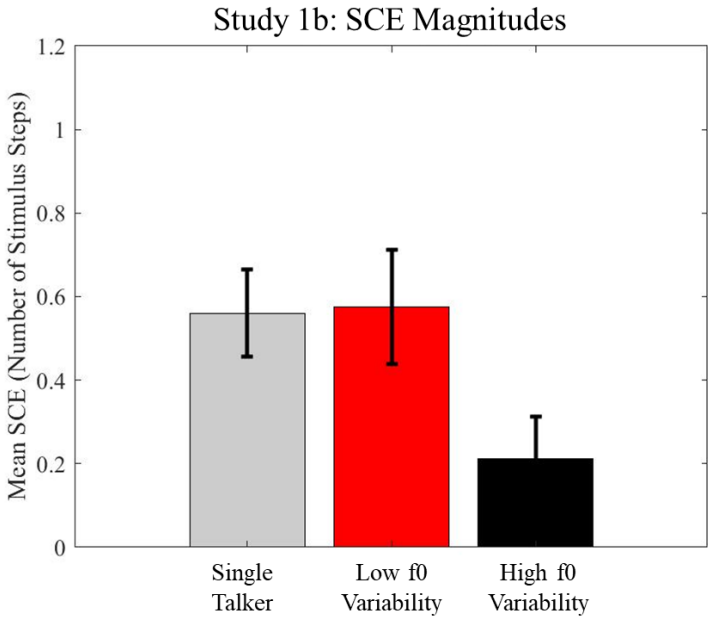

Figure 6. Contrast effect magnitudes from Study 1b. f0 variability conditions are along the x-axis and contrast effect magnitide is along the y-axis. The gray bar corresponds to the single talker condition, the red bar corresponds to the low f0 variability condition, the black bar corresponds to the high f0 variability condition. Error bars depict standard error of the mean.

Response Times

As previous reported, data from 23 listeners passed criterion from Study 1b and were included in the response time analysis. A 3 (Condition: high f0 variability, low f0

variability, single talker) X 10 (Vowel Target) repeated measures ANOVA was conducted on response times collected during Study 1 (see Figure 7). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the main effect of vowel ($p < 0.001$). As such, the main effect of vowel is reported with a Greenhouse-Geisser correction. The ANOVA indicated a significant effect of condition ($F(2,40) = 13.78$, $p < 0.001$, $\eta_p^2 = 0.41$). Bonferroni corrected post hoc pairwise t-tests for condition indicated that the single talker ($M = 832.89$, $SE = 35.97$, $p < 0.001$) and low f0 variability ($M = 888.95$, $SE = 30.77$, $p = 0.03$) elicited faster response times than the high f0 variability ($M = 935.08$, $SE = 32.72$). Differences in response times in the Low f0 variability and the single talker conditions were marginally significant ($p = 0.051$). The ANOVA also indicated a significant main effect of vowel ($F(2.73,54.55) = 13.276$, $p < 0.001$, $\eta_p^2 = 0.40$). The main effect of vowel was driven by faster response times to the endpoints relative to the mid-continuum vowels, as was expected (see Appendix C for pairwise t-tests with Bonferroni corrections). In addition to the main effects, the interaction between condition and vowel was also significant, ($F(18,360) = 1.85$, $p = 0.02$, $\eta_p^2 = 0.09$). This significant interaction suggests that the increases in response times across talker conditions depended on vowel. Some vowels showed larger increases in response times from single to multiple talker conditions relative to others. Particularly, vowels at the /ɪ/ endpoint showed greater sensitivity to f0 variability relative to vowels at the /ɛ/ endpoint (see Appendix D for pairwise t-tests with Bonferroni corrections).

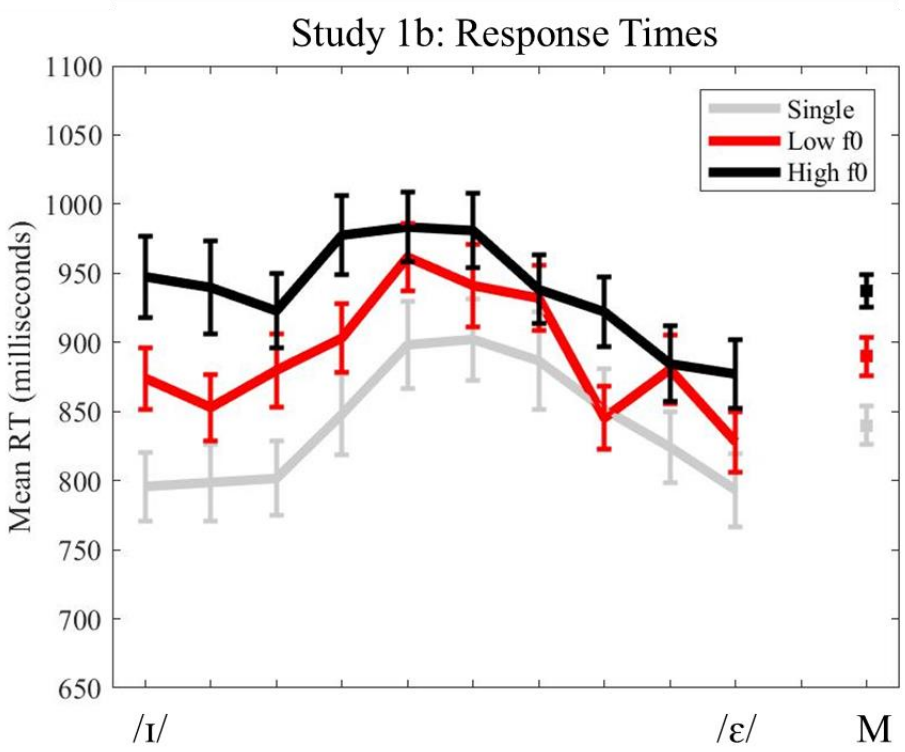

Figure 7. Response times by condition for Study 1b. The vowel continuum is represented along the x-axis with the /ɪ/ endpoint on the left and the /ɛ/ endpoint on the right. Response times in milliseconds are represented on the y-axis. Average response times collapsed across vowels are shown on the far right. The black line represents the high f0 variability condition. The red line represents the low f0 variability condition. The gray line represents the single talker condition. Error bars depict standard error of the mean.

Accuracy

As previous reported, data from 21 listeners were included in the analyses for Study 1b. A one-way repeated measures ANOVA of condition (high f0 variability, low f0 variability, single talker) indicated that accuracy did not change as a function of condition, ($F(2,40) = 0.723$, $p = 0.49$, $\eta_p^2 = 0.04$). Plots of accuracy can be found in Appendix P.

Discussion

The results of Study 1 suggest that acoustic variability and not gender variability influenced context effects. In Study 1a, talker gender and f0 variability were separated

and tested. Yet, only the main effect of f0 variability was significant. This suggests that talker gender alone does not influence contrast effects. In Study 1b, gender was mixed and f0 variability was tested. Here, f0 variability continued to influence contrast effects. Further, the results of Study 1b closely resemble the observed results in Assgari and Stilp (2015). In Assgari and Stilp (2015), smaller contrast effects were observed in a condition with 200 talkers relative to single talker conditions (0.26 stimulus steps vs. 0.52 stimulus steps respectively). In Study 1b, similar results were observed for a high f0 variability condition relative to a single talker condition (0.21 stimulus steps vs. 0.56 stimulus steps). This suggests that the smaller SCEs observed in the 200-talker condition of Assgari and Stilp (2015) was a result of the f0 variability, and not due to variability in talker gender.

Further, results from Study 1b indicate that response times to the target vowels are influenced by the f0 variability of the context with different-sounding talkers leading to slower response times. When the talkers sounded similar, response times are comparable to when hearing a single talker. These results are similar to what would be expected if listeners were experiencing talker normalization with different-sounding talkers (e.g., Mullenix, Pisoni, & Martin, 1989). The observed pattern of response times combined two of the predicted patterns reported in Figure 2. Response times increased across the entire continuum but more so for the /ɪ/ end of the continuum than the /ɛ/ end of the continuum. This suggest that the /ɪ/ endpoint vowels may be more susceptible to the influence of f0 variability than the /ɛ/ endpoint vowels in Study 1b.

The influence of f0 on SCEs is similar to the findings of Goldinger (1996) where f0 differences influenced talker normalization. In Goldinger's (1996) findings, listeners

were more accurate at recognizing old words when they were spoken by similar talkers (with similar f0s) than when they were spoken by different talkers (with different f0s). In Study 1, the benefit of disambiguation (i.e., larger SCEs that aid in disambiguating ambiguous phonemes) was greater when talkers were acoustically similar than when they were different. In addition, response times were faster when talker sounds similar relative to when talkers sound different. This suggests that talker normalization is more prevalent with different-sounding talkers than similar-sounding talkers. There are two ways to interpret the results that speech perception is better disambiguated when f0 variability is low. On one hand, it could be argued that less acoustic variability benefits the listeners. On the other hand, it could be argued that larger acoustic variability deprives the listener of benefits that would otherwise be present. While these are complementary interpretations, one suggests that f0 variability is detrimental while the other suggests it is the default. In assessing which interpretation is most valid, it is important to consider the 'natural state of affairs' of speech perception. It could be argued that natural speech perception arises from a quasi-random sample of talkers. It is unlikely that a listener would purposely select their conversational partners solely based on voice characteristics. When speech tokens are selected from a large corpus in a quasi-random fashion (as in Assgari & Stilp, 2015) they tend to be highly variable. However, controlling for f0 variability mandates that careful attention be paid to the acoustic characteristics of the tokens to be selected. While drawing analogies from methodological procedures to natural speech perception may be limited, it seems that quasi-random selection of speech will result in highly variable acoustic characteristics. Therefore, it is more likely that

highly variable speech is the default and benefits are gained from hearing similar talkers rather than lost when talkers are more variable.

In Study 1a, there was a general trend for female-talker sentences to show smaller contrast effects than male-talker sentences. This trend was not statistically significant yet it is relatively consistent with previous results. Results of Study 1b suggest that when genders are mixed but variability is low, that they can be treated similar to the single talker condition (inasmuch as both produced similar SCE magnitudes). Therefore, the non-significant trend of gender observed in Study 1a may be attributable to f0 differences between the context and the target rather than difference in gender. In Study 1a, the average f0 across both conditions containing women was 201.87, which is 101.87 Hz higher than the f0 of the target vowels (100 Hz). Thus, f0 was changing by 101.87 Hz, on average, from the context to the target when the context was spoken by a woman. The average f0 across conditions containing men was 119.71 Hz, which is only 19.71 Hz higher than the f0 of the target vowels. Thus, the f0 change from the context to the target was much smaller, on average when the contexts were spoken by men. It is likely that adjustments for f0 differences on trials where the context was spoken by women were much greater than when the contexts were spoken by men. In Study 1b, gender of the talkers was mixed and the average f0 of the conditions was more similar (high f0: 161.78 Hz, and low f0: 164.81 Hz). Any adjustments due to f0 differences from context to target would be expected be similar for both the high f0 and low f0 conditions. While the f0 change from context to target was not the focus of these experiments (and was an inevitable consequence of blocking by talker gender), it could potentially inform why a non-significant but consistent trend of talker gender is observed in these data.

It deserves mentioning that the relationship between gender and f0 is somewhat arbitrary for a few reasons. First, it is possible to view gender along a continuum rather than as a binary distinction, as is typical in psychological research. In reality, gender is more of a continuum with male and female at the endpoints. Individuals can fall anywhere on this continuum. When making our male/female distinctions we relied on the annotation provided with TIMIT. It is unclear how the gender of each individual speaker in TIMIT was assessed or if the corpus contains speakers that do not fall into the typical male/female distinctions. What is clear is that each sentence is labeled as either male or female suggesting that gender was treated as a binary distinction. Second, the relationship between f0 and gender is purely physiological and arises because, in general, those who identify as male have larger vocal folds producing lower f0s. So, while the claim that males tend to have lower f0 frequencies is generally correct, there will definitely be exceptions. It is possible to have a high-pitched male (i.e., a male with small vocal folds) and a low-pitched female (i.e., a female with large vocal folds). The ability to find both males and females with f0s similar enough to create a mixed gender, low f0 variability condition supports this claim. In this condition, we were able to find men and women whose f0s ranged a span of 30 Hz (from 150 Hz to 180 Hz), a relatively narrow range. Thus, gender and f0 are dissociable but follow predictable patterns. If a gender effect in speech perception or production is supported by the data, researchers should assess whether that relationship could be better explained by f0 variability. If an effect of f0 is also supported, it is likely preferable to interpret differences based on physiological measurements rather than an arbitrary societal distinction. Further, f0 is a continuous variable whereas gender, as discussed, is typically treated categorically. Continuous

variables allow for closer inspection of the relationship between two variables rather than relying on claims about group differences. Using f0 variability to assess the relationship between variables has been successful in this line of study and is discussed further in the interim general discussion following Study 2.

CHAPTER IV

STUDY 2 (F1 VARIABILITY)

Aims

It is apparent the low-level acoustic variability in f0 influences SCEs. Yet, there are many concurrent sources of low-level acoustic variability. As previously mentioned, there are well known differences in f0 based on talker gender. In addition, overall formant frequencies also differ based on talker gender. Men generally have lower f0s and overall lower formants while women's f0 and formants are generally higher (Peterson & Barney, 1952; Hillenbrand et al., 1995). As such, low f0s could be reflective of low F1s while high f0s could reflect high F1s. Thus, high f0 variability could be suggestive of high F1 variability. Of particular interest to the current experiment is the variability in the region of F1 because the target vowels are differentiated based on F1. If there is an effect of low-level acoustic variability, measuring variability in the region that differentiates our target vowels may better explain the influence of acoustic variability on SCEs. Study 2 investigated this possibility and looked specifically at the influence of F1 variability on SCEs.

Methods

Sentences from Study 1b produced different-sized SCEs when arranged according to variability in mean f0 across sentences. In Study 2, these sentences were rearranged based on measures of mean F1 variability. If F1 variability has a stronger influence on SCE occurring for vowels primarily distinguished by F1, then the differences between

low and high F1 variability conditions should be greater than the differences observed in Study 1b. The average F1 of entire sentences was measured in Praat (Boersma & Weenink, 2017). Formant contours were hand edited to ensure accuracy. Similar to Study 1a, these measures were converted to z-scores and a distribution was created based on these measurements. Groups were formed by pulling high variability sentences from the tails of this distribution (Mean F1 = 527.37, *SD* = 62.42) and low variability sentences from the center of this distribution (Mean F1 = 523.37, *SD* = 16.16; see Figure 8). Forty sentences were included in each condition.

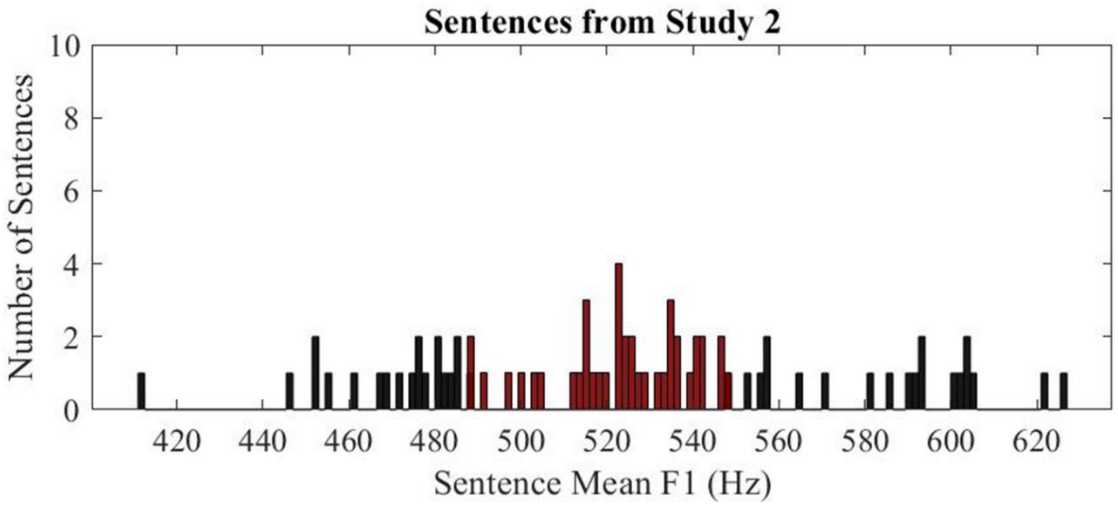

Figure 8. Distribution of Mean F1 of sentences for Study 2. Average sentence F1 is along the x- axis, number of sentences is along the y-axis. Sentences in red were chosen for the low F1 variability condition. Sentences in black were chosen for the high F1 variability condition.

Hypotheses

If F1 variability influences SCEs, a high F1 variability condition will produce smaller contrast effects than low F1 variability, parallel to the results in Study 1b. This experiment also determined whether F1 variability in context sentences has a larger influence on SCEs than does f0 variability. If this is the case, then the difference between

SCE magnitudes should be greater for low vs high F1 variability than for low vs high f0 variability.

Results

In Study 2, 25 listeners participated. One listener was removed for failing to maintain 80% accuracy on endpoints across the entire experiment leaving 24 listeners in the analysis. Deviance measures did not differ systematically across filter conditions or experimental conditions (see Appendix E). The confidence intervals and their overlap around midpoints did not differ systematically based on filter conditions and experimental conditions (see Appendix F).

SCEs

A one-way ANOVA (3 Levels: High F1 variability, Low F1 variability, Single talker) was conducted to assess whether SCE magnitude differed by condition. There were no significant differences between groups ($F(2,46) = 0.51$, $p = 0.60$, $\eta_p^2 = 0.02$; see Figure 9).

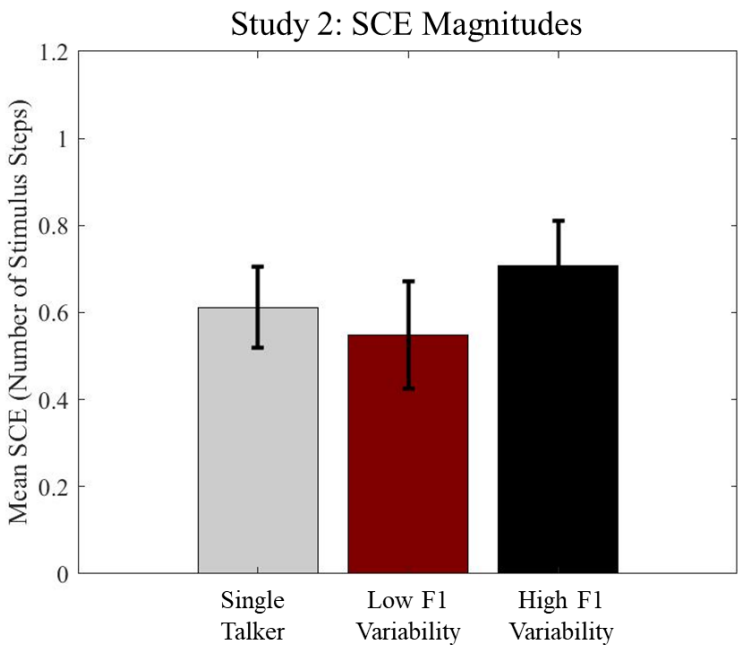

Figure 9. Contrast effect magnitudes from Study 2. The variability groups are represented on the x-axis. Contrast effect magnitude, as measured as number of stimulus steps, is represented on the y-axis. The gray bar represents the single talker condition, the maroon bar represents the low F1 variability condition and the black bar represents the high F1 variability condition. Error bars depict standard error of the mean.

Response Times

As previous reported, data from 24 listeners passed criterion from Study 2 and were included in the response time analysis. A 3 (Condition: high F1 variability, low F1 variability, single talker) X 10 (Vowel Target) repeated measures ANOVA was conducted on response times collected during Study 2 (see Figure 10). The assumption of sphericity was violated for all main effects and the interaction so a Greenhouse-Geisser correction is reported (all p's < 0.001). The ANOVA indicated a significant effect of condition ($F(1.279,29.421) = 8.370$, $p = 0.004$, $\eta_p^2 = 0.267$). Bonferroni corrected post-hoc pairwise t-tests indicated that the single talker condition ($M = 806.108$, $SE = 21.530$) elicited faster response times compared to the high F1 variability condition ($M = 892.395$, $SE = 28.670$, $p = 0.046$) and the low F1 variability condition ($M = 938.045$, $SE = 34.134$,

$p < 0.001$). Response times in the high F1 variability condition did not differ from the response times in the low F1 variability condition ($p = 0.87$). The ANOVA also indicated a significant main effect of vowel ($F(4.174,95.994) = 18.717$, $p < 0.001$, $\eta_p^2 = 0.449$). The main effect of vowel was driven by faster response times to the endpoints relative to the mid-continuum vowels, as was expected (see Appendix G for pairwise t-tests with Bonferroni corrections). There was no significant interaction ($F(8,183.90) = 0.513$, $p = 0.85$, $\eta_p^2 = 0.02$).

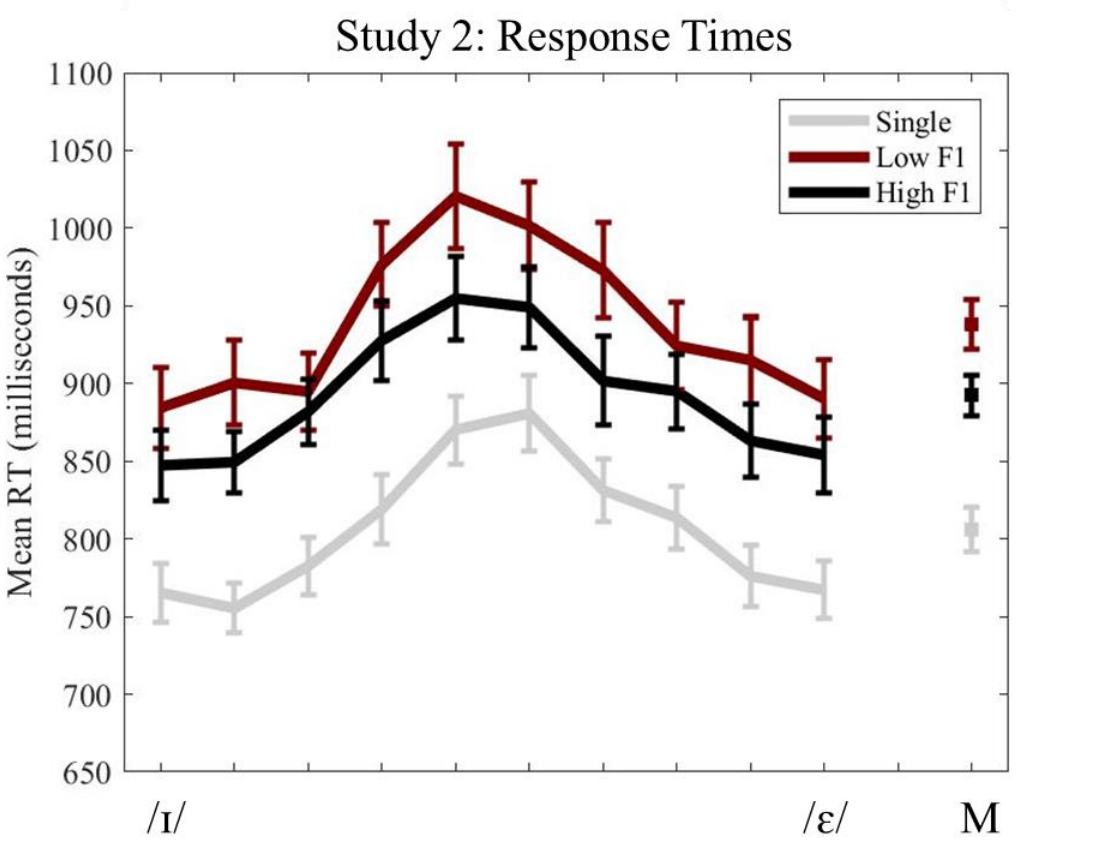

Figure 10. Response times by condition for Study 2. The vowel continuum is represented along the x-axis with the /ɪ/ endpoint on the left and the /ɛ/ endpoint on the right. Response times in milliseconds are represented on the y-axis. Average response times collapsed across vowels are shown on the far right. The black line represents the high F1 variability condition. The maroon line represents the low F1 variability condition. The

gray line represents the single talker condition. Error bars depict standard error of the mean.

