ACOUSTICS AND PERCEPTION OF CLEAR FRICATIVES

By

Kazumi Maniwa

B.A. Shimane University, Japan, 1995

M.A., University of Kansas, 2002

Submitted to the Department of
Linguistics and the Faculty of the Graduate School of the University
of Kansas in partial fulfillment of the requirements for the degree of Doctor of
Philosophy

____

____

____

____

____

Date defended: ____________________________

The Dissertation Committee for Kazumi Maniwa certifies that this is the approved version of the following dissertation:

ACOUSTICS AND PERCEPTION OF CLEAR FRICATIVES

Committee:

____

Chairperson*

____

____

____

____

Date approved: ___________________________

Abstract

Everyday observation indicates that speakers can naturally and spontaneously adopt a speaking style that allows them to be understood more easily when confronted with difficult communicative situations. Previous studies have demonstrated that the resulting speaking style, known as clear speech, is more intelligible than casual, conversational speech for a variety of listener populations. However, few studies have examined the acoustic properties of clearly produced fricatives in detail. In addition, it is unknown whether clear speech improves the intelligibility of fricative consonants, or how its effects on fricative perception might differ depending on listener population. Since fricatives are the cause of a large number of recognition errors both for normal-hearing listeners in adverse conditions and for hearing-impaired listeners, it is of interest to explore these issues in detail focusing on fricatives. The current study attempts to characterize the type and magnitude of adaptations in the clear production of English fricatives and determine whether clear speech enhances fricative intelligibility for normal-hearing listeners and listeners with simulated impairment.

In an acoustic experiment (Experiment I), ten female and ten male talkers produced nonsense syllables containing the fricatives /f, θ, s, ʃ, v, δ, z, and ʒ/ in VCV contexts, in both a conversational style and a clear style that was elicited by means of simulated recognition errors in feedback received from an interactive computer program. Acoustic measurements were taken for spectral, amplitudinal, and temporal properties known to influence fricative recognition. Results illustrate that (1) there were consistent overall clear speech effects, several of which (consonant duration, spectral peak location, spectral moments) were consistent with previous findings and a few (notably consonant-to-vowel intensity ratio) which were not, (2) 'contrastive' differences related to acoustic inventory and eliciting prompts were observed in key comparisons, and (3) talkers differed widely in the types and magnitude of acoustic modifications.

Two perception experiments using these same productions as stimuli (Experiments II and III) were conducted to address three major questions: (1) whether clearly produced fricatives are more intelligible than conversational fricatives, (2) what specific acoustic modifications are related to clear speech intelligibility advantages, and (3) how sloping, recruiting hearing impairment interacts with clear speech strategies. Both perception experiments used an adaptive procedure to estimate the signal to (multi-talker babble) noise ratio (SNR) threshold at which minimal pair fricative categorizations could be made with 75% accuracy. Data from fourteen normal-hearing listeners (Experiment II) and fourteen listeners with simulated sloping elevated thresholds and loudness recruitment (Experiment III) indicate that clear fricatives were more intelligible overall for both listener groups. However, for listeners with simulated hearing impairment, a reliable clear speech intelligibility advantage was not found for non-sibilant pairs. Correlation analyses comparing acoustic and perceptual style-related differences across the 20 speakers encountered in the experiments indicated that a shift of energy concentration toward higher frequency regions and greater source strength was a primary contributor to the

“clear fricative effect” for normal-hearing listeners but not for listeners with simulated loss, for whom information in higher frequency regions was less audible.

Acknowledgements

Though only my name appears on the cover of this dissertation, I could never have reached the heights or explored the depths without the help, support, guidance, and efforts of a lot of people. I am very grateful that they conspired to drag this author kicking and screaming towards her goal and would like to thank these people for their encouragement, patience, and understanding during the completion of work.

My sincere gratitude goes to the chief advisor for this study, Dr. Allard Jongman. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. Dr. Jongman has been an inspiration to me and is the reason I began studying acoustic phonetics and speech perception. His wisdom, patience and encouragement helped me overcome many crisis situations and finish this dissertation.

Likewise, Dr. Joan Sereno has been a great source of guidance and encouragement. Her precious comments and insightful directions helped me sharpen my ideas and critical tools. If it were not for her unending support and immense generosity, the dissertation work could not have been accomplished.

I am also deeply indebted to my committee members for their time, energy, useful comments and suggestions. I have benefited greatly from their advice and the points of vies from Phonology, Speech-language-hearing, and Psycholinguistics. I am very thankful to Dr. Jie Zhang for his invaluable insights and criticism, especially for his challenging and interesting questions during my defenses. His advice and suggestions will be of great help as I begin my research career. I would like to express my gratitude to Dr. Sarah Hargus Ferguson for providing me excellent advice on clear speech and speech-language-hearing and generous and insightful comments, which helped me to shape many of the ideas presented here. I am also thankful to Dr. Mike Vitevitch for sharing his scientific expertise with me and offering important suggestions during my defense.

I owe a special note of gratitude to my friends and colleagues at the Kansas University Phonetics and Psycholinguistics Laboratory and the Department of Linguistics for their intellectual and emotional support. My special thanks are to Mohammad Al-Masri, Yuwen Lai, Ricardo Benitez and Carmen Parafita-Couto for their contribution to the joyful time I had during my school years and sharing me with their views.

I also would like to extend my appreciation to people in the Phonology Lab at the University of California, Berkeley, especially to Dr. John Ohala and Dr. Keith Johnson for generously hosting me as a visiting student scholar in their lab and for providing me valuable suggestions and criticism. I also would like to mention my gratitude to Reiko Kataoka for her continuous encouragement and sharing her

linguistic knowledge. She is incredibly generous who always listened to my 愚痴. Her wit and smartness, and enthusiasm to research and strength always amazed me.

My very special thanks go to my parents, Katsuyoshi and Kiyoko Maniwa, and sister, Yasuko Maniwa, whom I owe everything I am today. Their unwavering faith and confidence in my abilities and in me is what has shaped me to be the person I am. They have continued to provide me with all kinds of unconditional support during all of my academic education and during my whole life. They are my inspiration and teachers in my life.

The last, and surely the most, I want to thank my husband, Travis Wade, for his wisdom, patience, and care. If it were not for his support, I would have been lost before the completion of the research. He has been a great ally, mentor, and rival as a researcher in the same field. He has been a friend and family to me, too. He cheered me up when I was stressed out. He smiled when I needed energy. He gave me a tremendous amount of the insightful criticism and suggestions for my work. He has been always there to sustain me when I was falling down. Thank you, Travis.

Chapter I

Introduction

1.1 Introduction

In everyday conversation, talkers speak casually and freely, often coarticulating heavily and without much concern for their enunciation or for the possibility that they might be misunderstood. However, talkers can naturally and spontaneously adopt intelligibility-enhancing articulatory strategies when they are aware of or experience communicative difficulties on the part of a listener. On the basis of experimental data, Lindblom and colleagues have stipulated that speakers make moment-by-moment estimates of listeners' needs for explicit signal information, and continuously adapt their speech production to address perceived listener demands while minimizing articulatory effort (Lindblom, 1990, 1996; Lindblom *et al.*, 1992). According to this theory (e.g., Lindblom, 1990), speech production is adaptively organized along a continuum of context-/knowledge-dependent variability from system-oriented, relaxed, conversational *hypo-speech* to output-oriented, energetic, clarified *hyper-speech*. Speakers, then, operate on the principle of yielding *sufficient discriminatory information* for a listener to recover their intended message while at the same time striving for low-cost form of articulatory performance. When speakers perceive or experiences no particular threat to their listeners' ability to identify speech sounds and reconstruct their intended meaning, the motor control system attempts to minimize the physical cost. When a speaker perceives or experiences listener's difficulty in comprehension, due to, e.g., background noise, reverberation, hearing impairment, or lack of linguistic/world knowledge, speakers will adapt their speech to deliver more explicit signal information and increase the perceptual distance.