Accuracy

As previous reported, data from 23 listeners were included in the analyses for Study 2. A one-way repeated measures ANOVA of condition (high F1 variability, low F1 variability, single talker) indicated that accuracy did not vary as a function of condition, ($F(2,44) = 1.158$, $p = 0.32$, $\eta_p^2 = 0.05$). Plots of accuracy can be found in Appendix P.

Discussion

The results from Study 2 suggest F1 variability in the context sentences does not influence SCEs. The listeners in this study responded to vowels primarily differentiated based on F1. Therefore, if listeners are capable of telling the vowels apart (which they are based on our inclusion criterion) they must be relying on F1 to make that distinction. If the influence of f0 variability observed in Study 1 was simply due to acoustic variability of any kind, F1 variability should have also influenced SCEs. Since contrast effects were equivalent in all groups, it is clear that not all sources of acoustic variability in the context sentences influence context effects in vowel categorization.

The response time results from Study 2 suggest that F1 variability also does not influence response times. Both conditions with multiple talkers produced slower response times relative to the single talker condition. The pattern is what is typically reported in talker normalization where any increase in the number of talkers elicits slower response time. Since the multiple talker conditions did not significantly differ from each other there was no influence of F1 variability on response times. The lack of a significant interaction suggests that response times increased relatively equally across the entire vowel continuum, following the first predicted pattern in Figure 2 (upper left panel).

The choice to rearrange sentences from Study 1b allowed for anecdotal observations of how f0 and F1 may relate to each other. In Study 1b, sentences were group based on mean f0 across the sentence. If F1 and f0 were closely related, then sentences with high f0 should also have high F1, and the high F1 variability group in Study 2 should contain the same sentences as the high f0 variability group in Study 1b. While measures of mean F1 and mean f0 are related ($r = 0.48$, $p < .001$), the groups in Study 2 did not resemble the groups in Study 1b. In fact, group assignment from the high f0 variability condition to the high F1 variability condition was at about chance (19 out of 40 sentences in the high f0 variability condition were in the high F1 variability condition). So, while f0 and F1 may share covariance, it does not appear that they are strongly related. It is possible that the ways in which f0 and F1 are used in speech perception could explain why f0 influences SCEs but F1 did not. F1 and f0 serve different purposes in speech perception. As previously mentioned, f0 is generally considered a cue to talker identity and is dictated by the vibration of the vocal folds. F1, on the other hand, is considered a primary cue to vowel identity and is related to jaw height. The differences in how these cues inform speech perception could explain why f0 influences SCEs but F1 does not.

In addition, the extent to which these cues vary in natural speech differs. There are multiple ways to quantify acoustic variability in speech. The approach taken here is to measure acoustic variability across speech tokens (e.g., trial-to-trial variability in average f0 across each sentences). This approach tends to characterize stable properties of a talker and can correspond to a talker's typical pitch and formant range. Another approach is to characterize acoustic variability within a speech token. This approach measures how

formants and pitch fluctuate when a person is talking. These measures can correspond to changes in intonation, semantics, prosody, etc. Listeners encounter both types of variability when perceiving speech and it is possible that how a cue varies within a sentence can influence how a listener responds to this cue's variability across sentences. Within a sentence, F1 is inherently more variable than f0 since talkers must change F1 to produce different speech sounds. When a phoneme changes, F1 will likely change with it. In an attempt to quantify the variability of F1 and f0 across many sentences, the average F1 and f0 was calculated for every sentence in TIMIT. The average standard deviation of f0 within a sentence for all of TIMIT is 21.38 Hz. The average standard deviation of F1 within a sentence for all of TIMIT is 127.76 Hz. It is possible that the higher within-sentence variability of F1 corresponds to lower sensitivity to F1 changes across sentences. It is also possible that since listeners have more experience with F1 variability, they are less sensitive to this cue overall and essentially weight trial-to-trial F1 variability less than f0 variability.

Study 1 and 2 Synthesis

It has been established that SCEs are linear in nature: when the size of the change between the context and target increases, so does the shift in categorization (Stilp et al., 2015; Stilp & Alexander, 2016; Stilp & Assgari, 2017a). In these studies, the magnitude of the change from context to target to sentences was varied. The resulting shifts in categorization also varied. Relating the magnitude of change (from context to target) to the size of the categorization shift revealed that they were highly related (e.g., $r = 0.88$, $p < 0.005$ in Stilp et al., 2015 [when filters were similar to those reported here]). This finding suggests that quantifying the magnitude of the contrast effect as a function of

other variables may reveal meaningful relationships. In Study 1 and Study 2, we have observed that the amount of f0 variability in a condition influences the size of the contrast effect. In Studies 1 and 2, f0 variability was treated as a binary variable (low f0 variability versus high f0 variability) but is actually continuous in nature. Since f0 variability can be treated as a continuous measure, it is possible to quantify the relationship between f0 variability and SCEs. Measures of f0 variability from Studies 1 and 2 (as well as preliminary data) were used to predict SCE magnitudes via a linear regression. There is a negative relationship ($r = – 0.63$, $p = 0.02$) between total mean f0 variability in a condition and SCE magnitude such that more f0 variability in a condition resulted in smaller SCE magnitude (see Figure 11). It is interesting to note that the preliminary contrast effect magnitudes observed in Study 2, which were grouped based on F1 variability, were well predicted by this linear regression when the f0 variability in those groups is taken into account. When the study was rerun to collect response times, the SCEs were larger than what is typically expected with the high F1 variability condition producing a SCE of 0.7 stimulus steps. This occurred despite the stimuli presented to listeners being the same. While a SCE magnitude of 0.7 stimulus steps is not outside of the range of SCEs observed with +5 dB peaks, it is larger than SCEs observed in the single talker conditions in the present studies. Since f0 was still variable in the high F1 variability condition, it is unusual that the high F1 variability condition would produce a contrast effect larger than a single talker condition. However, there were slight differences between the two data collections periods. These differences and their possible ramification are addressed in the general discussion.

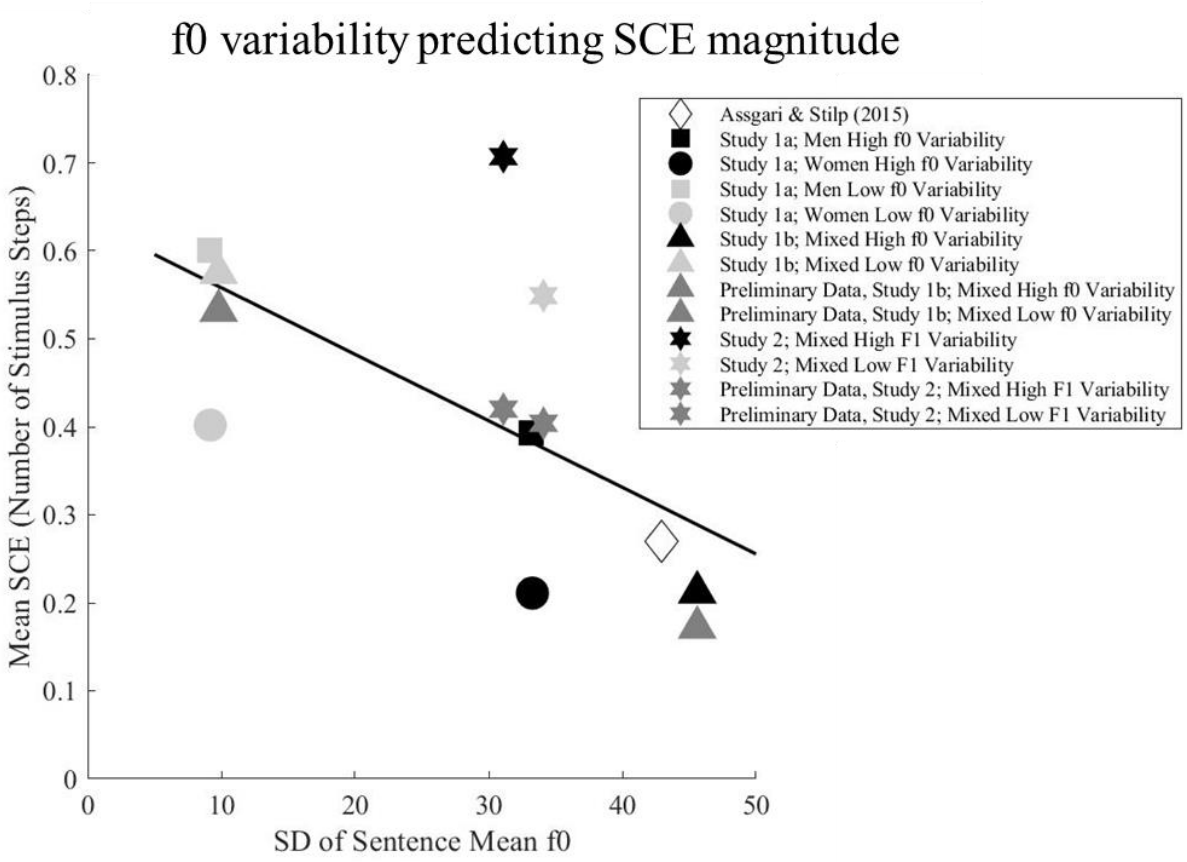

Figure 11. Predicting contrast effect magnitude from standard deviation of average f0 across a sentence within a condition. Black symbols represent high variability conditions. Light gray symbols represent low variability conditions. Dark gray symbols represent preliminary data. Squares represent men from Study 1a. Circles represent women in Study 1a. Triangles represent data from Study 1b (Mixed gender, f0 variability). Stars represent data from study 2 (F1 variability).

Even though results from Study 2 suggest that F1 does not influence SCEs, a similar linear regression was conducted to predict SCE magnitude from F1 variability. It is possible that a relationship between F1 and SCE would be illuminated if F1 measures from more than two groups were used to predict SCEs. As such, F1 measures from all groups in Studies 1 and 2 (as well as preliminary data) were obtained and entered into the linear regression. The results of the linear regression confirm that F1 variability in a condition is not a good predictor of SCE magnitude ($r = -0.09$, $p = 0.77$; see Figure 12)

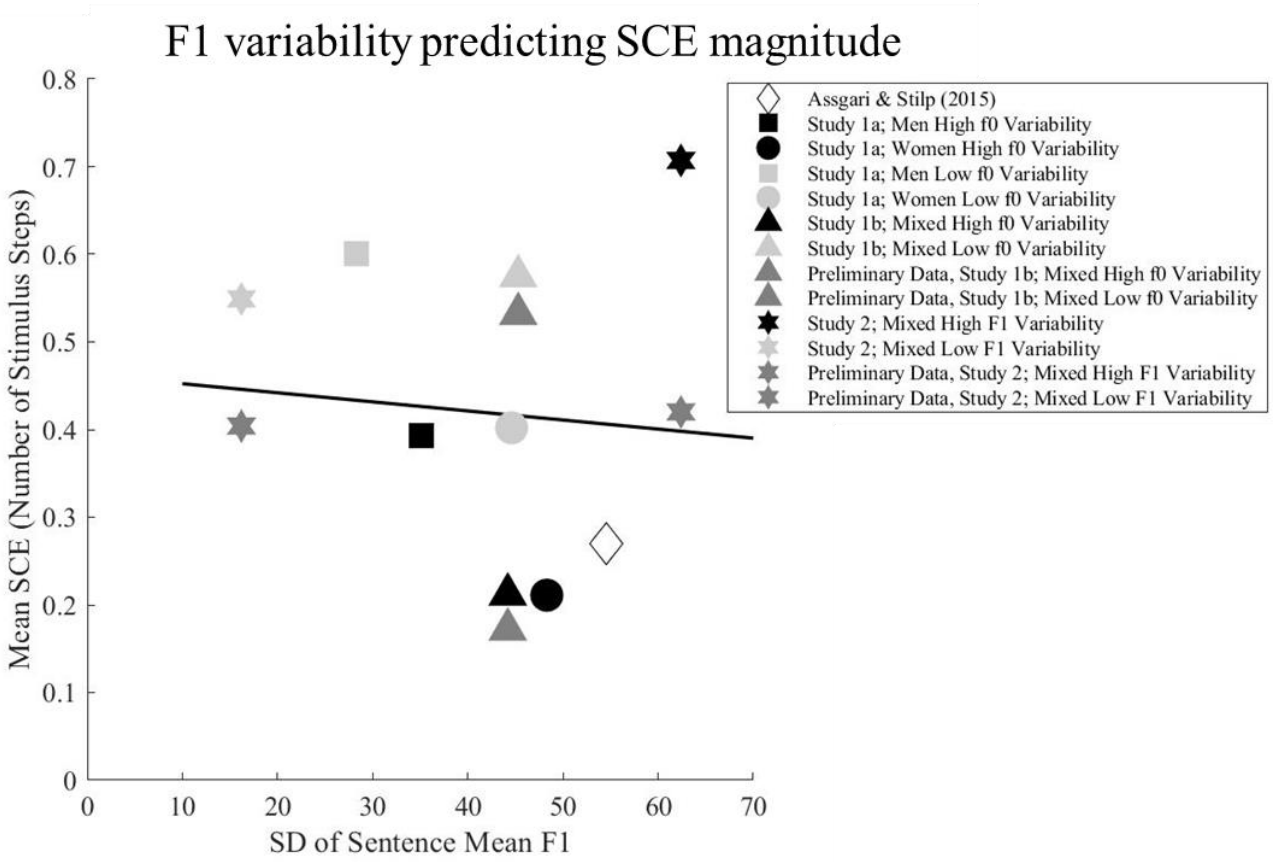

Figure 12. Predicting contrast effect magnitude from standard deviation of average F1 across a sentence within a condition. Black symbols represent high variability conditions. Light gray symbols represent low variability conditions. Dark gray symbols represent preliminary data. Squares represent men from Study 1a. Circles represent women in Study 1a. Triangles represent data from Study 1b (Mixed gender, f0 variability). Stars represent data from Study 2 (F1 variability).

The results of these regressions confirm the combined results of Studies 1 and 2: f0 variability influences SCEs in vowel categorization and F1 variability does not. It is even clearer that not all sources of acoustic variability influence SCEs. Based on this result, the remaining studies focus on how more fine-grained manipulations of f0 influence SCEs and how measures collected in these studies may suggest a connection between SCEs and talker normalization.

CHAPTER V

STUDY 3 (MANIPULATED F0)

Aims

In order to establish a stronger connection between f0 variability and smaller shifts in categorization, f0 variability can be acoustically manipulated. By acoustically manipulating f0, the same stimuli can be presented in each condition, but differ only on f0. By using the same stimuli in each condition, any differences in the size of SCEs can be directly attributed to the f0 manipulation. In Study 3, f0 of sentences was manipulated to reduce f0 variability in the sentences used in the high f0 variability of Study 1b.

Methods

Praat was used to manipulate f0 contours (Boersma & Weenink, 2017). Sentences were the same as those used in Study 1b, high f0 variability condition. In each condition, the f0 of each sentence was set to the grand average f0 for all sentences. In the f0-shifted condition, within-sentence f0 variability remained intact, maintaining natural f0 contours. In the f0-flattened condition, f0 contours were flattened and set at the grand average. In a third condition, f0 was not manipulated. This condition served as a within-subject comparison of low f0 variability versus high f0 variability, similar to other studies reported here. In the final condition, the context sentence was one sentence from a single talker (previously used in Assgari & Stilp, 2015).

Hypotheses

If f0 variability in a condition is responsible for smaller SCEs, then removing between-sentence f0 variability should produce SCEs that are at least comparable to, if not larger than, those in low-f0-variability conditions. In addition, if our manipulated f0 conditions produce SCEs similar to low variability conditions, this will establish a strong relationship between f0 variability in context sentences and the size of SCEs. If our manipulated conditions fail to produce SCEs similar to low variability conditions, then it is possible that other types of acoustic variability also influence SCEs.

Results

A total of 25 listeners participated in Study 3. Two participants were removed due to failing to maintain 80% accuracy over all endpoints, leaving 23 listeners in the analysis. In addition, upon inspection of the distribution of SCEs, one outlier was identified in the no manipulation condition. This participant showed a contrast effect of -4.33 stimulus steps, which is not only beyond the magnitude of what would be expected with +5 dB peaks in the context sentences but is also opposite of the predicted direction. This outlier was removed and the analysis was conducted on the remaining 22 listeners. Deviance measures did not differ systematically across filter conditions or experimental conditions (see Appendix H). The confidence intervals and their overlap around midpoints did not differ systematically based on filter conditions and experimental conditions (see Appendix I).

SCEs

A repeated measures ANOVA was conducted comparing mean SCEs across experimental conditions. The assumption of sphericity was violated as evidenced by a significant Mauchly's test of sphericity ($p = 0.021$). Thus, the ANOVA reported here is

reported with the Greenhouse-Geisser correction. The ANOVA revealed no significant differences between the means ($F(2.08, 43.75) = 1.870$, $p = 0.165$, $\eta_p^2 = 0.082$; see Figure 13).

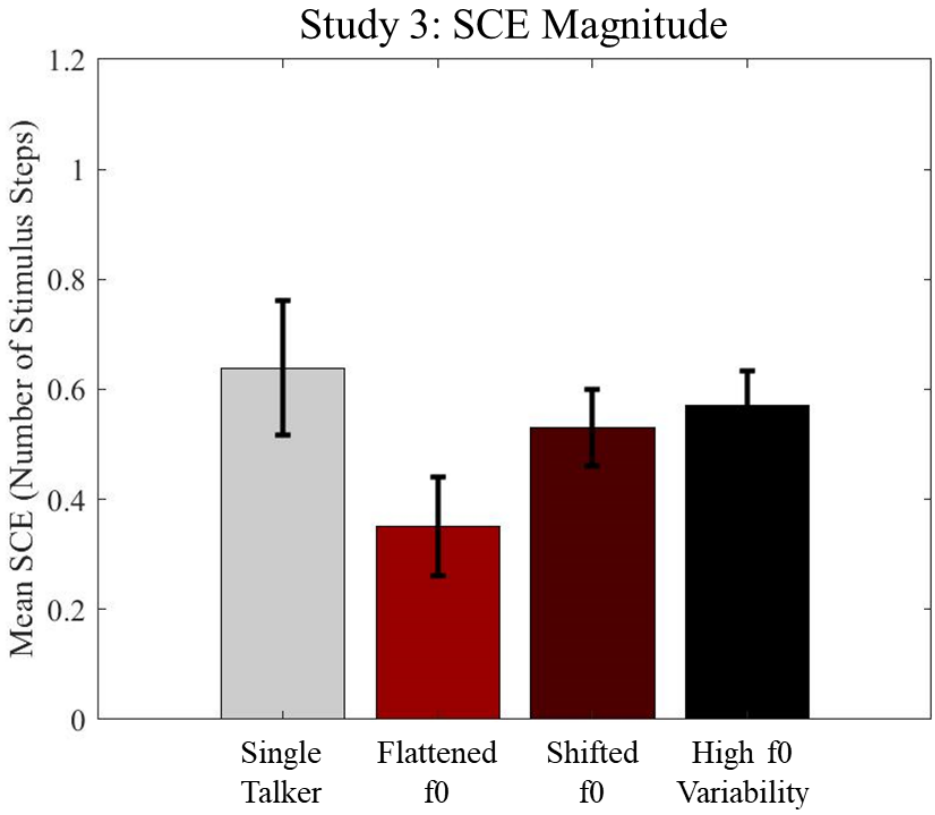

Figure 13. Contrast effect magnitudes from Study 3. Manipulated f0 conditions are along the x-axis and contrast effect magnitide is along the y-axis. The gray bar corresponds to the single talker condition, the light red bar corresponds to the flattened f0 condition, the dark red bar corresponds to the shifted f0 condition, the black bar corresponds to the high f0 variability condition. Error bars depict standard error of the mean.

Response Times

As previous reported, data from 22 listeners passed criterion from Study 3 and were included in the response time analysis. A 4 (Condition: no manipulation, flattened f0, shifted f0, single talker) X 10 (Vowel Target) repeated measures ANOVA was conducted on response times collected during Study 3 (see Figure 14). Mauchly's test of

sphericity was violated for vowel ($p < 0.001$). As such, a Greenhouse-Geisser correction is reported for the main effect of vowel. A significant main effect of condition was found ($F(3,63) = 3.04$, $p = 0.04$, $\eta_p^2 = 0.13$). This main effect of condition would be primarily driven by the difference in response times between the single ($M = 868.39$, $SE = 27.16$) and the flattened f0 condition ($M = 932.75$, $SE = 25.58$; $p = .06$), however, post hoc pairwise t-tests with Bonferroni corrections showed no significant differences (all p's > 0.06). A significant main effect of vowel was observed, $F(3.56,74.66) = 19.26$, $p < 0.001$, $\eta_p^2 = 0.48$). These results are primarily driven by the increase in response times to the ambiguous target vowels in the middle of the continuum (see Appendix J for pairwise t-tests with Bonferroni corrections). The interaction between vowel and condition was not significant ($F(10.20,214.26) = 0.98$, $p = 0.47$, $\eta_p^2 = 0.04$).

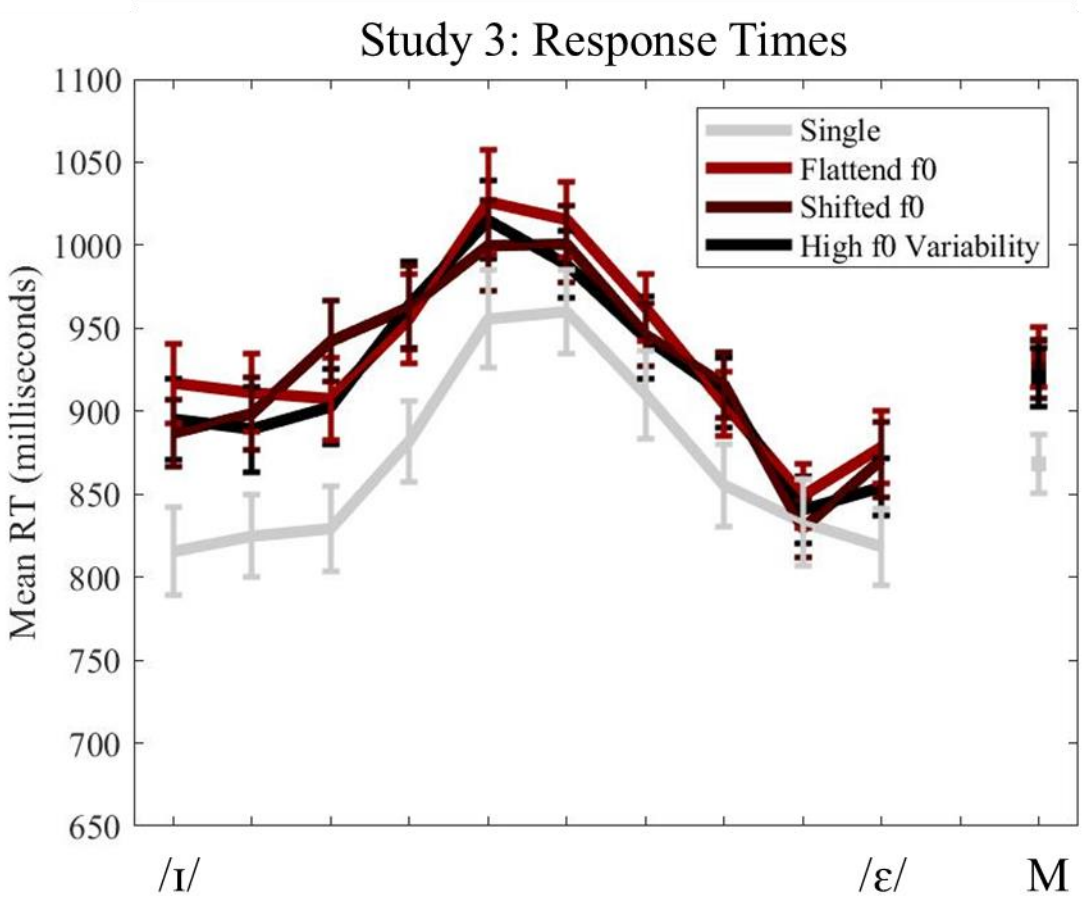

Figure 14. Response times by condition for Study 3. The vowel continuum is represented along the x-axis with the /ɪ/ endpoint on the left and the /ɛ/ endpoint on the right. Response times in milliseconds are represented on the y-axis. Average response times collapsed across vowels are shown on the far right. The black line represents the high f0 variability condition. The dark red line represents the shifted f0 condition. The light red line represents the flattened f0 condition. The gray line represents the single talker condition. Error bars depict standard error of the mean.

Accuracy

As previous reported, data from 22 listeners were included in the analyses for Study 3. A one-way repeated measures ANOVA of condition (no manipulation, flattened f0, shifted f0, single talker) indicated that accuracy did not vary as a function of condition, ($F(3,63) = 0.624$, $p = 0.602$, $\eta_p^2 = 0.03$). Plots of accuracy can be found in Appendix P.

Discussion

The magnitudes of SCEs in Study 3 indicate several issues. First, it is questionable whether the manipulations of f0 were successful in creating talkers that sounded similar. While the SCE in the single condition was the largest (M = 0.64, SE = 0.12), as predicted, the contrast effects in the shifted (M = 0.53, SE = 0.07) and flattened (M = 0.35, SE = 0.09) f0 conditions appeared smaller, although not significantly so. If the f0 manipulations were successful, the f0 variability in the shifted and flattened f0 conditions would have been equivalent to single talker condition (i.e., no f0 variability across trials). Thus, these conditions should have all produced large contrast effects. Even though the SCE magnitudes appear to differ in these conditions, a non-significant ANOVA supports that the hypothesis that shifted f0 and flattened f0 show similar contrast effect to the single-talker condition. A main effect of condition for response times suggested that response times may differ depending on condition. However, following corrected post hoc pairwise t-tests, there was no significant difference between the groups. This result provides further evidence that the manipulation of f0 may not have been successful. However, it is puzzling that the single talker condition also did not differ from the manipulated f0 conditions. It is possible that high variability in each of the conditions led to this result. Inspection of Figure 14 suggests that if differences between the condition had been significant, they would be most like the predicted pattern in the upper left panel of Figure 2, with response times increasing across the entire vowel continuum.

Second, the high f0 variability condition showed contrast effects much larger than what would be expected. These stimuli are the same stimuli as in Study 1b. In Study 1b, a contrast effect of 0.21 stimulus steps was observed with these stimuli while in this study a

contrast effect of 0.57 was observed. Thus, this condition failed to replicate past results. While it is not entirely clear why this condition failed to replicate, there are a few suggestions that deserve consideration. First, different listeners participated in Study 1b compared to the current study. Individual differences in contrast effect magnitude are relatively common and could explain why the SCE magnitude for this condition was larger in this study relative to Study 1b. However, high f0 variability producing the same size contrast effect as the single talker condition suggests that f0 variability had no influence on SCEs in these participants, which is problematic. Second, it is possible that this failure to replicate occurred because the alternate conditions presented in this study differed from those presented in Study 1b. In Study 1b, a single talker condition and a low f0 variability condition were presented alongside the high f0 variability. As such, there was no overlap in sentences across conditions. In the current study, listeners heard the same sentences in all the multi-talker conditions. It is possible that this allowed listeners to become familiar with the sentences and the talkers despite the manipulations of f0. Why familiarity with sentences may influence contrast effects is addressed further in the general discussion.

While acoustic manipulations of speech are ideal for experimental control, these manipulations can sometimes have unintended consequences in the speech signal. As previously discussed, speech is a combination of the source and formant filters. The source is a complex sound that is comprised of a fundamental frequency and harmonics that fall at integer multiples of that fundamental frequency. For example, if a source has a fundamental frequency of 100 Hz, the harmonics will fall roughly at 200, 300, 400, 500 Hz, etc. Shifting the fundamental frequency up to 200 Hz would result in harmonics that

now fall at 400, 600, 800, 1000 Hz, etc. The formant filter contains peaks that correspond to resonances of the vocal tract. In order to produce a speech sound, the source is pushed through the formant filters. The degree to which the peaks in the formant filter are expressed depends on if there is harmonic energy near those peaks. Thus, the amplitude of the harmonics in the final speech signal depend on where the formants filter peaks fall. If a harmonic falls directly under a formant peak, this harmonic will be higher in amplitude than a harmonic that falls farther from a formant peak. Thus, the distribution of energy in the final speech sound is closely tied to both f0 and the formant filter. If one or both changes, there will be consequences for the distribution of energy in the sound. If the formant filters remain the same, the relationship between the harmonics and the formant filters will change causing the overall distribution of energy in the signal to change as well.

The method of manipulating f0 changed the fundamental frequency and its corresponding harmonics while attempting to maintain the formant filters. This had consequences for the overall distribution of energy in the speech signal. SCEs are argued to be driven by the distribution of energy in the frequency region that differentiates the target phonemes, here the F1 region. Shifting energy through manipulations could have had unintended consequences on SCEs. To explore this possibility, the distribution of energy in the F1 region was measured and compared in the multi-talker conditions (Stilp & Assgari, 2017b). The energy in the low-F1 and high-F1 regions was measured in a single sentence using two band pass filters (low-F1: 100-400 Hz; high-F1: 550-850 Hz) with a 5-Hz transition between the passband and the stopband. Filters were made using the fir2 command in Matlab with 1000 coefficients. The amplitude envelope of both

regions was obtained by rectifying the signal and low-pass filtering using a 2nd order Butterworth filter with a 30 Hz cutoff frequency. The root-mean-square of each envelope was converted to dB. A single measure from each sentence was obtained by subtracting the energy in high-F1 region from the energy in the low-F1 regions. This difference is referred to at the Mean Spectral Difference (MSD). A negative MSD indicates relatively more energy in the high-F1 region than the low-F1 region in the sentence. Importantly, MSD is argued to be vital for producing SCEs (Stilp & Assgari 2017b). To evaluate whether manipulating f0 changed the MSDs in the current study, average MSDs across all sentences were compared for three of the conditions in Study 3: no manipulation, shifted f0 and flattened f0. Since the same sentences were used in each of these conditions, any differences in MSDs is a direct result of f0 manipulation.

The first measure obtained was the average MSD when comparing all low-F1-amplified sentences to all high-F1-amplified sentences in each condition. In the current studies, adding + 5 dB spectral peaks targeted an overall MSD of approx. 10 dB. This ensured that the overall MSD was roughly equivalent for all groups (i.e., Overall MSD in Table 1). However, when overall MSD was broken down further and the MSD for low-F1 and high-F1 sentences was measured individually, rather than averaged together, there was a clear effect of f0 manipulation. First, MSDs in sentences clearly shift depending on the f0 manipulation being conducted (see Table 1). This demonstrates that while the relative differences between low F1 and high F1 regions were preserved, the spectral composition within sentences are being changed in unintended ways by the f0 manipulations. Shifting f0 pushes a little bit more energy toward low F1 region (more positive MSD), and flattening f0 is pushing a lot more energy toward low F1 region (a lot

more positive MSD). Further, the variability of MSDs increased as a result of manipulations (SD of Overall MSD in Table 1). It is possible that increasing variability of MSDs in manipulated f0 conditions was inadvertently decreasing SCEs since MSDs were changing more from trial-to-trial. However, if increased MSD variability accounted for all differences in SCEs, then both shifted and flattened f0 should have smaller contrast effect, which was not the case. Therefore, MSD variability cannot fully capture the consequences that our manipulations had on SCEs. These changes in the distribution of energy in the F1 regions because of manipulation were unforeseen and could not be controlled.

Table 1.

Mean spectral differences in multi talker conditions in Study 3

	MSD of Low F1 Sentences	MSD of High F1 Sentences	Overall MSD (Low F1 – High F1)	SD of Overall MSDs
Not Manipulated	1.73 dB	-7.38 dB	9.11 dB	2.47 dB
Shifted f0	2.59 dB	-6.61 dB	9.20 dB	3.38 dB
Flattened f0	4.67 dB	-4.45 dB	9.12 dB	3.38 dB

Finally, as previously mentioned, talkers differ acoustically on more parameters than just f0. Previous research has demonstrated that when a cue is no longer informative for a phoneme distinction, listeners will decrease reliance on that cue in favor of a more informative cue (e.g., Stilp, Anderson, Assgari, Ellis & Zahorik, 2016; Stilp & Anderson, 2014). It is possible that in the absence of f0 variability, listeners can use other available

cues to differentiate between talkers. While Study 2 suggests that F1 variability alone is not enough to influence SCE magnitude, there are other acoustic parameters that are argued to cue talker changes. For example, F3 has been argued to be related to vocal tract length (Johnson, 2005). Vocal tract is a stable property of a talker and differences in vocal tract length may cue listeners to changes in talkers. If listeners are able to exploit F3 in the absence of the f0, it is possible that talker information related to F3 is influencing SCEs. Again, this discussion point will be revisited in further detail in the general discussion.

Despite the unintended consequences that the manipulations may have had on the stimuli, using Praat to manipulate pitch was still the preferred method. During the design phase of this study, several other methods were explored with far less optimal results. For example, one alternative method used Praat to decompose the sound into the source and filter components. The source was manipulated independently to adjust f0. The manipulated source and original filter were recombined in an attempt to reproduce the original sound with a different f0. In theory, this manipulation should have allowed for an independent manipulation of f0 without influencing the formants. However, the resulting speech sounds were quite terrible. First, the sounds were no longer intelligible and, while intelligibility is not necessary to observe contrast effects, this pointed to larger issues in the manipulation. Particularly, it appeared that the formant filter was not extracted cleanly and there was evidence that source material was still present in the filter component and vice-versa. Second, even if this method had worked, much of the unintended consequences regarding the redistribution of energy would have still been an issue. Other widely used methods (e.g., STRAIGHT; Kawahara, Takahashi, Morise, &

Banno, 2009) were also explored but these methods tend to assume that manipulations of f0 reflect the desire to switch the gender of the talker. As such, these methods also shift formant frequencies, which was not the goal in these studies.

CHAPTER VI

STUDY 4 (ORDERED F0)

Aims

Results from Study 1 make a strong case for f0 variability influencing SCEs in natural speech. However, in Study 1, both global (i.e., total variability in a condition) and local (i.e., trial-to-trial variability) variability were high. Thus, it was unclear whether global or local variability drives the influence of f0 variability on SCEs. Study 4 investigated whether local or global variability influences SCEs by manipulating local (trial-to-trial) variability in the mean f0 of context sentences.

Methods

The same stimuli used in Study 1b, High f0 Variability condition were presented in the order of ascending or descending f0. Measures of f0 have already been obtained in Study 1b. In a third condition, stimuli were arranged so that there is the maximum local f0 variability. With regard to the tails of the distribution shown in Figure 4, in the maximum local variability condition, the order of presentation was as follows: the sentence with the lowest f0 from the lower tail of the distribution, the sentence with the lowest f0 in the upper tail of the distribution, the sentence with the second-lowest f0 of the lower tail, the sentence with the second-lowest f0 in the higher tail, and so on. The Matlab script was edited to control the order of stimulus presentation.

Hypotheses

If trial-to-trial variability in average f0 of context sentences influences SCEs, then ascending and descending conditions should produce larger SCEs because trial-to-trial variability has been minimized. In addition, the maximum variability condition should show the smallest SCE. On the other hand, if global variability is more influential than local variability, then all multi-talker conditions should have the same size SCE because global variability is matched across conditions. Further, since global variability is high in all conditions, SCEs should be smaller than single talker conditions.

Results

Twenty-five listeners participated in Study 4. One listener failed practice and two were removed for failing to maintain 80% accuracy across all endpoints, leaving a total of 22 listeners in the analysis. In the exploration of the data, an outlier was identified in the descending group. This participant exhibited a contrast effect with the magnitude of 6.93 stimulus steps which is, once again, beyond what would be expected with +5 dB peaks. This outlier was removed and the analysis was conducted with the remaining 21 listeners. Deviance measures did not differ systematically across filter conditions or experimental conditions (see Appendix K). The confidence intervals and their overlap around midpoints did not differ systematically based on filter conditions and experimental conditions (see Appendix L).

SCEs

A repeated measures ANOVA was conducted to test whether mean SCEs are different based on condition. The ANOVA revealed no significant difference between the mean SCEs based on condition, ($F(3, 60) = 0.908$, $p = 0.442$, $\eta_p^2 = 0.043$; see Figure 15).

It is possible that this omnibus ANOVA was underpowered considering that 3 out of 4 group means were predicted to be equal to each other. As a way to probe further into this claim, one-sample t-tests comparing contrast effect magnitudes in Study 4 against 0 were conducted. If one-sample t-tests are not significant, it indicates that no contrast effect was observed in that condition. Results of 4 one-sample t-tests reveal a significant contrast effect in the single (M = 0.386, SE = 0.12), ascending (M = 0.400, SE = 0.12), and descending (M = 0.350, SE = 0.10) conditions (all p's < 0.003, Bonferroni corrected $\alpha = 0.0125$). However, the contrast effect in the maximum condition (M = 0.176, SE = 0.10) does not differ significantly from 0, $t(20) = 1.697$, $p = 0.105$.

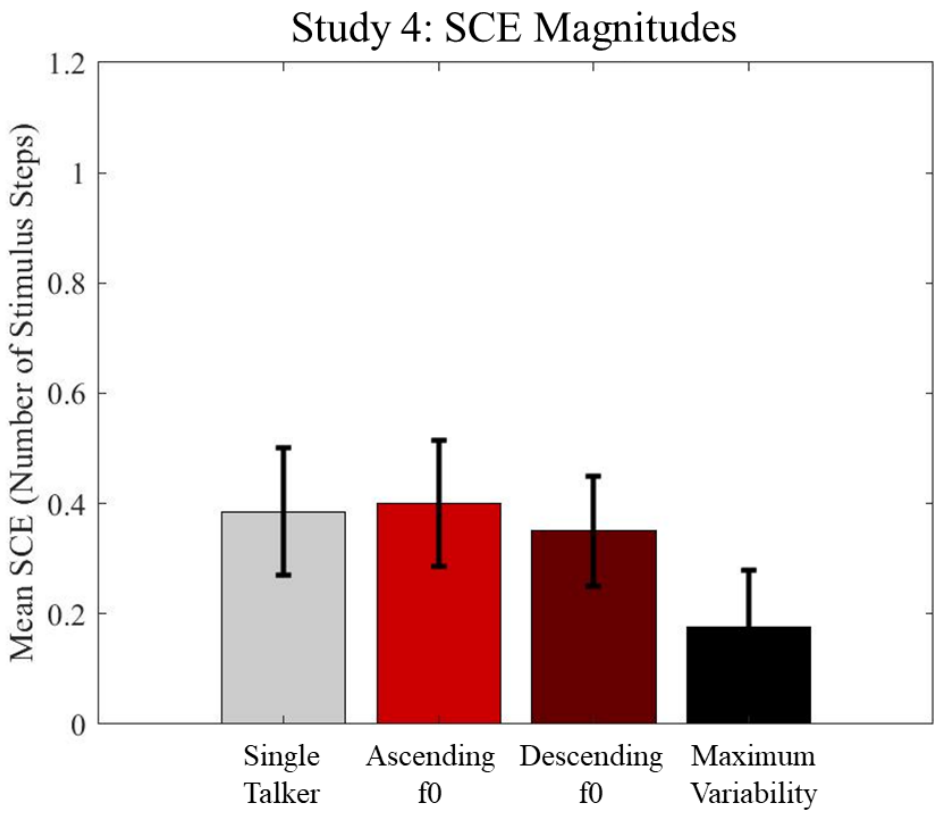

Figure 15. Contrast effect magnitudes from Study 4. Ordered f0 conditions are along the x-axis. Contrast effect magnitide is along the y-axis. The gray bar corresponds to the single talker condition, the light red bar corresponds to the ascending f0 condition,

the dark red bar corresponds to the descending f0 condition, the black bar corresponds to the maximum f0 variability condition. Error bars depict standard error of the mean.

Response Times

As previously reported, 21 participants had passed criterion in Study 4 and were included in the response time analysis. A 4 (Condition: single talker, ascending f0, descending f0, and maximum f0) X 10 (Vowel Target) repeated measures ANOVA was conducted on response time data collected during Study 4 (see Figure 16). Mauchly's test of sphericity indicated that sphericity was violated for vowel, ($p < 0.001$). As such, the main effect and interaction are reported with a Greenhouse-Geisser correction. A main effect of condition was significant, ($F(3, 60) = 3.74$, $p = 0.016$, $\eta_p^2 = 0.157$). Bonferroni corrected post hoc pairwise t-tests indicate that participants are significantly slower in the maximum variability condition relative to the single talker condition ($p = 0.046$) and the descending condition ($p = 0.04$). The ascending condition did not differ from the maximum variability condition ($p = 0.30$). The main effect of vowel was also significant, ($F(3.04,60.72) = 13.77$, $p < 0.001$, $\eta_p^2 = 0.41$). Again, this main effect was primarily driven by increases in response times in the middle of our vowel continuum where stimuli are most ambiguous (see Appendix M for pairwise t-tests with Bonferroni corrections). The interaction between condition and vowel was significant ($F(27, 540) = 4.24$, $p < 0.001$, $\eta_p^2 = 0.18$). Post hoc pairwise t-tests suggest that this result is driven primarily by the increase in response times observed in the ascending and maximum conditions on the /ɪ/ end of the continuum (see Appendix F for pairwise t-tests with Bonferroni corrections).

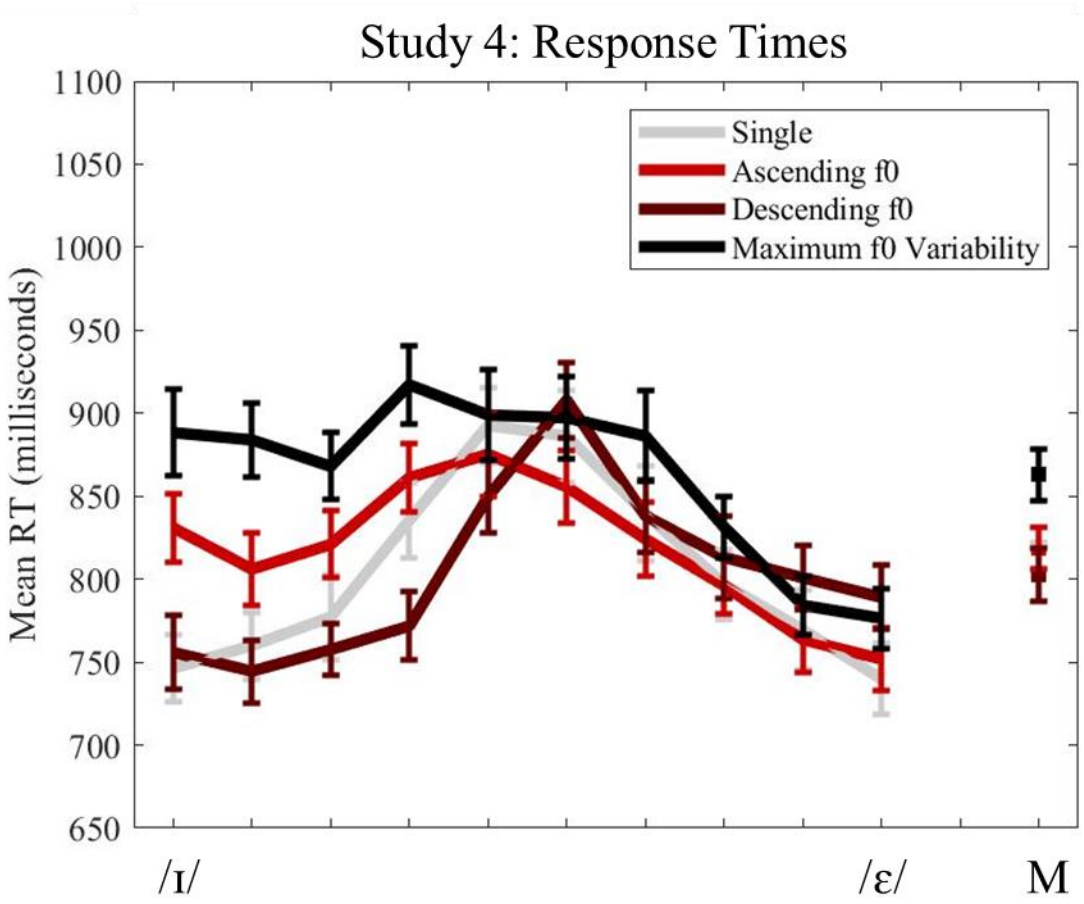

Figure 16. Response times by condition for Study 4. The vowel continuum is represented along the x-axis with the /ɪ/ endpoint on the left and the /ɛ/ endpoint on the right. Response times in milliseconds are represented on the y-axis. Average response times collapsed across vowels are shown on the far right. The black line represents the maximum f0 variability condition. The dark red line represents the descending f0 condition. The light red line represents the ascending f0 condition. The gray line represents the single talker condition. Error bars depict standard error of the mean.

Accuracy

As previous reported, data from 21 listeners were included in the analyses of Study 4. A one-way repeated measures ANOVA of condition (single talker, ascending f0, descending f0, and maximum f0) was conducted to see if endpoint accuracy differed as a function of condition. Mauchly's test of sphericity indicated that the assumption of

sphericity was violated in Study 4 (p = 0.001). As such, the ANOVA results are reported with a Greenhouse-Geisser correction. There was no significant effect of condition on accuracy, (F(1.73,34.69) = 1.535, p = 0.230, η_p^2 = 0.071). Average accuracy in each condition can be found in Appendix P.

Discussion

In Study 4, f0 variability was controlled on a trial-by-trial basis by playing the stimuli in order of ascending f0, descending f0 or maximally alternating f0. The magnitudes of SCE in Study 4 follow the predicted pattern. The maximum variability condition showed the smallest SCE. The single, ascending, and descending conditions showed similar SCEs that were larger than the maximum variability condition. However, the differences between SCE magnitudes were not significantly different.

The response time results affirm that response times are influenced by f0 variability, with slower response times observed in the maximum f0 condition relative to the single talker condition. Once again, these differences seem to be driven primarily by larger increases in response times at the /ɪ/ end of the vowel continuum suggesting that the influence of f0 variability is greater for /ɪ/ than for /ɛ/. The pattern of response times in Study 4 do not follow any of the predicted patterns in Figure 2. Instead, response times in Study 4 increased only for the /ɪ/ endpoint vowels.

It is worth mentioning that the lack of differences found between SCEs in these studies may be because SCEs with + 5 dB peaks are already small. Since the majority of the studies reported here attempt to detect a decrease in SCE magnitude following contexts that are highly variable, starting with small SCEs limits the decrease that can be observed. Thus, in our analyses using ANOVA, conditions that fail to produce a contrast

effect do not differ significantly from conditions that do produce contrast effects. One-sample t-tests were used to determine whether listeners experienced a contrast effect in each condition. The results suggested that the maximum variability condition did not produce a contrast effect but all other conditions did. Yet, the ANOVA did not find any significant differences between any of the conditions.

Even though the maximum variability condition did not produce a contrast effect greater than zero, it is possible that smaller SCEs could be observed with a few adjustments. Using the same stimuli as in the Study 1b high variability condition meant that only 40 talkers were used in the all the multi-talker conditions reported here. By reusing these stimuli, it was possible to compare the effect of ordering sentences by f0 with conditions where the same sentences were randomized in previous studies. However, only using 40 talkers meant that each talker was repeated 4 times. To ensure maximum variability from trial-to-trial, the lowest-pitched man (e.g. Sentence 1) was followed by the lowest-pitched woman (e.g., Sentence 2). This order was repeated three additional times (e.g., 1, 2, 1, 2, 1, 2, 1, 2) before moving on to the second lowest man followed by the second lowest woman (e.g., 3, 4, 3, 4, 3, 4, 3, 4). Repetition of trails also occurs in the ascending and descending conditions. Each talker is repeated 4 times before moving to the next talker (e.g., 1, 1, 1, 1, 2, 2, 2, 2, etc.). During the repetition of trials, the order of the stimuli was predictable. It is possible that high predictability could negate or at least limit the influence of high acoustic variability. If a listener could predict what the f0 of the next talker would be, perhaps they would be less sensitive to the changes in f0. In order to eliminate predictability in the maximum variability condition, future studies should increase the number of talkers so that a new talker is heard on every trial.

In addition, the trials could be randomized so that they would not follow a low pitch then high pitch pattern. This way, no talker is familiar and the pattern is not predictable in the high f0 variability condition.

Results from Study 4 still point to the importance of trial-to-trial variability over global variability. Conditions with low trial-to-trial variability look more like single talker conditions than the maximum variability condition. Ascending and descending f0 conditions had high global f0 variability whereas the single talker condition has no global f0 variability. Yet, no significant differences were found between single talker and ascending or descending conditions in the omnibus ANOVA. In addition, contrast effects were significantly above zero when trial-to-trial variability was low, but not when trial-to-trial variability was high. In conjunction, these results suggest that high global variability did not drive down contrast effects as much as high trial-to-trial variability.

CHAPTER VII

GENERAL DISCUSSION

Overview

The context in which an object is perceived influences how that object is identified. This is true in all perceptual modalities. In audition, there are many ways that context can influence perception. Here, two well established contextual influences were tested in conjunction: talker normalization and SCEs.

Talker normalizations occurs when listeners hear different talkers and results in speech perception being slower and less accurate (e.g., Creelman, 1957; Fourcin, 1968;Assmann, Nearey, & Hogan, 1982; Geiselman & Bellezza, 1976; Mullenix, Pisoni, & Martin, 1989; Mullenix & Pisoni, 1990; Logan & Pisoni, 1987). Spectral contrast effects occur when the acoustic context of speech biases subsequent perception. A recent study suggested that these contextual effects may be related to one another by observing that SCEs were restrained when the context was spoken by different talkers (Assgari & Stilp, 2015). However, why hearing different talkers reduced contrast effects was not immediately clear.

The studies reported here served two primary purposes. The first purpose was to establish what talker characteristics restrain the influence of spectral content on speech sound categorization. The second purpose was to investigate what links may exist between SCEs and talker normalization. Both of these contextual influences appear to be

modulated by low-level acoustic variability (e.g., Assgari & Stilp, 2015 and Goldinger, 1996). The common influence of low-level acoustic variability suggests that these processes may somehow be related. The current studies attempted relate these effects by collecting measures of both SCEs and talker normalization in SCE experiments.

Recap of Results

Study 1 sought to tease apart the influence of changes in mean f0 across context sentences and talker gender. Results from Study 1 demonstrated that mean f0 variability across sentences but not talker gender influenced SCEs. This suggested that low-level acoustic variability influenced contrast effects. Study 2 assessed whether another source of acoustic variability, F1 variability, also influenced SCEs. Results demonstrated that mean F1 variability across sentences did not influence SCEs, suggesting that not all sources of low-level variability influence contrast effects. Study 3 manipulated f0 to equate mean f0 across sentences. Results suggested that manipulating f0 variability had mixed effects likely due to issues with the manipulation method. Study 4 ordered the presentation of sentences so that changes from trial-to-trial would be either minimal or maximal. Results found that increasing the trial-to-trial variability in f0 eliminated contrast effects. Response time results revealed that listeners were slower at responding to the target vowels when the contexts were highly variable but no influences on accuracy were found.

Influence of f0 variability on SCEs

Overall, in these studies when talkers' mean f0 was more variable from trial to trial, SCEs that biased vowel categorization were smaller. If SCEs are smaller in magnitude, then ambiguous vowel targets are less disambiguated by context. Thus, when

talkers sound more different, the preceding sentence context informs the perception of the subsequent vowel target to a lesser degree. The influence of f0 variability on SCEs may be adaptive. As previously mentioned, SCEs disambiguate otherwise ambiguous targets. Thus, SCEs likely serve as a way to increase the accuracy of speech perception. The influence of context should only occur when past perception is informative for current perception. When the acoustics of the sound change dramatically as it can across talkers, it is unlikely that past speech perception is still informative to the perception of current speech. If the influence of context was maintained when sounds are extremely acoustically different, using past experience to inform current speech perception could be detrimental by unreliably disambiguating sounds.

However, it is possible that when the acoustics of the talkers are similar, previous perception could still be informative. The data reported here show that, at least in the present vowel categorization task, similar-sounding talkers can be treated like a single talker regardless of gender (Study 1b). This suggests that even when hearing (acoustically similar) different talkers, previous experience with similar-sounding talkers will still inform perception. One reason this may occur is that talkers that are similar in f0 may share other similar spectral properties in their speech. When f0 is more similar between talkers, their distribution of energy in the F1 region may also be more similar. In these experiments, the low F1 region (100-400 Hz) encompasses typical f0 values. Therefore, it is possible that talkers with similar f0s in these studies will also have similar energy in the F1 regions thought to influence SCEs with these target vowels. It has been established that the long-term average spectrum (LTAS) of speech is consistent across many languages (Byrne et al., 1994). In these studies, speech from both men and women across

17 different languages was averaged and the LTAS was measured. The most consistent departures from the consistent average were based on gender. The authors reported for all of the languages surveyed, males had LTASes with a lower frequency bias than females. Particularly, these differences were observed below 250 Hz and is likely a result of f0 differences between genders (Byrne et al., 1994). As previously mentioned, when gender is treated as a binary distinction, men have lower f0s and women have higher f0s. Thus, within-gender similarity of f0 is likely greater than across-gender similarity. This suggests that talkers with similar f0s may have more similar LTASes than talkers with different f0s.

Another reason why similar sounding talkers may show larger influences on subsequent speech perception is that the listener might not realize the talker has changed. It has been demonstrated that context effects can differ based on listener expectation of the number of talkers even when the stimuli are the same (Magnuson & Nusbaum, 2007). Further, demonstrations of 'change deafness' suggests that it is possible for changes in talkers to go undetected if attention is allocated elsewhere (Vitevitch, 2003). In this study, listeners were asked to repeat words in a word list. Importantly, for some of the participants the talker producing the word lists changed about halfway through the experiment. At the end of the study, listeners were asked to answer 3 follow-up questions to gauge whether they had noticed the talker change. At least 40% of the listeners in each experiment failed to notice the talker change (Vivevitch, 2003). Neuhoff et al., (2015) demonstrated that if acoustic changes are more gradient, as in Study 4, this change is harder to detect. In this study, listeners heard continuous speech that slowly increased or decreased in pitch. Listeners were asked three follow up questions to determine if they

had noticed the pitch change. Less than half of the participants noticed any changes in pitch (Neuhoff et al., 2015). Therefore, it is possible that our listeners were not noticing talker changes especially when talkers were ordered by f0 in Study 4. However, listeners were never told to pay attention to talker and whether they noticed talker changes was not assessed.

If f0 variability can cue to the listener that past perception should not bias current perception, then perhaps other types of context effects in speech will also be restrained by f0 variability. Another type of context effect, temporal contrast, occurs when temporal cues differentiate two response options. For example, the consonants /b/ and /p/ are primarily differentiated by the duration of voice onset time (VOT). VOT is the duration of time that passed between the opening of the lips and the beginning of vocal fold vibration in stop consonants. The perception of these phonemes can be pushed around based on the speaking rate of the context (e.g., Summerfield, 1981). If the speaking rate of the context is slow, the VOT will be perceived as faster and participants will report hearing the shorter VOT option (i.e., /b/). However, if the speaking rate is fast, the same VOT will be perceived as slower and listeners will report the longer VOT option (i.e., /p/). If f0 variability cues to listeners that this past perception is no longer informative, then high f0 variability in a temporal contrast paradigm should also result in smaller contrast effects. If temporal contrast effects do not decrease with increased f0 variability across contexts, then the influence of f0 variability may be specific to spectral contrast effects.

Since differences in f0 can be related to talker changes, perhaps other acoustic cues related to talker changes would have a similar influence on SCEs. Another acoustic

parameter suggested to be related to talker changes is the third formant (F3). F3 is said to correspond to the length of the vocal tract of the individual (Johnson, 2005), a property that is relatively stable over time. If manipulating F3 variability (akin to Studies 1b and 2) produces results similar to manipulating f0 variability (i.e., high variability results in diminished SCEs), it is likely that cues to talker changes influence context effects and not merely f0 variability. However, if manipulating F3 variability fails to show similar results (as observed in Study 2 with F1), the influence of f0 on context effects may not be due to cueing listeners to changes in talkers per se, but rather acoustic ramification of varying f0.

SCEs are argued to be very low-level acoustic effects (e.g., Lotto & Holt, 2006; Sjerps and Reinisch, 2015). It has repeatedly been demonstrated that SCEs can occur with non-speech stimuli (Holt, 2005; 2006, Stilp, Alexander, Kiefte, & Kluender, 2010, Watkins, 1991; Assgari, Frazier, & Stilp, 2018) suggesting that speech is not necessary to observe SCEs. Further, it has been argued that the long-term average spectrum is one of the most important considerations when determining if an SCE will occur (e.g., Laing et al., 2012). The LTAS characterizes the distribution of energy in key frequency regions over the entire sentence. When there is a difference in energy in key frequency regions that differentiates target sounds, an SCE will be observed (e.g., Stilp, Anderson & Winn, 2015; Assgari & Stilp, 2015; Stilp & Assgari, 2017, Laing et al., 2012). In the current studies, these key frequency regions are low-F1 (100-400 Hz) and high-F1 (550-850 Hz). When the energy in these regions differ within a sentence, categorization of the target vowels /ɪ/ and /ɛ/ is biased. Research in our lab has shown the difference in these specific regions produce SCEs (Stilp & Assgari, 2017b). To quantify these differences, MSDs are

measured (see discussion of Study 3). In all studies reported here, spectral peaks were added to sentences ensuring that MSDs across sentences were the equivalent in each condition (approx. 10 dB). If average MSDs was the only parameter necessary for SCEs, SCEs should have been equivalent in all conditions. This was clearly not the case. Thus, other acoustic parameters must also influence SCEs. Here, f0 variability was manipulated and was shown to influence SCE magnitude. It is possible that f0 variability is also influencing MSDs. In the discussion of Study 3, it was shown that manipulating f0 (while all other aspects of the sentence was held constant) had clear consequences for MSDs. However, because the LTAS is determined by a complicated reaction between the speech source and the formant filter, measuring the influence of f0 on MSDs in sentences that are not manipulated can be difficult. In addition, our experimental sentences are changing from trial-to-trial, making them far too variable to illuminate a relationship between f0 and MSDs. Therefore, two exploratory analyses, measuring of f0 and MSDs, were conducted for two sentences in the TIMIT corpus (SA1: "She had your dark suit in greasy wash water all year" and SA2: "Don't ask me to carry an oily rag like that"). These sentences were spoken by each of the 630 talkers in the TIMIT database and should theoretically contain similar phonetic content, keeping formant filters relatively consistent across sentences. When correlating f0 with MSDs, both correlations were significant (p's < 0.03) but the strength of the correlations were minimal (SA1: r = -0.17, SA2: r = -0.08). So, while there does appear to be a statistically significant relationship between f0 and MSDs, f0 does not explain a meaningful amount of variance in MSDs for these sentences. It is important to mention that while semantic content was the same for each talker producing these sentences, idiosyncrasies in pronunciation might have caused

variations in formant filters. Since formant frequencies have an impact on the shape of the LTAS, and by proxy the MSD, varying formant frequencies could be obscuring the relationship between f0 and MSDs. Thus, future exploration of the relationship between f0 and MSDs where the formants can be controlled is warranted.

The results of the current studies suggest that f0 variability is a key factor when experiencing context effects. Listeners are generally extremely sensitive to changes in f0 (e.g., Klatt, 1973; Hart, 1981). However, there are certain populations of hearing impaired listeners who cannot encode f0. Specifically, cochlear implant users are argued to have limited access to spectral pitch cues (Başkent, Mächler, Bolker, & Walker, 2014). This is primarily due to two reasons. First, the electrode array of the implant cannot be inserted deep enough into the cochlea to encode the frequencies that correspond to f0 (Faulker, Rosen, & Stanton, 2003). Second, the frequency resolution of CI is low causing difficulty in discriminating pitch differences (Başkent, Mächler, Bolker, & Walker, 2014). Importantly, it has been demonstrated that listeners hearing noise-vocoded single talker stimuli do demonstrate SCEs (Stilp, 2017). Here, noise-vocoding was used to model the processing of a CI so that normal hearing individuals would respond as CI users would. This suggests that CI users will experience contrast effects. Testing whether f0 variability influences the size of SCEs in listeners with cochlear implants would allow a deeper assessment of whether or not these effects are driven by f0 specifically. If f0 information is necessary to show the influence of variability on SCEs, CI users should not show differences in contrast effect magnitudes for our low and high f0 variability groups presented in Study 1b. However, if other sources of acoustic variability cueing

talker changes also influence SCEs, the CI listeners may still show smaller contrast effects when talkers are highly variable.

Response Times and Accuracy

Reaction times were slower overall in conditions where there were multiple talkers. This pattern corresponds to what would be expected based on talker normalization (e.g., Mullenix, Pisoni, & Martin, 1989). Thus, it appears that listeners are adjusting to differences between talkers in the context sentences, and this adjustment is influencing how fast they respond to the target vowels. Further, the increases in response time seem to combine two of our predicted patterns (see Figure 2). First, there are general increases in response times when hearing different talkers that affect the entire vowel continuum, following the predicted pattern in the upper left corner of Figure 2. Second, it appears that the influence of f0 variability is stronger on the /ɪ/ end of the continuum, in part following the predicted pattern in the upper right panel of Figure 2. Therefore, it appears that the /ɪ/ endpoint vowel is more strongly influenced by context. Prior to these analyses, there was no reason to expect that /ɪ/ would be influenced more by context then /ɛ/. In fact, there was no reason to expect that either of the endpoint vowels would be influenced by context. This surprising result suggests that even prototypical speech sounds (i.e., vowels categorized with a high degree of accuracy) can be influenced by context.

It is worth mentioning that assessing accuracy on endpoint vowels may be limiting our ability to observe differences between conditions. In all the studies presented here, a performance criterion dictated that participants had to maintain 80% accuracy when labeling endpoints. In addition, listeners had to pass a practice block that required

80% accuracy performance on endpoints before moving to the main experiment. This is a necessary condition of these experiments to ensure that listeners can identify the vowels as two different categories. If listeners cannot identify the endpoint vowels as different categories, a categorization shift cannot be observed. The performance criterion may have limited the ability to observe changes in accuracy due to the narrow range (80-100% accuracy). However, inspections of average accuracy at endpoints in each condition (see Appendix P) shows that listeners did not fall below 95% in any of our conditions. This suggests listeners that can tell the endpoint vowels apart are very accurate at labeling them.

The results reported here suggest that accuracy and response times are dissociable. Here response times were slower with multiple highly variable talkers but no change in accuracy was observed. Talker normalization literature often treats these measures as related. If an increase in response times is observed, a decrease in accuracy should also be observed. Here we showed that does not have to be the case; response times can increase or decrease independent of accuracy. Since the sensitivity of these measures to task difficulty differs, this is perhaps not surprising. In general, response times are a more sensitive measure of task difficulty. Accuracy can still be at ceiling levels even if the task has increased in difficulty. This has led speech research to alternative approaches to measure task difficulty.

Interestingly, to the author's knowledge, this is the first demonstration of contexts influencing endpoint vowels. The significant interactions between vowel and condition in Study 1b and Study 4 indicated that /ɪ/ vowels were more susceptible to the influence of f0 variability. Clear differences on response times between the conditions on the /ɪ/

endpoint are illustrated in Figures 13 and 16. Through measures of SCE magnitude, it appeared that endpoints were relatively impervious to the influence of context. This finding was so consistent that the influence of context is only assessed in the middle of the vowel continuum, where stimuli are intentionally ambiguous. In addition, accuracy revealed no influence of context on endpoints: endpoints are consistently labeled as their intended categories regardless of what context preceded them. Here, it was demonstrated that context can influence endpoint vowels when using a more sensitive measure of task difficulty. When contexts were more variable, listeners were slower at responding to the endpoints as well as mid-continuum vowels. In addition, the influence of f0 variability on response times seemed to be greater at the /ɪ/ endpoint relative to the other parts of the continuum. This suggests that even prototypical speech sounds, which can be consistently and accurately identified, can be influenced by context.

Accuracy has generally been used as a measure of task difficulty. As previously mentioned, speech perception is generally accurate, particularly when tested in quiet. Thus, ceiling effects of accuracy in speech, like the ones observed here, are common. There are a variety of methods that speech researchers can use to pull performance off of ceiling if ceiling effects are predicted to occur (e.g., adding noise to the signal). However, since the current studies were the first to quantify the influence of talker variability on accuracy of endpoints in SCE experiments, they were more exploratory in nature. Particularly, these studies attempted to measure if changes in accuracy would occur in multiple-talker conditions relative single-talker conditions. This was clearly not the case with none of our conditions showing differences in accuracy at endpoints.

A key difference between the measures of accuracy reported here and those typically reported in talker normalization experiments was when accuracy was assessed relative to the manipulation of talker. In the current studies, accuracy was assessed when the listener responded to the target vowel. However, the manipulation of talker was in the context that preceded the target. Thus, listeners were not responding directly to the acoustically varying talkers. In traditional talker normalization paradigms, the manipulation of talker is often in the stimulus the listener is identifying. For example, in Goldinger (1996), listeners were asked to recall if they had previous heard words in a list. Listeners were worse at identifying previously heard words if the talker changed between the first presentation of the word and the second presentation of the word. In this case, the listener is responding to the word spoken by a different talker. In the current studies, the listeners responded to targets spoken by the same talker in all conditions. While other measures collected here suggest that the influence of f0 variability can still be observed at the time of responding to the target (SCEs, response time), it is possible that vowel endpoint accuracy cannot. This key difference could have influenced the ability to detect changes in accuracy. However, if the purpose of measuring changes in accuracy is to infer task difficulty, measures that are more sensitive to changes in task difficulty may be more appropriate.

Response times can also be used as a measure of task difficulty. As task difficulty increases, listeners are slower to respond and response times increase. These increases in response times can be related to level of processing (e.g., Chabot, Miller, & Juola, 1976). If a task is more difficult, it may require a higher level of processing and response times increase. If response times are slow, then it can be inferred that processing the stimulus

required a deeper level of processing. Thus, response times can be used to infer both task difficulty and level of processing.

In these studies, response times were measured as how long it took the listener to respond to the target vowel once it began playing. Of primary interest was whether there are differences between response times following single talker contexts versus multiple talker contexts. The differences observed in the current studies can be broadly compared to those observed in talker normalization. In the current studies, response times increased when comparing a single talker to multiple talkers. This is similar to what is observed in studies of talker normalization: increased reaction times with an increase in the number of talkers. The current studies expanded on this finding, demonstrating that response times were greater when talkers were acoustically variable versus when they were acoustically similar. When these results are interpreted in conjunction with SCE magnitudes, they suggest that talker normalization and spectral contrast effects are influencing the same task. Further, the pattern of response times observed in the current studies suggest that response times increase as f0 variability increases. Study 1b demonstrated response times in high f0 variability conditions are slower than both single-talker and low f0 variability conditions. Thus, as with SCEs, simply hearing different talkers is not sufficient to slow speech perception: the talkers have to be acoustically different to observe an effect of changing talkers.

Further, it possible to assess whether talker normalization and spectral contrast effects may be related to one another using response times. If contrast effect magnitudes are decreasing at the same time that response times are increasing, it is possible that these effects are related to each other. To assess this relationship, a correlation was conducted

on individual's average deviation in response times and a measure of their contrast effect magnitude. There are clear idiosyncrasies in individual's response times with some listeners responding faster than others. In order to account for these differences, each individual's mean response time over all trials was subtracted from each response time. These differences were then averaged to obtain one response time measure for each individual in each condition. In order to make response times and SCEs more comparable, a measure of SCEs was used that accounts for changes in responses across the entire vowel continuum. Average difference in response times is measured over the entire vowel continuum. However, SCEs, as measured by mid-point shifts, tend to be measured in the middle of the vowel continuum. A different measure of SCE magnitude, percent shift in /ɛ/ responses, accounts for changes in responses across the entirety of the vowel continuum. Following low-F1 context sentences, listeners should have a higher percentage of /ɛ/ responses than following high-F1 context sentence. Thus, changes in percent /ɛ/ responses measure categorization shifts following different contexts, similar to mid-point shifts, but encompass the entire vowel continuum. Percent shifts in /ɛ/ responses are highly related to mid-point shifts for the data presented in these studies (r = 0.96, $p < .001$). In addition, percent shifts in phoneme responses has been used in previous literature as measures of contrast effects (e.g., Laing et al., 2012; Holt 2005). Again, using this measure has the benefit of encompassing the entire vowel continuum making measures of SCEs more comparable to measures of response times. Correlating each individual's average difference in response times with their contrast effect magnitudes (measured as changes in percent /ɛ/) produced a non-significant correlation (Figure 17; r = -0.004, p = 0.94). An important consideration when interpreting the

results is that different listener groups participated in each experiment. In order to control for this, a linear mixed effects regression was conducted in R using average difference in response times to predict changes in percent /ɛ/ responses while including a random effect of subject. Average difference in response times was not a significant predictor of SCEs ($t(1) = -0.08$, $p = 0.94$) confirming that response times were not related to SCE magnitudes. Therefore, it appears that, when considered at the level of the individual, SCEs and response times were not related. However, both of these effects are still influenced by f0 variability. As previously mentioned, SCEs in this study were measured when the listener was responding to the target vowel but the manipulation of talker occurred in the context. This may have limited the relationship between response times and SCEs in these studies. It is possible that a future study with stimuli where the target sound occurs within the context sentence, where the talker varies, might be more sensitive to the relationship between response times and SCEs.

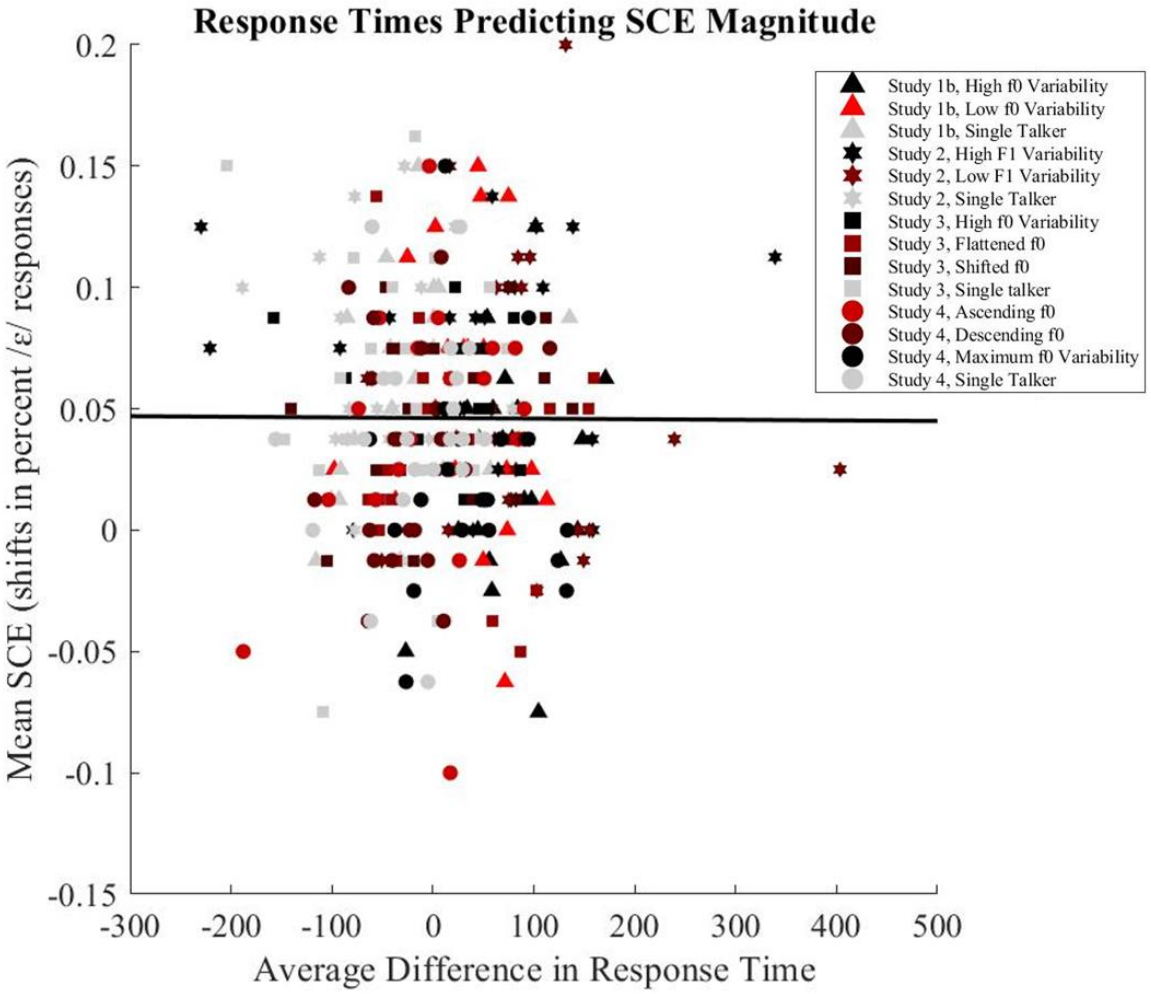

Figure 17. The relationship between individual's average deviance in response times and their shifts in percent /ɛ/ responses. Triangles represent data from Study 1b. Stars represent data from Study 2. Squares represent data from Study 3. Circles represent data from Study 4. Gray symbols represent single talker conditions. Black symbols represent high variability conditions. Various shades of red represent low variability conditions.

Recently, researchers claimed that the decrements in speech perception observed when hearing different talkers were obligatory (Choi, Hu, & Perrachione, 2018). In this study, the authors assessed whether talker normalization influences phonetically ambiguous stimuli more than non-ambiguous stimuli. Decreases in response times were observed in both conditions. This led the authors to claim that talker normalization was obligatory since it occurred with both non-ambiguous and ambiguous stimuli.

Importantly, their multi-talker conditions mixed both male and female talkers suggesting that acoustic variability was high in these conditions although no measures of acoustic variability were reported. Previous research (e.g., Goldinger, 1996) and the research results reported here challenge that claim by finding that the influence of context does not occur in all conditions. Particularly, the occurrence of a context effect seems to be related to how variable the talkers are. Goldinger (1996) demonstrated that listeners were less accurate at recognizing words previously heard when the talkers producing the words were more acoustically different than when the talkers were more acoustically similar. Here, similar results were observed with response times. Study 1b demonstrated that listeners were slower at responding to targets when the contexts were spoken by highly variable talkers relative to single and similar talkers. When listeners heard similar sounding talkers, they were not slower than single talkers ($p = 0.051$), suggesting that similar-sounding talkers can elicit response times broadly similar to single talkers. Thus, the decrements in speech perception, in terms of response times, are not obligatory but only occur when talkers are acoustically different.

As mentioned in the introduction, speech acoustics are immensely variable. There are a number of parameters that can vary in speech, some of which are related to talker differences. Talker normalization has been suggested to be a means to adjust for these differences between talkers (e.g., Joos, 1948; Ladefoged & Broadbent, 1957; Mullenix, Pisoni, & Martin, 1989; Martin, Mullenix, Pisoni, Summers, 1989). It appears that this need to adjust is driven by the acoustic differences between talkers (Goldinger, 1996). If speech is highly acoustically variable, greater adjustments need to be made. If speech is less acoustically variable, less adjustment need to be made. Here, it is suggested that

SCEs are also related to acoustic variability. If talkers are acoustically variable, there is less of an influence of context. As evidenced by the regression reported in the synthesis of Studies 1 and 2 (Figure 11), a significant portion of the variability observed in SCE magnitude is explained by the degree to which mean f0 varies across context sentences. In addition, the effects of talker normalization and SCEs may both lead to more accurate perception when talkers are similar. In talker normalization, accuracy is a primary measure of the effect. Listeners are more accurate when talkers do not change or are less acoustically different. In SCEs, similar talkers lead to a greater influence of context. This influence helps to disambiguate otherwise ambiguous sounds, leading to more accurate perception. Thus, it is possible that these effects work in conjunction to lead to accurate perception in the presence of minimal acoustic variability. Restraining the influence of context when acoustic variability is high may also lead to more accurate speech perception. When acoustic variability is high, there is less of an influence of context to disambiguate otherwise ambiguous sounds. As previously mentioned, if the contexts are acoustically distinct, disambiguating sounds based on previous context could be maladaptive as that context is likely no longer informative for perception of the current sound(s).

Other Possible Influences on SCEs

In the studies reported here, there were 40 unique talkers in each of the multi-talker conditions. Previous work in our lab demonstrated that the influence of talker is observed with as few as 16 talkers (Mohiuddin, Assgari & Stilp, 2016). Using 40 talkers well exceeds this number, ensuring that an influence of talker variability can be measured in all conditions. However, using 40 talkers required that the sentence from each talker be

repeated 4 times per block in a given condition. When the sentences were randomized, it is unlikely that the listener could learn the characteristics of any one talker and a range of contrast effects was observed based on f0 variability. However, repeating sentences on successive trials in Study 4 may have led to order of talkers being somewhat predictable. In the ascending and descending conditions, this may have not been an issue since increasing or decreasing pitch would have also lead to predictability. However, in the maximum variability condition, listeners may have caught on to the pattern and known what to expect during the repeated trials. So, while it seems that a sample of 40 talkers is enough to demonstrate decreases in SCEs, the influence of hearing different talkers could have potentially been mitigated by talker similarity and predictability. Study 1a and 1b showed that the influence of hearing different talkers was mitigated by talker similarity even if the trials were randomized. In the discussion of Study 4, it was suggested that even when talkers are acoustically different at the level of the condition, predictability may restrain the influence of talker variability.

Another way that the current results support the low-level nature of SCEs is that semantic content of the sentence had no influence on the size of the context effects. In previous literature and in our single talker conditions, context sentences would often be related to the task (e.g., “Please say what vowel this is” Assgari & Stilp, 2015 or “Please say what word this is” Ladefoged & Broadbent, 1957). Stilp & Assgari (2017) argued that using a single sentence whose semantic content was unrelated to the task, did not negatively impact the size of contrast effects. In the multiple talker conditions of the current studies, not only was the semantic content of the sentences unrelated to the task, it also varied wildly. Importantly, despite semantic variability in low f0 variability

conditions, SCEs observed were similar to single-talker conditions. This suggests that drawing the listeners' attention to the task via the content of the sentence is not necessary. Again, higher-level influences did not influence the magnitude of lower-level context effects in speech perception.

While the results reported here suggest that low-level acoustic cues are the primary influence on SCEs, it is possible that other higher-order information, not tested here, may also influence SCEs. Previous research has demonstrated that listener expectations can override the influence of low-level acoustic variability on speech perception (Magnuson & Nusbaum, 2007). In these studies, two sets of listeners heard the same stimuli synthesized with a 10 Hz difference in f0. Importantly, the authors manipulated listeners' expectations of talkers by telling one group they would hear two talkers and the other group that they would hear a single talker. When listeners were told to expect to hear two different talkers, speech perception was slower. This suggests that listeners were experiencing talker normalization and adjusting for differences between talkers. Despite hearing the same stimuli, when listeners were told to expect a single talker, speech perception was faster. Thus, listener expectations about talkers may override the influence that f0 variability can have on speech perception. While intriguing, these results have limited applicability to the current students for two reasons. First, in the current experiment, listeners were not told to expect anything about the number of talkers they would encounter. Despite acoustic similarity, it was possible to tell the talkers apart. Second, the context sentences reported here often differed by much more than 10 Hz. It is likely that listener expectations can influence context effects in so far as the expectation of talkers is reasonably supported by acoustics. In Magnuson and Nusbaum (2007), the

stimuli differed in pitch by 10 Hz. This difference can be reasonably attributed to a pitch change within talker. For example, in the single talker/200 sentence condition of Assgari and Stilp (2015) the standard deviation of mean f0 across sentences was 12.15 Hz. This suggests that a 10 Hz difference between two sentences falls well within the range of f0 differences that can be attributed to a single talker. In the current studies, pitch variability in the high f0 variability conditions was much greater and it is possible that this variability would not be able to be modified by listeners' expectations.

Another important consideration is whether listeners are familiar with the talker. Nygaard and colleagues have demonstrated that speech perception is more accurate if the listener is familiar with the talker (Nygaard, Sommers, & Pisoni, 1994; Nygaard & Pisoni, 1998). Familiarity with talker has been defined as either extensive experience (e.g., family members) or familiarizing listeners with lab training. Both types of familiarity lead to increased performance in speech perception tasks. Since increases in speech perception are observed when the listener is familiar with the talker, this could offset the decreases in speech perception observed when talker changes. In addition, if predictability of stimuli influences SCEs (as discussed earlier) then perhaps being able to predict the acoustics of the talker through familiarization would also influence SCEs. However, both of these claims have yet to be addressed in either talker normalization or SCEs.

Limitations

When comparing the results of the current experiments to past experiments, there were some fundamental differences that deserve consideration. Each of the current experiments included a single talker condition. This allowed the SCEs magnitudes from

the current studies to be compared with previous results (e.g., Assgari & Stilp, 2015). In addition, other between-study comparisons were possible. Preliminary data were collected for two experiments: Study 1b and Study 2. Further, in Study 1b and Study 4, the high variability conditions were exactly the same. Thus, it was possible to check if conditions replicated results across experiments. Differences in the SCEs observed in these conditions suggest that groups of listeners may have been responding differently. While some of these differences may be attributed to individual differences between listener groups, two main methodological differences deserve consideration. First was the inclusion of a single-talker condition in all experiments. Single-talker conditions served as a control condition to compare each set of listeners to each other. These conditions should also estimate the upper limit of SCE magnitudes for each listener group. Since there is no acoustic variability due to talker changes, SCEs should be largest in the single-talker condition. However, it is possible that the presence of a single-talker condition may have caused the listener to perceive low-variability conditions as more variable than in past experiments where no single-talker conditions was present. Second, listeners in the current study were asked to respond using a button box rather than a mouse click. The button box was used to allow for the collection of response times. Listeners were told to press the left button to respond /ɪ/ and the right button to respond /ɛ/. However, no further instructions were given. It is possible that listeners were using the button box differently (e.g., one vs two hands, finger placement after responding, etc.). Due to the repeated measures nature of our designs, differences between response methods should be controlled for, but it is also possible that a listener switched response methods mid-experiment, changing their patterns of results. These differences in utilization of the

button box should have likely only influenced response times, but it is possible that they also produced more variability within subjects than previous reported. Further, which button corresponded to which vowel response was not counterbalanced. It is possible that this may have caused differences in the time it took to respond to one category relative to the other.

Table 2.

Observed power when testing SCEs in each experiment as output by SPSS

Study	Observed Power
Study 1b	0.601
Study 2	0.129
Study 3	0.376
Study 4	0.238

In the current studies, small effect sizes and large individual variability were observed. This may be resulting in the statistical analyses being underpowered (see Table 2 for a list of observed power as outputted by SPSS). Effect sizes in the proposal were estimated based on the differences between single-talker and the 200-talker conditions in Assgari and Stilp (2015). In the current studies, comparisons were drawn between multi-talker conditions that varied in terms of their variability in mean f0. The differences between means are smaller than previously considered. As previously mentioned, each of the studies reported here included an additional single-talker condition that has not always been included in the past. Comparing additional means that are expected to be similar may be introducing issues with the ANOVA method of testing group differences.

An alternative approach that may better account for individual differences is using Mixed Effects Models with subject as a random variable. Mixed effects models were conducted in R (R Development Core Team, 2016) using the lme4 package (Bates et al., 2014) to analyze the results for each study. Previous work in our lab has utilized mixed effect models to analyze results from SCE experiments (Stilp et al., 2015; Stilp & Assgari, 2017a). The models conducted here were maximal following the suggestions of Barr, Levy, Scheepers and Tily (2013). Thus, each model had every fixed effect also included as a random effect. The patterns of results with mixed effects models did not differ from the results of ANOVAs. Therefore, attempts to account for variability due to individual differences did not change the interpretation of our results. Further, differences between multiple talker-conditions and single-talker conditions only arise when spectral peaks are modest (here, +5 dB). Modest peaks produce modest SCEs. When there is a small change from context to the target, there is a smaller shift in categorization (Stilp, Anderson & Winn, 2015; Stilp & Assgari, 2017a). Detecting a decrease in an already modest effect requires more power. This is evident by the results of Study 4. In Study 4, the maximum variability condition did not produce a contrast effect significantly different than 0, while all other conditions did. Despite this fact, the omnibus ANOVA failed to detect any differences between means. This suggests that even if the effect is decreased to the point that it cannot be claimed that SCEs occurred, it cannot be claimed that this effect is smaller than conditions that did produced an SCE. Several of these issues may be addressed by adding listeners to the data set in an attempt to reduce variability and increase power. However, the number of listeners in the current study met the proposed sample sizes and similar studies generally do not test more than 20 listeners.

The target vowels in the current studies are primarily differentiated based on their F1 frequencies. Differences in the F1 regions in our low-F1 (100-400 Hz) and high-F1 (550-850 Hz) sentences are argued to produce the SCE. The amplitude in these regions was manipulated to ensure MSDs that would produce contrast effects. It is worth noting that the low-F1 region often encompasses the fundamental frequency. In these studies, the average f0 fell between 80-260 Hz. Since our manipulation of talker was based on f0 variability, it is possible that these manipulations were somehow confounded with the energy in the F1 regions. Previously, the argument was made that changing f0 could have an influence on MSDs. However, it is possible that these differences are only confined to the low-F1 region where f0 is typically located. If f0 variability influences context effects for reasons other than its effect on the distribution of energy in the F1 regions considered here, then f0 variability should also influence other types of context effects. Demonstrations of f0 variability influencing non-spectral context effects would allow for the spectral consequences of changing f0 to be separated from its influence on context effects. This idea will be discussed further in the Future Directions section.

Future Directions

Several future directions have already been discussed but will be reviewed briefly here. First, future studies could explore the extent to which other context effects are also influence by f0 variability. For example, it would be interesting to measure how f0 variability influences other types of contrast effects (e.g., temporal; e.g., Summerfield, 1981) or spectral calibration (e.g., Stilp et al., 2016). Spectral calibration occurs when a spectral property does not change from context to target. In this case, listeners rely less on the unchanging parameter in favor of a more informative cue to phoneme distinction.

Measuring the influence of f0 variability on other types of context effects would have two main advantages. First, if context effects that are not spectrally based show an influence of f0, then it is clear that the spectral consequences of changing f0 are not solely responsible for the influence of f0 reported here. However, if f0 variability does not influence context effects that are not spectrally based, then it is likely acoustic changes driven by changes in f0 that are responsible for the influence of f0 variability on SCEs. Second, if other context effects are also influenced by f0 variability, the hypothesis that f0 changes cue that past perception should no longer bias current perception is further supported.

Second, future studies could explore how f0 influences context effects when f0 is not as strong a cue to talker changes. Here, the case was made that f0 cueing changes in talker may help explain the influence of f0 on context effects. However, there are situations where f0 is not as strong a cue to talker changes. In tonal languages, like Mandarin Chinese, f0 is also a cue to lexical changes (Wong & Diehl, 2003). Adjusting f0 to cue lexical changes leads to more overlap in f0 of different talkers (Wong & Diehl, 2003). This suggests that differences in f0 between talkers may be of lesser magnitude in tonal languages. As previously mentioned, it is possible that f0 cueing talker changes is what drives the influence of f0 variability on SCEs. If this is the case, then in situations where f0 changes are not as strong a cue to talker changes, there may be less of an influence of f0 variability on SCEs. However, if f0 variability still influences contrast effects in tonal languages, then this lends more credibility to the hypothesis that acoustic consequences of changing f0 drive the influence of f0 variability on SCEs.

Third, if it is f0 cueing different talkers that influences context effects, then other acoustic cues that relate to differences between talkers may also influence context effects. For example, other acoustic cues, such as F3, have been argued to be important for distinguishing between talkers (Johnson, 2005). If F3 variability has a similar influence on SCEs in F1 regions as f0 variability, then the hypothesis that acoustic correlates of talker changes influence SCEs is further supported. However, if F3 variability does not have a similar influence, then future efforts should concentrate on how the acoustic consequences of varying f0 may be impacting SCEs.

Finally, other measures of task difficulty could be utilized to fully explore when f0 variability is influencing speech perception. As previously mentioned, accuracy and response times are generally used as a way to measure task difficulty. If the task is sufficiently difficult, participants are less accurate and slower at responding. More recently, it has become apparent that difficulty of listening tasks may not be sufficiently quantified using accuracy (e.g., Winn, Edwards, & Litovsky, 2015). There are cases where performance is at ceiling but other indicators suggest the task was difficult (e.g., reaction time, listener self-report, etc). In these cases, researchers have begun to measure listening effort through physiological responses such as pupil dilation (e.g., Winn, Edwards, & Litovsky, 2015) and skin conductance (e.g., Mackersie & Cones, 2011). These measures are more sensitive to task difficulty. In the SCE paradigm described in these studies, it would theoretically be possible to measure listening effort at two vital times during each trial. First, is during the context sentences where talkers are changing. If hearing different talkers leads to increases in task difficulty, listening effort should increase when hearing the contexts spoken by different talkers. Further, the results

reported here suggest that acoustically different talkers should lead to increases in listening effort over acoustically similar talkers. Second, listening effort could also be measured when the participant is listening to the target. In most conditions, talker is changing from context to target as well as from context to context and target to context. However, in single-talker conditions, the sentence and the target were spoken by the same individual. Thus, any increases in listening effort can be attributed to changing the talker from context to target. Further, if a single talker condition is included where the talker changes from context to target, it is possible to assess how changing talker from context to target affects listening effort. An increase in listening effort from this condition to a multi-talker condition could be directly attributable to changing talkers from trial-to-trial and not a context and target mismatch. The ability to measure listening effort during passive listening allows a more nuanced analysis of how changing talker is affecting task difficulty. Through this method, it could be possible to assess where the influence of hearing different talkers is first observed. Further, it would be possible to assess if task difficulty continues to increase as talkers continue to vary.

Conclusion

Speech is immensely variable with several sources of acoustic variability. Contextual influences can serve as a way to accommodate this variability. The current studies sought to assess to what degree two of these contextual influences, talker normalization and spectral contrast effects, may be related. Talker normalization and spectral contrast effects can be measured simultaneously suggesting that they may be acting on speech at the same time. In addition, f0 variability modulates the degree to

which context influences speech perception, suggesting that these effects are related to low-level acoustic variability.

REFERENCES

Ainsworth, W. (1975). Intrinsic and extrinsic factors in vowel judgments. *Auditory analysis and perception of speech*, 103-113.

Assgari, A.A., Frazier, J.M., & Stilp, C.E. (2018, May). "Musical instrument categorization is highly sensitive to spectral properties of earlier sounds." Paper accepted for presentation at the 175th Meeting of the Acoustical Society of America, Minneapolis, Minnesota.

Assgari, A. A., & Stilp, C. E. (2015). Talker information influences spectral contrast effects in speech categorization. *The Journal of the Acoustical Society of America, 138*(5), 3023-3032.

Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *The Journal of the Acoustical Society of America, 71*(4), 975-989.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Başkent, D., Gaudrain, E., Tamati, T. N., and Wagner, A. (2016). "Perception and psychoacoustics of speech in cochlear implant users," In A. T. Cacace, E. de Kleine, A. G. Holt, and P. van Dijk (Eds.), Scientific foundations of audiology: perspectives from physics, biology, modeling, and medicine, Plural Publishing, Inc, San Diego, CA, pp. 285–319. ISBN13:978-1-59756-652-0

Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 5.3.61, retrieved 1 January 2014 from http://www.praat.org/

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Nasser Kotby, M., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., and Ludvigsen, C. (1994). 'An international comparison of long-term average speech. The Journal of the Acoustical Society of America, 84(3), 1100– 1104.

Chabot, R. J., Miller, T. J., & Juola, J. F. (1976). The relationship between repetition and depth of processing. Memory & Cognition, 4(6), 677-682.

Choi, J. Y., Hu, E. R., & Perrachione, T. K. (2018). Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. Attention, Perception, & Psychophysics, 80(3), 784-797.

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

Fairbanks, G., House, A. S., & Stevens, E. L. (1950). An experimental study of vowel intensities. *The Journal of the Acoustical Society of America*, *22*(4), 457-459.

Fant, G., (1960): Acoustic theory of speech production. Mouton, The Hague.

Faulkner, A., Rosen, S., & Stanton, D. (2003). Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth. The Journal of the Acoustical Society of America, 113(2), 1073-1080.

Fourcin, A. (1968). Speech source inference. *IEEE Transactions on Audio and Electroacoustics, 16*(1), 65-67.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., & Dahlgren, N. (1990). "DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM", National Institute of Standards and Technology, NIST Order No. PB91-505065.

Geiselman, R. E., & Bellezza, F. S. (1976). Long-term memory for speaker's voice and source location. *Memory & Cognition, 4*(5), 483-489.

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*(5), 1166.

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

Hall, J. F. (1954). Learning as a function of word-frequency. *The American journal of psychology, 67*(1), 138-140.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America, 97*(5), 3099-3111.

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science, 16*(4), 305-312.

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America, 120*(5), 2801-2817.

Johnson, K. (1991). Differential effects of speaker and vowel variability on fricative perception. *Language and speech, 34*(3), 265-279.

Johnson, K. (2005). Speaker normalization in speech perception. The handbook of speech perception, 363-389.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*(4), 359-384.

Joos, M. (1948). Acoustic phonetics. *Language 24*, 1–136.

Kang, S., Johnson, K., & Finley, G. (2016). Effects of native language on compensation for coarticulation. *Speech Communication, 77*, 84-100.

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009, October). Development of exploratory research tools based on TANDEM-STRAIGHT. In Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference (pp. 111-120). Asia-Pacific Signal and

Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee.

Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication, 41*(1), 59-69.

Ladefoged, P. (1989). A note on ‘‘Information conveyed by vowels’’. *The Journal of the Acoustical Society of America, 85*(5), 2223-2224.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America, 29*(1), 98-104.

Laing, E. J., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in psychology*, 3.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review, 74*(6), 431.

Logan, J.S. & Pisoni, D.B. (1987) Talker variability and the recall of spoken word lists: A replication and extension. In Research on Spoken Language Processing Progress Report No. 13 (pp. 307-328). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Attention, Perception, & Psychophysics, 60*(4), 602-619.

Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization (No. 8). Oxford University Press on Demand.

Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indexes of listening effort in a competing-talker task. Journal of the American Academy of Audiology, 22(2), 113-122.

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(4), 676.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human perception and performance*, *33*(2), 391.

Mohiuddin, A., Assgari, A.A., & Stilp, C.E. (2016). "Effects of talker variability on spectral contrast effects". Poster presented at the 2016 Undergraduate Research and Community Engagement Symposium, University of Louisville, Louisville, Kentucky.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*(4), 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365-378.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of experimental psychology*, *64*(5), 482.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America, 85*(5), 2088-2113.

Neuhoff, J. G., Wayand, J., Ndiaye, M. C., Berkow, A. B., Bertacchi, B. R., & Benton, C. A. (2015). Slow change deafness. Attention, Perception, & Psychophysics, 77(4), 1189-1199.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology, 47*(2), 204-238.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Attention, Perception, & Psychophysics, 60*(3), 355-376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*(1), 42-46.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America 24*, 175–184.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693-703.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech communication, 13*(1), 109-125.

Rand, T. C. (1971). Vocal tract size normalization in the perception of stop consonants. *The Journal of the Acoustical Society of America, 50*(1A), 139-139.

Roberts, T. P., Flagg, E. J., & Gage, N. M. (2004). Vowel categorization induces departure of M100 latency from acoustic prediction. *Neuroreport, 15*(10), 1679-1682.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33*(3), 441.

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics, 73*(4), 1195-1215.

Sjerps, M. J., & Reinisch, E. (2015). Divide and conquer: How perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 710.

Sjerps, M. J., & Smiljanić, R. (2013). Compensation for vocal tract characteristics across native and non-native languages. *Journal of Phonetics, 41*(3), 145-155.

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., ... & Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear and hearing, 33*(1), 112.

Stilp, C.E. (2017). Acoustic context alters vowel categorization in perception of noise-vocoded speech. Journal of the Association for Research in Otolaryngology, 18(3), 465-481.

Stilp, C. E., & Alexander, J. M. (2016, May). Spectral contrast effects in vowel categorization by listeners with sensorineural hearing loss. In Proceedings of Meetings on Acoustics 171ASA (Vol. 26, No. 1, p. 060003). ASA.

Stilp, C. E., Alexander, J. M., Kiefte, M., & Kluender, K. R. (2010). Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets. *Attention, Perception, & Psychophysics*, *72*(2), 470-480.

Stilp, C. E., & Anderson, P. W. (2014). Modest, reliable spectral peaks in preceding sounds influence vowel perception. The Journal of the Acoustical Society of America, 136(5), EL383-EL389.

Stilp, C. E., Anderson, P. W., Assgari, A. A., Ellis, G. M., & Zahorik, P. (2016). Speech perception adjusts to stable spectrotemporal properties of the listening environment. Hearing research, 341, 168-178.

Stilp, C. E., Anderson, P. W., & Winn, M. B. (2015). Predicting contrast effects following reliable spectral properties in speech perception. *The Journal of the Acoustical Society of America*, *137*(6), 3466-3476.

Stilp, C. E., & Assgari, A. A. (2017a). Consonant categorization exhibits a graded influence of surrounding spectral context. The Journal of the Acoustical Society of America, 141(2), EL153-EL158.

Stilp, C.E. & Assgari, A.A. (2017b). "Filtered and unfiltered sentences produce different spectral context effects in vowel categorization." Poster presented at the 174th Meeting of the Acoustical Society of America, New Orleans, Louisiana.

Studebaker, G. A. (1985). A rationalized arcsine transform. Journal of Speech, Language, and Hearing Research, 28(3), 455-462.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. Journal of Experimental Psychology: Human Perception and Performance, 7(5), 1074.

Verbrugge, R., Strange, W., & Shankweiler, D. (1974). What information enables a listener to map a talker's vowel space? *The Journal of the Acoustical Society of America, 55*(S1), S53-S54.

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. Journal of Experimental Psychology: Human Perception and Performance, 29(2), 333.

Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *The Journal of the Acoustical Society of America, 90*(6), 2942-2955.

Watkins, A. J., & Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *The Journal of the Acoustical Society of America, 96*(3), 1263-1282.

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. Ear and hearing, 36(4), e153.

Wong, P. C., & Diehl, R. L. (2003). Perceptual normalization for inter-and intratalker variation in Cantonese level tones. Journal of Speech, Language, and Hearing Research, 46(2), 413-421.

Wong, P. C., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience, 16*(7), 1173-1184.

APPENDICES

Appendix A.

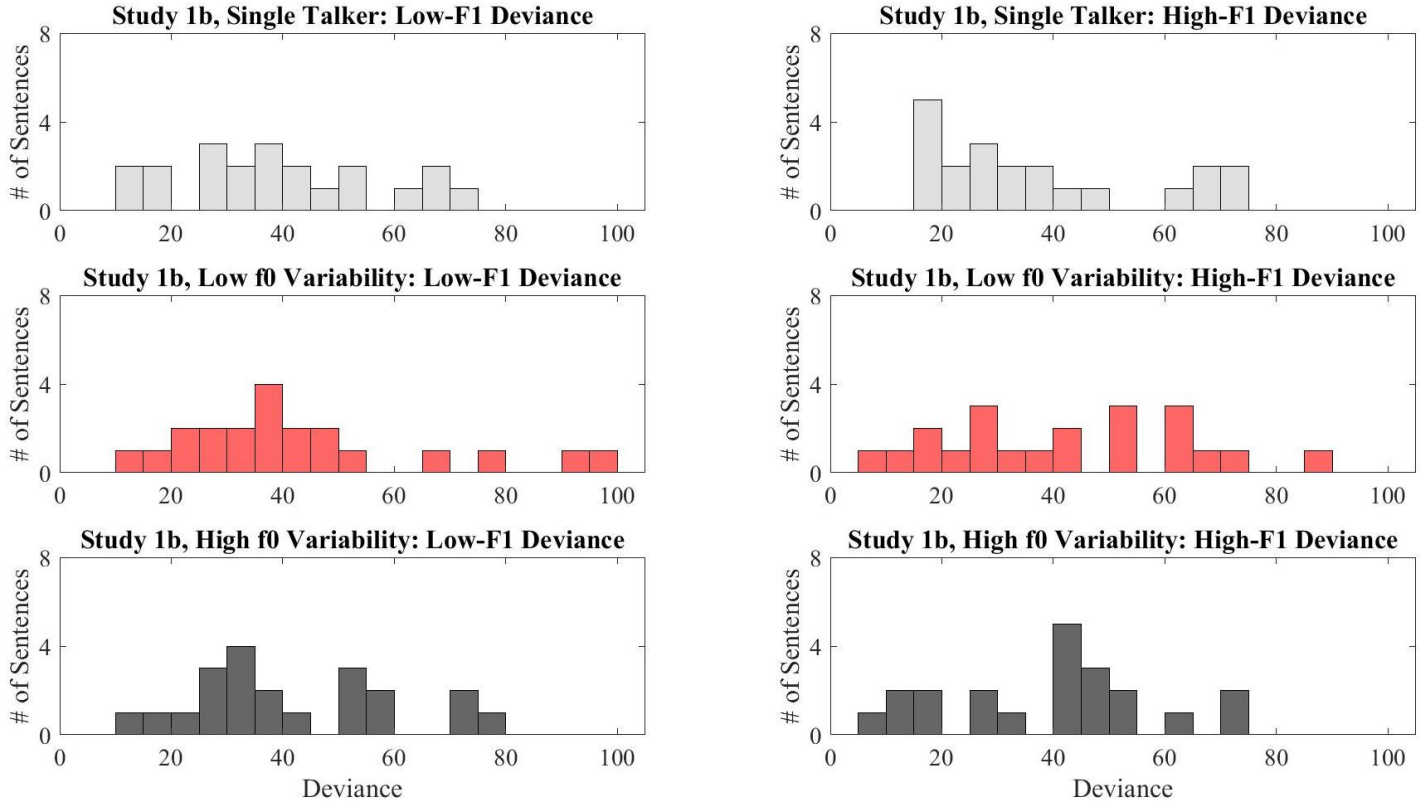

Figure 18. Histograms of deviance measures for the conditions in Study 1b. The y-axis displays number of sentences. The x-axis displays deviance measures as output by the glmfit command in Matlab. Low-F1 deviances are on the left and High-F1 deviances are on the right. The top row contains deviance measures for the single talker condition. The middle row contains deviance measures from the Low f0 variability condition. The bottom row contains deviance measures from the High f0 variability condition.

Appendix B.

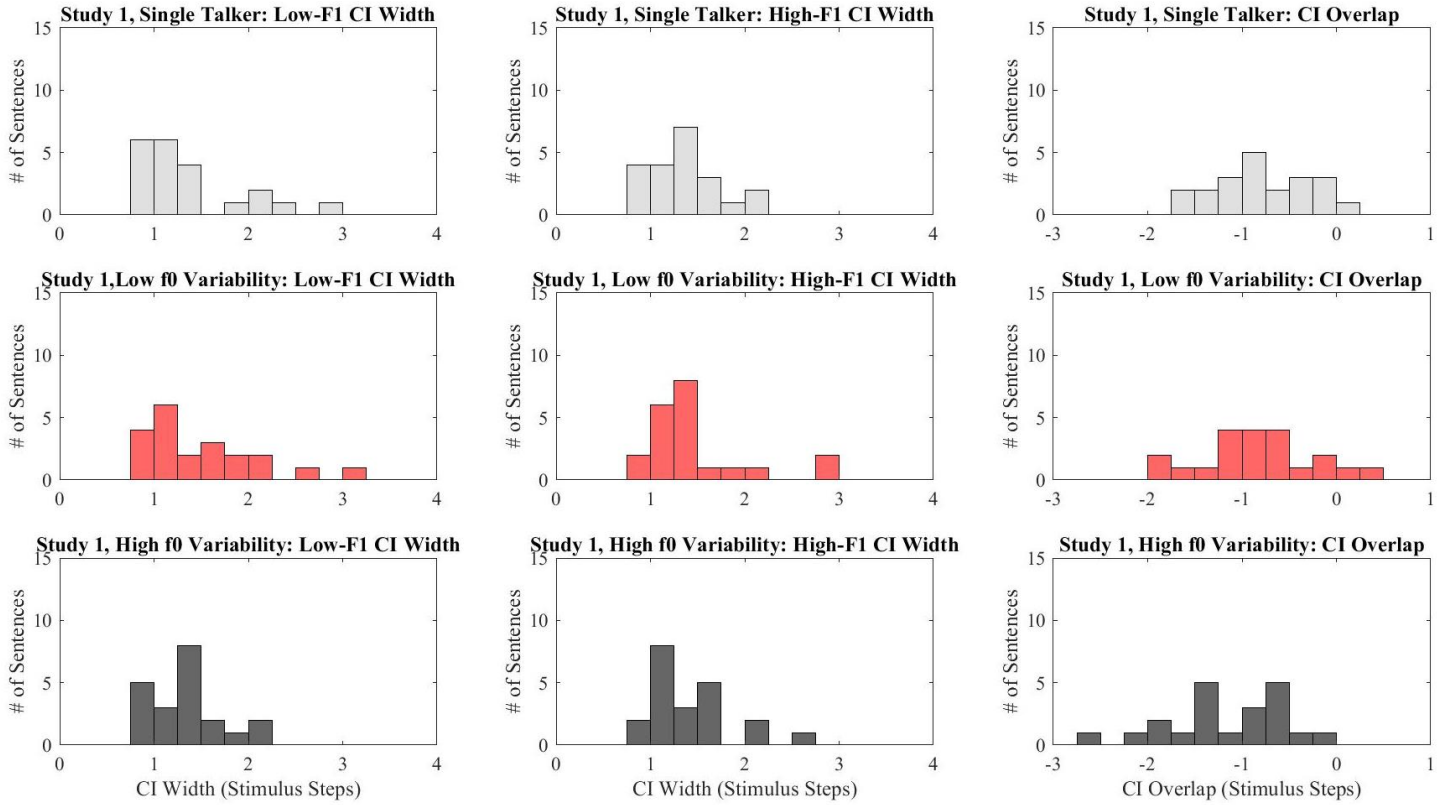

Figure 19. Histograms of confidence intervals around midpoints of logistic functions for the conditions in Study 1b. The y-axis displays number of sentences. For the first two columns, the x-axis displays the width of the 95% confidence interval around the midpoint of the logistic function in stimulus steps. The left column contains low-F1 confidence intervals. The middle column contains high-F1 confidence intervals. The x-axis of the right column displays the overlap of the low-F1 and high-F1 CIs (lower bound of high-F1 CI minus upper bound of low-F1 CI). Negative values indicate overlap. The top row contains measures from the single talker condition. The middle row contains measures from the low f0 variability condition. The bottom row contains measures from the high f0 variability condition.

Appendix C.

Table 3

Pairwise t-tests for the main effect of vowel for Study 1b

Reference Vowel	Comparison Vowel	Mean Difference	SE	p-value
/ɪ/	2	5.12	12.63	1.00
	3	-5.71	10.54	1.00
	4	-49.85	12.74	0.04*
	5	-86.83	19.70	0.01*
	6	-72.76	26.49	1.00
	7	-49.00	20.92	1.00
	8	-4.45	14.55	1.00
	9	13.62	12.00	1.00
	/ɛ/	42.56	13.09	0.18
2	3	-10.83	11.44	1.00
	4	-54.97	14.53	0.05
	5	-91.95	20.88	0.01*
	6	-77.88	27.58	0.47
	7	-54.12	22.07	1.00
	8	-9.57	16.13	1.00
	9	8.50	15.75	1.00
	/ɛ/	37.45	15.91	1.00
3	4	-44.14	10.87	0.03*
	5	-81.116	18.67	0.01*
	6	-67.04	24.98	0.64
	7	-43.29	19.67	1.00
	8	1.26	16.42	1.00
	9	19.34	13.17	1.00
	/ɛ/	48.28	14.97	0.19
4	5	-36.98	13.72	0.63
	6	-22.91	20.89	1.00
	7	0.85	13.84	1.00
	8	45.40	13.76	0.16
	9	63.47	13.39	0.01*
	/ɛ/	92.42	14.39	<0.001*
5	6	14.07	14.45	1.00
	7	37.83	10.73	0.10
	8	82.38	11.79	<0.001*
	9	100.45	14.31	<0.001*
	/ɛ/	129.39	14.36	<0.001*
6	7	23.76	10.55	1.00
	8	68.31	17.42	0.04*
	9	86.38	18.43	0.01*
	/ɛ/	115.32	20.56	0.001
7	8	44.55	12.18	0.07

	9	62.62	13.81	0.01*
	/ɛ/	91.56	15.16	<0.001*
8	9	18.07	9.30	1.00
	/ɛ/	47.02	8.70	.001*
9	/ɛ/	28.94	9.01	0.20

Appendix D.

Table 4

Pairwise t-tests for the main effect of vowel for Study 1b

Vowel	Reference Condition	Comparison Condition	Mean Difference	SE	p-value
/ɪ/	1	2	54.53	22.31	0.07
		3	135.85	22.51	<0.001*
	2	3	81.32	25.22	0.01*
2	1	2	66.27	31.13	0.14
		3	127.89	29.14	0.001*
	2	3	61.62	26.77	0.10
3	1	2	55.33	18.68	0.02*
		3	132.90	21.90	<0.001*
	2	3	77.57	24.61	0.02*
4	1	2	85.45	25.80	0.01*
		3	156.57	25.80	<0.001*
	2	3	71.12	29.63	0.08
5	1	2	9.78	25.67	1.00
		3	92.47	37.60	0.07
	2	3	82.68	41.83	0.19
6	1	2	42.18	26.02	0.36
		3	72.00	25.65	0.03*
	2	3	29.82	26.17	0.80
7	1	2	19.02	23.87	1.00
		3	65.23	34.27	0.21
	2	3	46.21	33.97	0.57
8	1	2	69.20	24.90	0.04*
		3	78.11	31.07	0.06
	2	3	8.91	25.59	1.00
9	1	2	18.13	19.86	1.00
		3	68.40	26.32	.05
	2	3	50.27	25.26	.18
/ɛ/	1	2	41.36	21.98	.22
		3	92.40	31.62	.03*
	2	3	51.04	22.37	.10

Appendix E.

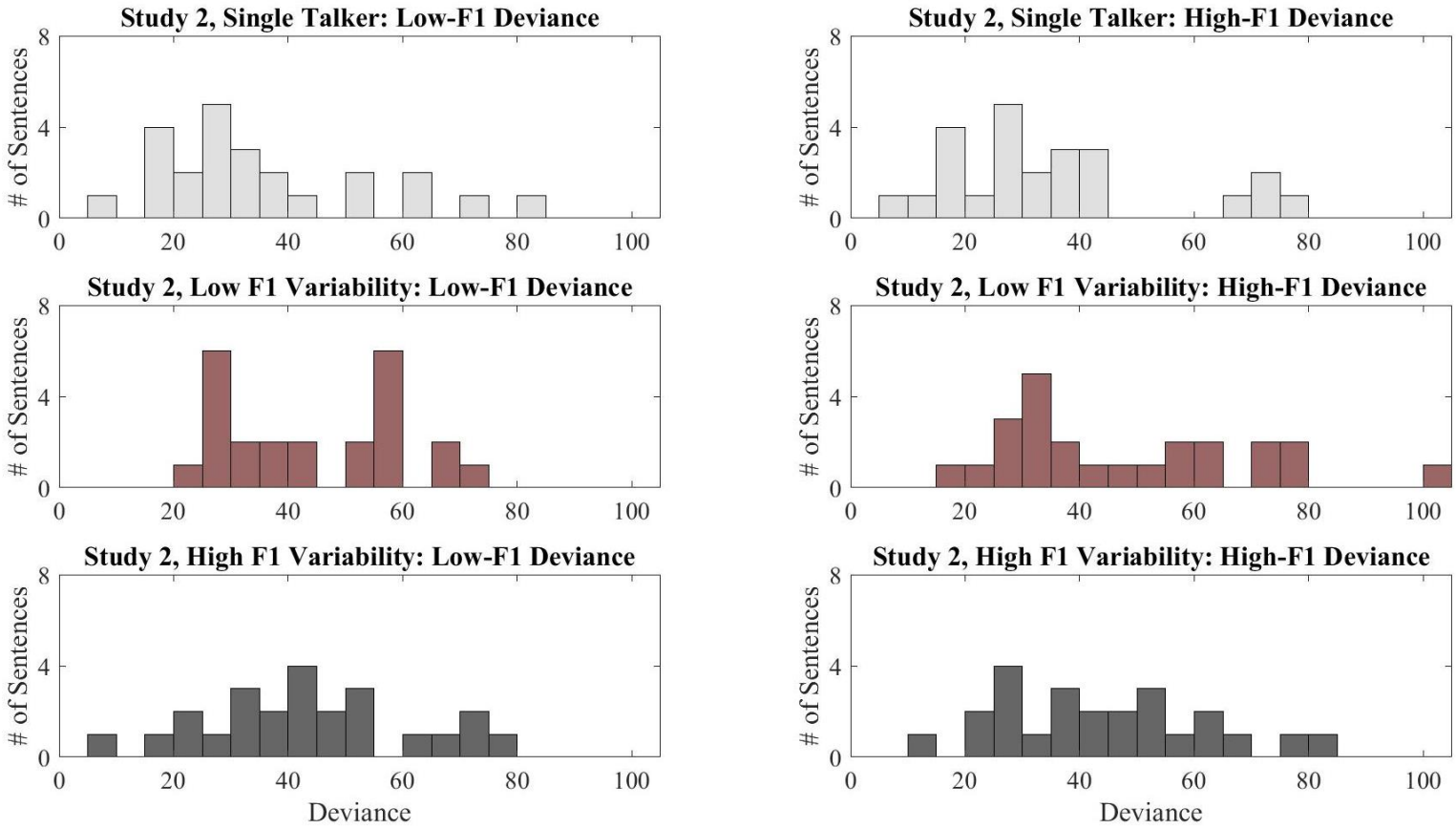

Figure 20. Histograms of deviance measures for the conditions in Study 2. The y-axis displays number of sentences. The x-axis displays deviance measures as output by the glmfit command in Matlab. Low-F1 deviances are on the left and High-F1 deviances are on the right. The top row contains deviance measures for the single talker condition. The middle row contains deviance measures from the Low F1 variability condition. The bottom row contains deviance measures from the High F1 variability condition.

Appendix F.

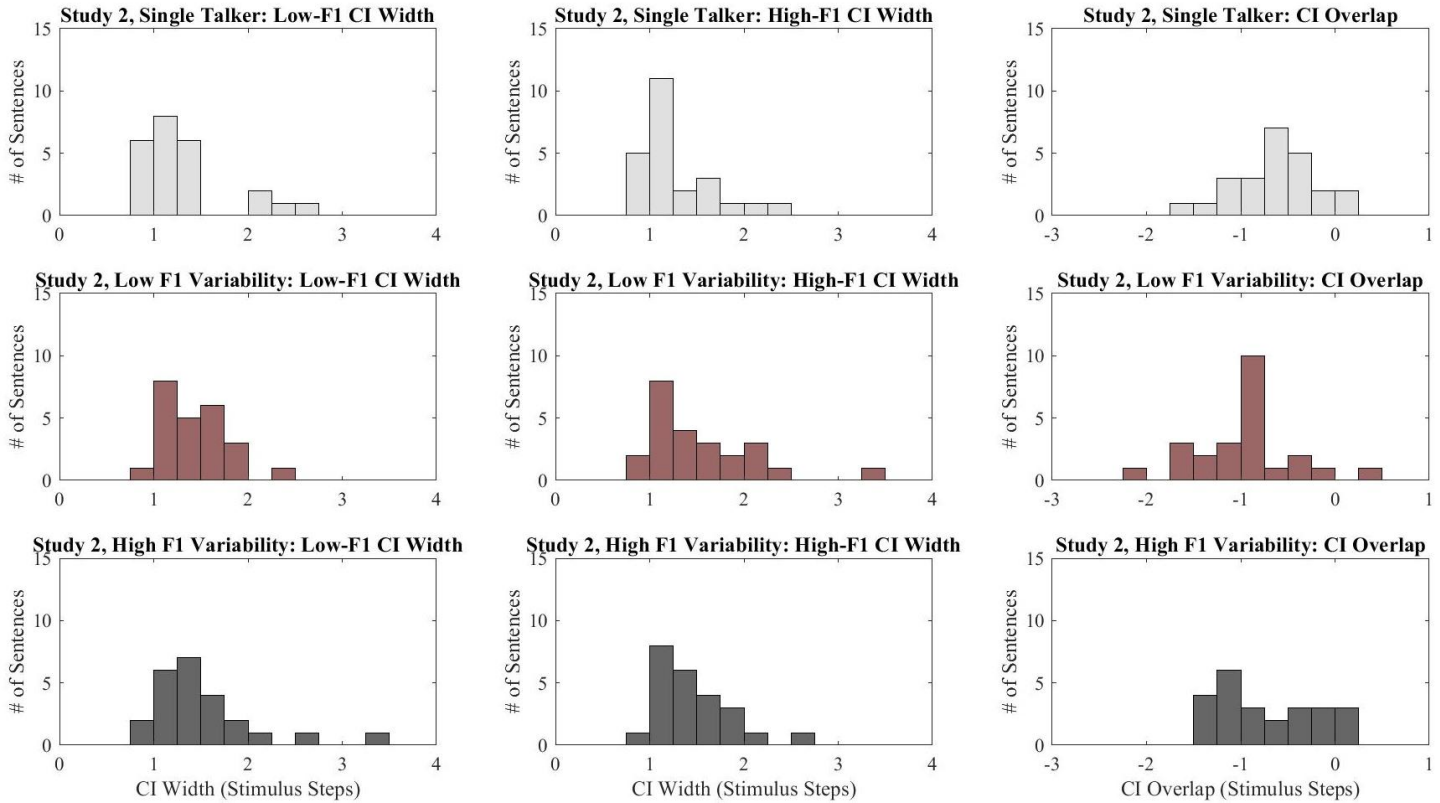

Figure 21. Histograms of confidence intervals around midpoints of logistic functions for the conditions in Study 2. The y-axis displays number of sentences. For the first two columns, the x-axis displays the width of the 95% confidence interval around the midpoint of the logistic function in stimulus steps. The left column contains low-F1 confidence intervals. The middle column contains high-F1 confidence intervals. The x-axis of the right column displays the overlap of the low-F1 and high-F1 CIs (lower bound of high-F1 CI minus upper bound of low-F1 CI). Negative values indicate overlap. The top row contains measures from the single talker condition. The middle row contains measures from the low F1 variability condition. The bottom row contains measures from the high F1 variability condition.

Appendix G.

Table 5

Pairwise t-tests for the main effect of vowel for Study 2

Reference Vowel	Comparison Vowel	Mean Difference	SE	p-value
/ɪ/	2	-2.70	10.51	1.00
	3	-20.69	12.14	1.00
	4	-75.42	14.79	0.002*
	5	-116.18	19.17	<0.001*
	6	-111.45	21.34	0.001*
	7	-69.54	17.48	0.03*
	8	-45.28	14.84	0.26
	9	-19.06	13.19	1.00
	/ɛ/	-4.75	11.54	1.00
2	3	-18.00	11.33	1.00
	4	-72.72	15.24	0.004*
	5	-113.49	18.01	<0.001*
	6	-108.75	18.77	<0.001*
	7	-66.85	14.75	0.007*
	8	-42.59	14.51	0.33
	9	-16.36	11.29	1.00
	/ɛ/	-2.05	11.40	1.00
3	4	-54.73	11.20	0.003*
	5	-95.49	15.93	<0.001*
	6	-90.76	18.21	0.002*
	7	-48.85	14.28	0.11
	8	-24.59	14.00	1.00
	9	1.64	13.45	1.00
	/ɛ/	15.94	11.28	1.00
4	5	-40.76	14.13	0.38
	6	-36.03	17.20	1.00
	7	5.88	13.92	1.00
	8	30.13	15.04	1.00
	9	56.36	14.09	0.03*
	/ɛ/	70.67	12.06	<0.001*
5	6	4.73	9.31	1.00
	7	46.64	12.70	0.06
	8	70.90	16.68	0.01*
	9	97.12	16.32	<0.001*
	/ɛ/	111.43	16.14	<0.001*
6	7	41.91	12.31	0.11
	8	66.16	17.54	0.045*
	9	92.39	18.31	0.002*
	/ɛ/	106.70	17.43	<0.001*
7	8	24.26	12.58	1.00

	9	50.49	14.25	0.08
	/ɛ/	64.79	12.56	0.001*
8	9	26.22	9.26	0.42
	/ɛ/	40.54	8.04	0.002*
9	/ɛ/	14.31	9.81	1.00

Appendix H.

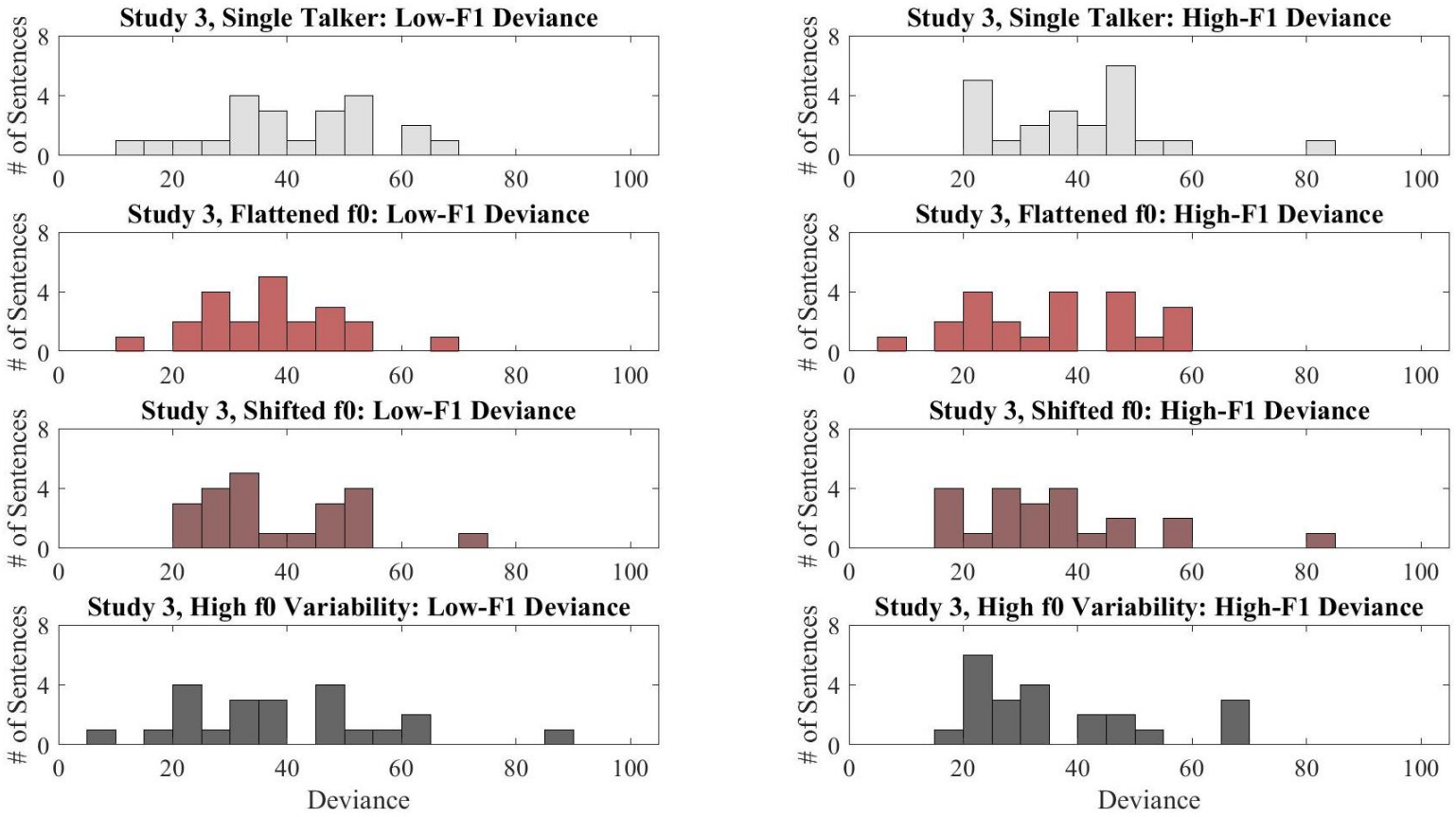

Figure 22. Histograms of deviance measures for the conditions in Study 3. The y-axis displays number of sentences. The x-axis displays deviance measures as output by the glmfit command in Matlab. Low-F1 deviances are on the left and High-F1 deviances are on the right. The top row contains deviance measures for the single talker condition. The second row contains deviance measures from the Flattened f0 condition. The third row contains deviance measures from the Shifted f0 variability condition. The bottom row contains deviance measures from the High f0 variability condition.

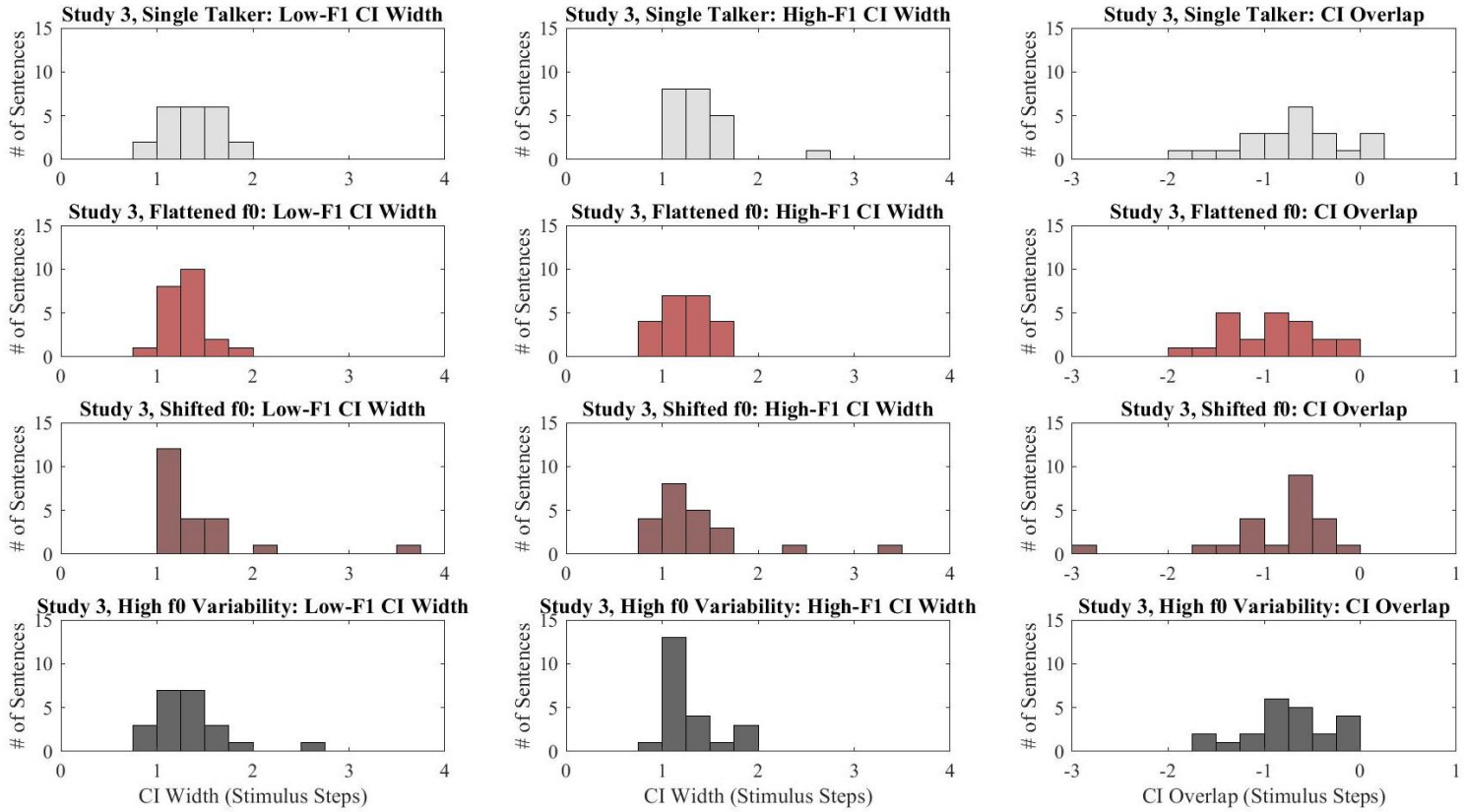

Figure 23. Histograms of confidence intervals around midpoints of logistic functions for the conditions in Study 3.The y-axis displays number of sentences. For the first two columns, the x-axis displays the width of the 95% confidence interval around the midpoint of the logistic function in stimulus steps. The left column contains low-F1 confidence intervals. The middle column contains high-F1 confidence intervals. The x-axis of the right column displays the overlap of the low-F1 and high-F1 CIs (lower bound of high-F1 CI minus upper bound of low-F1 CI). Negative values indicate overlap. The top row contains measures from the single talker condition. The second row contains measures from flattened f0 condition. The third row contains measures from the shifted f0 condition. The bottom row contains measures from the high f0 variability condition.

Appendix J.

Table 6

Pairwise t-tests for the main effect of vowel for Study 3

Reference Vowel	Comparison Vowel	Mean Difference	SE	p-value
/ɪ/	2	-2.29	7.88	1.00
	3	-16.77	10.38	1.00
	4	-62.61	15.64	0.03*
	5	-120.52	22.58	0.001*
	6	-112.54	23.56	0.01*
	7	-62.16	18.85	0.15
	8	-18.10	19.72	1.00
	9	40.61	17.42	1.00
	/ɛ/	23.14	14.96	1.00
2	3	-14.49	8.82	1.00
	4	-60.32	14.49	0.02*
	5	-118.24	21.84	0.001*
	6	-110.25	22.90	0.004*
	7	-59.87	18.01	0.15
	8	-15.81	17.47	1.00
	9	42.89	13.98	0.26
	/ɛ/	25.42	13.16	1.00
3	4	-45.84	13.03	0.09
	5	-103.75	19.77	0.002*
	6	-95.77	20.54	0.01*
	7	-45.39	17.70	0.81
	8	-1.33	17.81	1.00
	9	57.38	16.05	0.08
	/ɛ/	39.91	16.32	1.00
4	5	-57.92	11.52	0.003*
	6	-49.93	16.88	0.34
	7	0.45	16.88	1.00
	8	44.51	19.02	1.00
	9	103.21	19.73	0.002*
	/ɛ/	85.74	18.02	0.01*
5	6	7.99	14.08	1.00
	7	58.36	20.25	0.40
	8	102.42	21.98	0.01*
	9	161.13	24.22	<0.001*
	/ɛ/	143.66	23.57	<0.001*
6	7	50.38	14.81	0.12
	8	94.44	17.49	0.001*
	9	153.14	22.22	<0.001*
	/ɛ/	135.67	23.82	0.001*
7	8	44.06	12.35	0.08

	9	102.77	15.52	<0.001*
	/ɛ/	85.30	15.71	0.001*
8	9	58.70	13.65	0.01*
	/ɛ/	41.24	14.60	0.46
9	/ɛ/	-17.47	10.25	1.00

Appendix K.

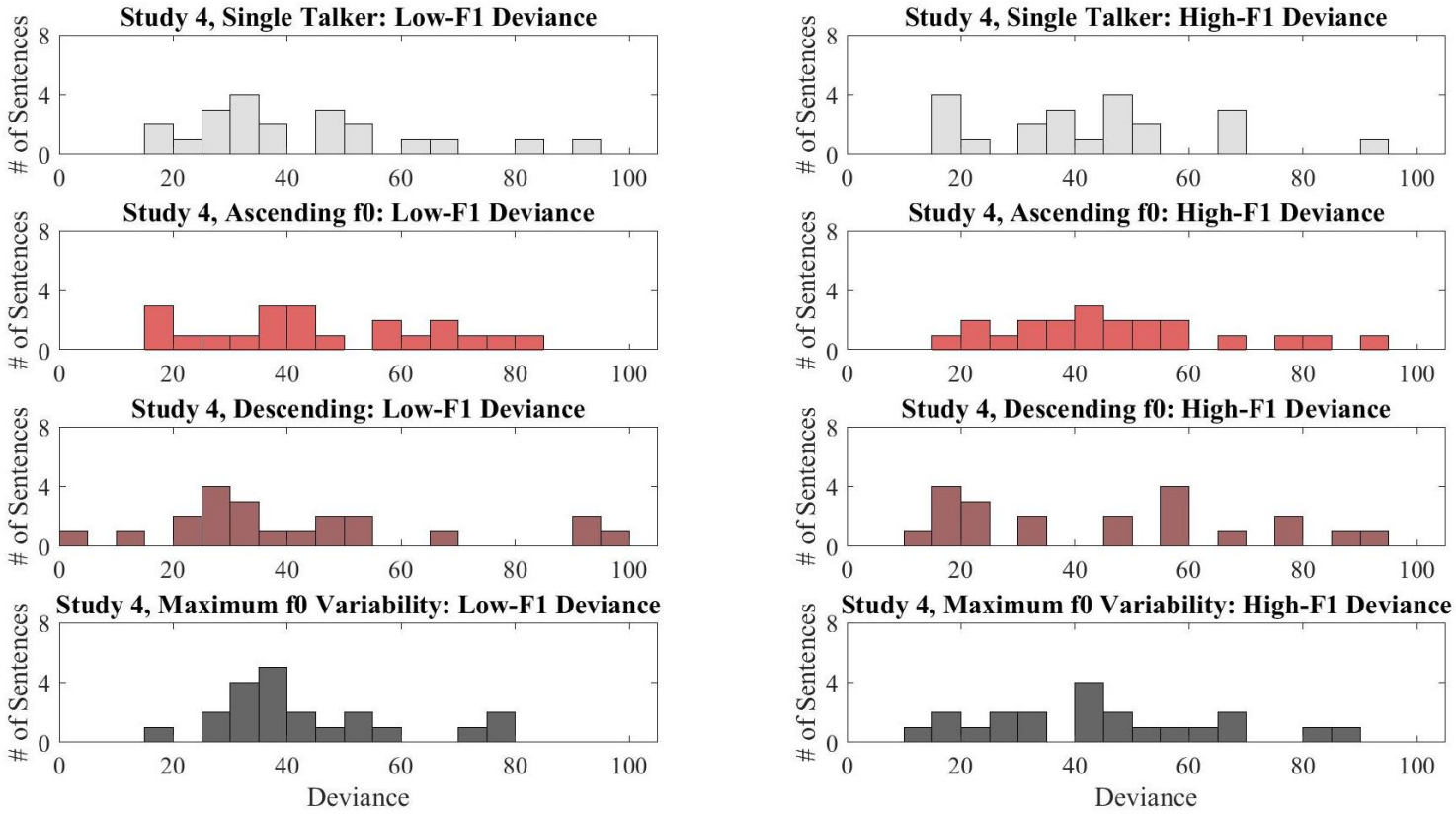

Figure 24 Histograms of deviance measures for the conditions in Study 4. The y-axis displays number of sentences. The x-axis displays deviance measures as output by the glmfit command in Matlab. Low-F1 deviances are on the left and High-F1 deviances are on the right. The top row contains deviance measures for the single talker condition. The second row contains deviance measures from the ascending f0 condition. The third row contains deviance measures from the descending f0 variability condition. The bottom row contains deviance measures from the maximum f0 variability condition.

Appendix L.

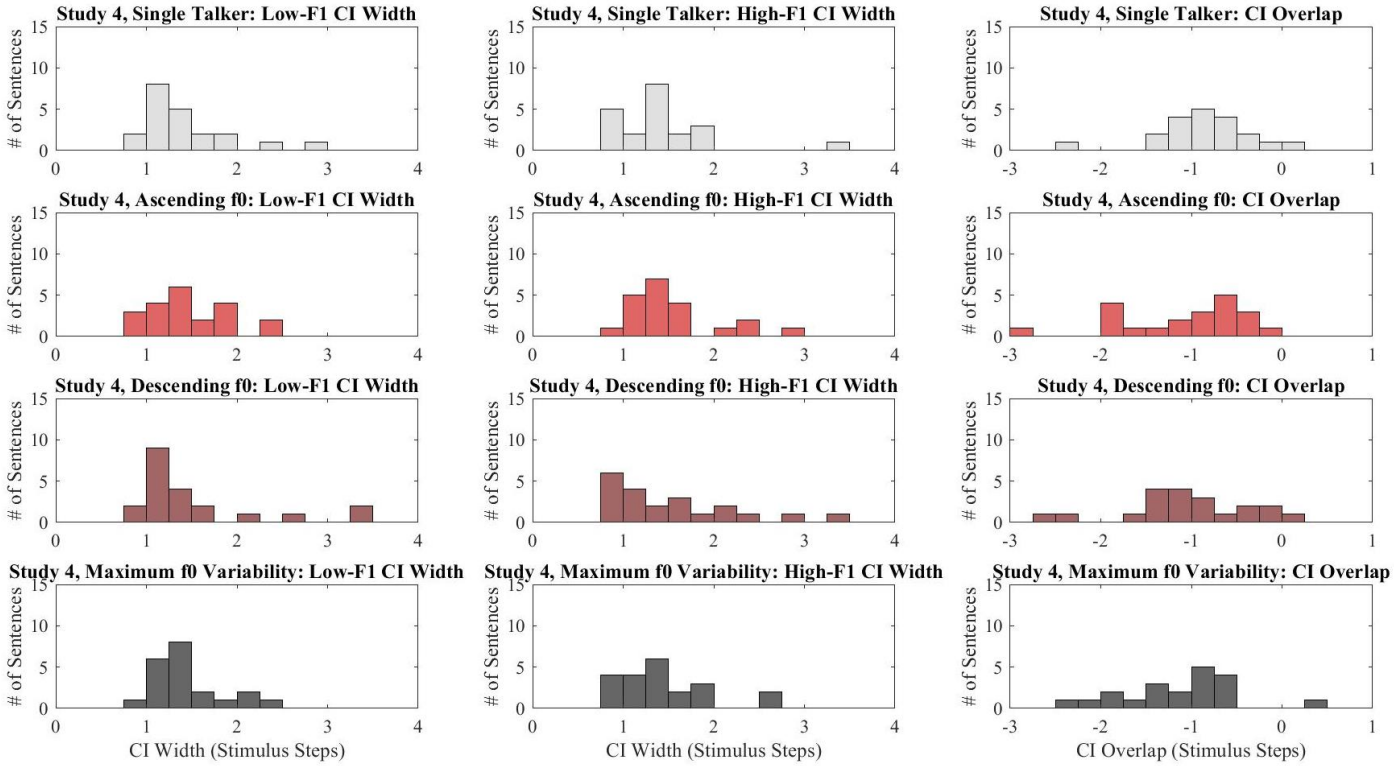

Figure 25 Histograms of confidence intervals around midpoints of logistic functions for the conditions in Study 4. The y-axis displays number of sentences. For the first two columns, the x-axis displays the width of the 95% confidence interval around the midpoint of the logistic function in stimulus steps. The left column contains low-F1 confidence intervals. The middle column contains high-F1 confidence intervals. The x-axis of the right column displays the overlap of the low-F1 and high-F1 CIs (lower bound of high-F1 CI minus upper bound of low-F1 CI). Negative values indicate overlap. The top row contains measures from the single talker condition. The second row contains measures from the ascending f0 condition. The third row contains measures from the descending condition. The bottom row contains measures from the maximum f0 variability condition.

Appendix M.

Table 7

Pairwise t-tests for the main effect of vowel for Study 4

Reference Vowel	Comparison Vowel	Mean Difference	SE	p-value
/ɪ/	2	6.74	8.44	1.00
	3	-0.51	11.64	1.00
	4	-41.02	13.31	0.27
	5	-73.78	23.37	0.22
	6	-81.40	21.78	0.06
	7	-41.61	22.17	1.00
	8	-4.00	18.62	1.00
	9	25.83	15.70	1.00
	/ɛ/	41.12	16.80	1.00
2	3	-7.25	7.06	1.00
	4	-47.76	8.48	0.001*
	5	-80.52	19.31	0.02*
	6	-88.14	19.13	0.01*
	7	-48.35	20.74	1.00
	8	-10.74	15.80	1.00
	9	19.09	13.10	1.00
	/ɛ/	34.38	14.42	1.00
3	4	-40.51	8.44	0.01*
	5	-73.27	18.67	0.04*
	6	-80.89	19.71	0.03*
	7	-41.10	20.62	1.00
	8	-3.49	16.42	1.00
	9	26.34	14.38	1.00
	/ɛ/	41.63	15.85	0.73
4	5	-32.76	15.96	1.00
	6	-40.38	18.88	1.00
	7	-0.60	18.97	1.00
	8	37.02	15.00	1.00
	9	66.85	13.46	0.003*
	/ɛ/	82.14	13.64	<0.001*
5	6	-7.62	10.76	1.00
	7	32.17	13.66	1.00
	8	69.78	14.72	0.01*
	9	99.61	13.86	<0.001*
	/ɛ/	114.90	16.07	<0.001*
6	7	39.78	11.17	0.09
	8	77.39	13.06	<0.001*
	9	107.23	12.79	<0.001*
	/ɛ/	122.52	15.86	<0.001*
7	8	37.61	13.76	0.58

	9	67.44	15.63	0.02*
	/ɛ/	82.74	18.93	0.01*
8	9	29.83	9.24	0.19
	/ɛ/	45.13	10.82	0.02*
9	/ɛ/	15.29	6.82	1.00

Appendix N.

Table 9

Pairwise t-tests for the interaction for Study 4

Vowel	Reference Condition	Comparison Condition	Mean Difference	SE	p-value
/ɪ/	1	2	-84.67	30.39	0.07
		3	-9.58	19.35	1.00
		4	-141.97	28.70	<0.001*
	2	3	75.09	31.71	0.17
		4	-57.30	40.72	1.00
	3	4	-132.39	28.02	.001*
2	1	2	-46.07	31.40	0.95
		3	15.47	25.34	1.00
		4	-123.95	23.16	<0.001*
	2	3	61.54	33.45	0.48
		4	-77.88	32.62	0.16
	3	4	-139.42	21.42	<0.001*
3	1	2	-43.74	33.32	1.00
		3	19.62	30.81	1.00
		4	-90.79	31.82	0.06*
	2	3	63.36	32.31	0.38
		4	-47.05	28.69	0.70
	3	4	-110.41	26.38	0.003*
4	1	2	-25.54	28.22	1.00
		3	63.81	25.50	0.13
		4	-81.58	26.33	0.03*
	2	3	89.36	32.04	0.07
		4	-56.04	31.15	0.52
	3	4	-145.40	26.69	<0.001*
5	1	2	17.36	27.71	1.00
		3	42.99	24.51	0.57
		4	-6.27	33.10	1.00
	2	3	25.63	36.59	1.00
		4	-23.62	38.81	1.00
	3	4	-49.26	30.29	0.72
6	1	2	30.65	37.91	1.00
		3	-21.67	28.89	1.00
		4	-10.96	37.14	1.00
	2	3	-52.32	31.18	0.65
		4	-41.61	29.54	1.00
	3	4	10.71	34.83	1.00
7	1	2	15.21	32.95	1.00
		3	1.23	29.33	1.00
		4	-46.83	29.05	0.74

	2	3	-13.98	33.08	1.00
		4	-62.04	33.61	0.48
	3	4	-48.06	27.87	0.60
8	1	2	0.817	26.44	1.00
		3	-16.23	30.36	1.00
		4	-34.42	28.07	1.00
	2	3	-17.04	25.99	1.00
		4	-35.24	20.74	0.63
	3	4	-18.20	30.48	1.00
9	1	2	5.86	16.63	1.00
		3	-31.40	30.44	1.00
		4	-14.81	28.59	1.00
	2	3	-37.25	24.57	0.87
		4	-20.67	21.23	1.00
	3	4	16.59	23.83	1.00
/ɛ/	1	2	-12.03	30.70	1.00
		3	-49.16	24.61	0.36
		4	-36.55	25.00	0.96
	2	3	-37.13	29.72	1.00
		4	-24.51	19.71	1.00
	3	4	12.62	26.53	1.00

Appendix O.

Table 9

Proportion accuracy in all conditions of all studies

Study 1b	M	SE	Study 2	M	SE	Study 3	M	SE	Study 4	M	SE
Single	0.97	0.01	Single	0.98	0.01	Single	0.98	0.01	Single	0.98	0.01
Low f0 Variability	0.96	0.01	Low F1 Variability	0.97	0.01	Flattened f0	0.99	0.00	Ascending	0.98	0.01
High f0 Variability	0.98	0.01	High F1 Variability	0.96	0.01	Shifted f0	0.97	0.02	Descending	0.96	0.02
						High f0 Variability	0.98	0.01	Maximum Variability	0.97	0.01

Appendix P.

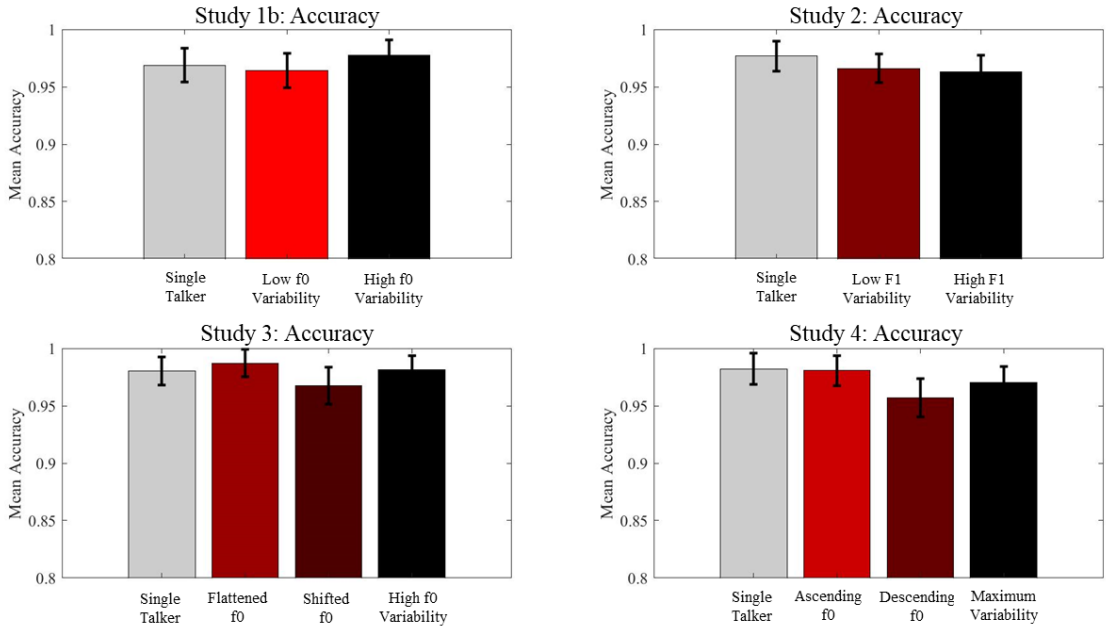

Figure 26. Accuracy at endpoints in all studies. f0 variability conditions are represented on the x-axis. Accuracy in proportion correct is represented on the y-axis. The y-axis only encompasses the range between 80% and 100% since listeners with less than 80% accuracy were removed from further analyses. The gray bars represent the single talker conditions. The red bars represent the low variability conditions. The black bars represent the high variability condition. Error bars depict standard error of the mean.

CURRICULUM VITAE

Ashley (Lily) Assgari
Email: ashley.assgari@louisville.edu
Phone: (443) 340-5151

Permanent Address:
2035 Periwinkle Dr.
Finksburg, MD 21048

Current Address:
1211 S 2nd St. Apt 3
Louisville, KY 40203

Education:

Doctorate of Philosophy, Expected May 2018
Master of Science, May 2016
Psychological and Brain Sciences, Vision and Hearing Sciences
University of Louisville
Louisville, KY

Master of Arts, May 2014, Psychological Sciences, Experimental Psychology
Bachelor of Arts, Cum Laude, May 2011, Psychology, Cumulative GPA: 3.62
James Madison University
Harrisonburg, VA

Honors and Organizations:

Experimental Psychology Award for Excellence in Service, 2016-2017
Best Student Advocate, Student Government Superlatives, 2016-2017
Graduate Student Council Travel Award, Summer 2016
The Graduate Dean's Citation Award, Spring 2016
Graduate Student Council Travel Award, Summer 2015
Graduate Student Council Travel Award, Fall 2014
Research Fellowship, University of Louisville, 2014-2016
Out-of-state Graduate Assistantship for Psychological Sciences Program, 2013-2014
Outstanding Scholarship Award, Spring 2013
Awarded Outstanding Undergraduate Research Travel Grant, Spring 2013
In-state Graduate Assistantship for Psychological Sciences Program, 2012-2013
Dean's List, Spring 2008-Fall 2009, Fall 2010-Spring 2011
Student Member
Kentucky Academy of Science
Acoustical Society of America
Member of Association for Psychological Science (Graduate Student Affiliate)

Psi Chi, Psychology Honor Society
Golden Key International Honour Society

Research Experience:

Auditory Perception and Processing Lab, Fall 2014-present, University of Louisville
Auditory Perception Laboratory, Fall 2012-Spring 2014, James Madison University
Animal Research Facility, Fall 2012- Spring 2014, James Madison University
The Cognitive Science Lab, Spring 2011, James Madison University

---

Teaching Experience:

Spring 2018, Graduate Teaching Assistant, Psych 611, Advanced Statistics II
Fall 2017, Guest Lecture, Psyc 610, Advanced Statistics I, Independent Samples T-test
Fall 2017, Graduate Teaching Assistant, Psych 610, Advanced Statistics I
Summer 2017, Guest Lecture, Psyc 301, Quantitative Methods in Psychology, Independent Samples t-test
Spring 2017, Guest Lecture, Psyc 301, Quantitative Methods in Psychology, One-way t-tests
Fall 2016/Spring 2017, Graduate Teaching Assistant, Psyc 301, Quantitative Methods in Psychology
Fall 2015, Guest Lecture, Psyc 331, Sensation and Perception, Speech Perception
Fall 2013/Spring 2014, Teaching Assistant, Psyc 211, Psychological Research Methods
Summer 2013, Teaching Assistant, Psyc 210, Psychological Measurement and Statistics
Spring 2013/Spring 2014, Teaching Assistant, GPsyc 122, The Science of Vision and Audition
Fall 2012/Spring 2013, Teaching Assistant, Psyc 210, Psychological Measurement and Statistics

---

Service Experience:

Spring 2018, Student Member, Decanal Review Committee
Represent the student perspective in the assessment of the effectiveness of Dean Beth Boehm
Fall 2017-Spring 2018, Student Member, Graduate Faculty Council
Represent the Graduate Student Council and student interests to Graduate Faculty
Fall 2017-Spring 2018, Student Member, Budget Advisory Committee
Represent student interests regarding budgetary planning and decision-making
Helped design a scoring matrix for objective comparisons of similar budget proposals
Fall 2017-Spring 2018, Student Member, Taskforce on Tuition and Fees
Represent student interests regarding tuition and fees
Fall 2017-Spring 2018, Senate Speaker, Student Government Association
Lead Student Government Senate meetings through parliamentary procedure
Serve as the primary contact for Senate related matters
Fall 2017-Spring 2018, President, Graduate Student Council
Serve as the primary contact for Graduate Student Council related matters
Oversee Graduate Student Council spending and budgeting
Prepare monthly meeting agendas
Lead Graduate Student Council meetings through parliamentary procedure
Fall 2016-Spring 2018, Senator, Student Government Association

Represent the graduate student perspective in Student Government Association

Fall 2016-Spring 2017, Director of Graduate Travel, Graduate Student Council
Awarded Travel Grants to Graduate Students
Balanced funding sources

Fall 2016-Spring 2017, Member, Student Government Appropriations Committee
Approved the disbursal of Student Government Association funds

Fall 2015-Spring 2017, Graduate Student Representative, Graduate Student Council
Represented the Department of Psychological and Brain Sciences in the Graduate Student Council

Fall 2012-Spring 2014, Graduate Assistant, Psychological Sciences
Built Psychological Sciences SharePoint webpage
Assisted in the coordination of Psychological Sciences Interview Day
Maintained updates to Psychological Sciences Websites
Prepared task lists for incoming Psychological Sciences graduate assistants

Fall 2012, Graduate Assistant, Department of Psychology
Contacted Alumni
Updated alumni profiles on the Department of Psychology Webpage
Alumni Data Entry in ACCESS

Publications:

Assgari, A. A. (in preparation). Assessing the Relationship between Talker Normalization and Spectral Contrast Effects in Speech Perception. (Dissertation), University of Louisville.

Assgari, A. A., Theodore, R.M., & Stilp, C.E. (in preparation). Differential effects of talker acoustic variability on context effects in speech perception.

Stilp, C.E. & Assgari, A.A. (in press). Perceptual sensitivity to spectral properties in earlier sounds during speech categorization. *Attention, Perception, & Psychophysics.*

Stilp, C.E., & Assgari, A.A. (2017). Consonant categorization exhibits graded influence of surrounding spectral context. *Journal of the Acoustical Society of America, 141*(2), EL153-EL158.

Stilp, C.E., Anderson, P.W., Assgari, A.A., Ellis, G.M., & Zahorik, P. (2016). Speech perception adjusts to reliable spectrotemporal properties in the listening environment. *Hearing Research, 341*, 168-178.

Chan, K. Y., Hall, M. D., & Assgari, A. A. (2016). The role of vowel formant frequency and duration in the perception of foreign accent. *Journal of Cognitive Psychology*.

Assgari, A.A., & Stilp, C.E. (2015). Talker information influences spectral contrast effects in speech categorization. *Journal of the Acoustical Society of America, 138*(5), 3023-3032.

Stilp, C.E., & Assgari, A.A. (2015). Languages across the world are efficiently coded by the auditory system. *Proceedings of Meetings on Acoustics, 23*, 060003.

Assgari, A. A. (2014). Effects of temporal parameters on the perception of foreign accent in synthesized speech. (Master's thesis), James Madison University.

Presentations and Abstracts:

Assgari, A.A. & Stilp, C.E. (2018, May). "Trial-to-trial variability in talkers' fundamental frequencies restrains spectral context effects in vowel categorization." Paper accepted for presentation at the 175th Meeting of the Acoustical Society of America, Minneapolis, Minnesota.

Assgari, A.A., Frazier, J.M., & Stilp, C.E. (2018, May). “Musical instrument categorization is highly sensitive to spectral properties of earlier sounds.” Paper accepted for presentation at the 175th Meeting of the Acoustical Society of America, Minneapolis, Minnesota.

Stilp, C.E. & Assgari, A.A. (2018, May). “Natural signal statistics and the timecourse of spectral context effects in consonant categorization.” Paper accepted for presentation at the 175th Meeting of the Acoustical Society of America, Minneapolis, Minnesota.

Stilp, C.E. & Assgari, A.A. (2017, December). “Filtered and unfiltered sentences produce different spectral context effects in vowel categorization.” Paper submitted for presentation at the 174th Meeting of the Acoustical Society of America, New Orleans, Louisiana.

Assgari, A.A., Theodore, R.M., & Stilp, C.E. (2017, June). “Isolating sources of acoustic variability that diminish spectral contrast effects in vowel categorization.” Poster presented at the 173rd Meeting of the Acoustical Society of America, Boston, Massachusetts.

Assgari, A.A., Mohiuddin, A., Theodore, R.M., & Stilp, C.E. (2016). “Dissociating contributions of talker gender and acoustic variability for spectral contrast effects in vowel categorization.” Poster presented at the 171st Meeting of the Acoustical Society of America, Salt Lake City, Utah.

Stilp, C.E., Anderson, P.W., Assgari, A.A., Ellis, G.M., & Zahorik, P. (2015). “Reverberation increases perceptual calibration to reliable spectral peaks in speech.” Presented at the 169th Meeting of the Acoustical Society of America, Pittsburgh, PA.

Stilp, C.E., & Assgari, A.A. (2015). “Languages across the world are efficiently coded by the auditory system”. Presented at the 169th Meeting of the Acoustical Society of America, Pittsburgh, PA.

Assgari, A.A., & Stilp, C.E. (2015). “Talker normalization and acoustic properties both influence spectral contrast effects in speech perception.” Presented at the 169th Meeting of the Acoustical Society of America, Pittsburgh, PA.

Stilp, C.E., & Assgari, A.A. (2015). “Does the auditory system efficiently code all languages or just American English?” Paper presented at the 38th Annual MidWinter Meeting of the Association for Research in Otolaryngology, Baltimore, Maryland.

Chan, K. Y., Hall, M.D., & Assgari, A. (2014) An evaluation of the role of vowel formant frequencies and vowel duration on the perception of foreign accent. Poster presented at the Auditory Perception Cognition Action Meeting conference in Long Beach, California.

Assgari, A., & Hall, M. D. (2013) Effects of vocal source characteristics on vowel perception. Talk presented at the Auditory Perception Cognition Action Meeting conference in Long Toronto, Cananda.

Chan, K. Y., Hall, M.D., & Assgari, A. (2013) An evaluation of the role of vowel formant frequencies on the perception of foreign accent. Poster presented at the Auditory Perception Cognition Action Meeting conference in Toronto, Canada.

McVay, S., O'Malley, J.J., Shemery, A.M., Assgari A.A., Clasen, M, Sequeira, S.N., Crosby, T.R., Tiry, A.M., Whitehurst, L., Holt, D.D., & Dyche, J. (2013). Effects of brief paradoxical sleep deprivation on radial arm maze performance in spontaneously hypertensive rats. Sleep, 35, 0252.

Assgari, A., Becker, C., & Hall, M. D. (2013) Effects of vocal source characteristics on vowel perception. Poster presentation at the Association of Psychological Sciences conference in Washington, D.C.

McVay, S., O’Malley, J. J. , Shemery, A., Assgari, A., Tiry, A., Clasen, M., Sequerira, S., Cooke, C., Yohn, C,. Houhoulis, S., Williams, D., Jeter, A., Dyche, J., Holt, D. (2013). Strategic

Differences on a Radial Arm Maze Task in a Rodent Model of ADHD. Poster presentation at the Association of Psychological Sciences conference in Washington, D.C.

McVay, S., Clasen, M., Assgari, A., Gross, J., Meccariello, M., Vassallo, M., O'Malley, J.J., Tiry, A., Williamson, C., Holt, D., Dyche, J. (2013) Effects of Brief Paradoxical Sleep Deprivation on Radial Arm Maze Performance. Poster presentation at the meeting of Virginia Association for Behavior Analysis, Harrisonburg, VA

Hall, M. D., Becker, C., Redpath, T., & Assgari, A. (2012) Auditory Stimulus Generation Tools in MaxforLive, Poster presentation at the Auditory Perception and Cognition Action Meeting in Minneapolis, MN.