In line with these theoretical notions, there is evidence from a variety of studies in laboratory settings that, when speakers are instructed to speak as though talking to listeners with hearing impairment, aged listeners, or non-native listeners, their productions do in fact become more intelligible (e.g., Bradlow and Bent, 2002; Bradlow *et al.*, 2003; Ferguson and Kewley-Port, 2002; Gagné *et al.*, 1994, 1995, 2002; Helfer, 1997, 1998; Iverson and Bradlow, 2002; Krause and Braida, 2002; Liu *et al.*, 2004; Payton *et al.*, 1994; Picheny *et al.*, 1985; Schum, 1996; Uchanski *et al.*, 1996). The intelligibility advantage of *clear speech* over *conversational speech* averaged across talkers, listeners, stimulus types, and experimental conditions such as speech-to-noise ratio has ranged from 3 to 38 percentage points across experiments.

A fair amount of previous research has been dedicated to the acoustic properties of clear speech, as well. In general, most studies have focused on either global (sentence-level) measurements or phonetic modifications of vowel sounds. There have been very few studies that investigated acoustic changes found in the fine acoustic-phonetic components of consonant sounds in detail. Phonetic modifications of consonant sounds reported so far have been mostly limited to temporal characteristics such as larger distinction of the voice-onset-time (VOT) between voiced and voiceless stops and longer duration of plosives, fricatives, nasals, and semi-vowels, and amplitudinal characteristics such as increased consonant-to-vowel ratio (CVR). Similarly, clear speech intelligibility benefits were reported for sentence (*e.g.* Bradlow and Bent, 2002; Bradlow *et al.*, 2003; Helfer, 1997, 1998; Liu *et al.*, 2004), CV/VCV-syllable (*e.g.* Chen, 1980; Gagné *et al.*, 2002) and vowel stimuli (*e.g.* Ferguson and Kewley-Port, 2002*),* but few previous studies have investigated whether and to what extent clear speech enhances the intelligibility of specific consonants, e.g. fricatives and plosives. In addition, it is not clear yet which acoustic-phonetic changes made by speakers in clear speech may be responsible for the intelligibility advantage, since previous studies either used few speakers (Bradlow *et al.*, 2003; Chen, 1980) or presented limited acoustic data (Gagné *et al*, 1994, 1995, 2002; Schum, 1996; *cf.* Ferguson, 2002), especially for consonants. Considering that one of the most common perceptual errors in noisy environments or reduced presentation levels (Dubno and Levitte, 1981; Miller and Nicely, 1955; Singh and Black, 1966; Soli and Arabie, 1979; Wang and Bilger, 1973) and by hearing-impaired listeners (Bilger and Wang, 1976; Dubno et al., 1982) involves place of consonant articulation, it is important to describe which acoustical modifications occur when speakers try to make consonant sounds more intelligible and how these changes affect listeners' perception, especially in adverse situations.

The present study was designed to determine whether and which systematic changes in the production of English fricatives talkers tend to make as a result of online feedback indicating comprehension difficulty, and to assess the range of effects that these strategies actually have on the intelligibility of their productions. Fricative consonants were selected for analysis since previous studies on consonant recognition and confusion have reported that fricative sounds, especially non-sibilants, are less distinguishable for normal-hearing listeners in noise. Likewise, listeners with sensorineural hearing loss often have difficulty perceiving place of articulation for fricatives with a preponderance of high-frequency energy (Boothroyd, 1984; Dubno et al., 1982; Owens, 1978; Owens et al., 1972; Sher and Owens, 1974; Hedrick, 1997; Zeng and Turner, 1990).

The research described here introduces a new technique for eliciting conversational and clear productions of fricative sounds, in order to examine more carefully than has been possible in previous studies the different types of adaptations that might be made in specific situations. The method replicates and extends the one described by Ohala (1994), in which talkers produce and then repeat a target stimulus after receiving pseudo-response feedback indicating that it has been misidentified as a similar, confusable sound. Given the claims of *hyper-hypo* speech (H&H) theory, it seems likely that talkers might make effort not only to improve the intelligibility of the target stimulus but also to make it sound more unlike the sound it was mistaken for.

The following two chapters outline three experiments that were designed with the major goals of: (1) delineating the types and magnitudes of linguistic adaptations in the clear speech of consonants sounds, focusing on English fricatives, and (2) assessing the range of effects that speakers' strategies actually have on the intelligibility of their productions. Chapter 2 describes Experiment 1, which addresses three questions regarding the fine-grained acoustic-phonetic properties of clearly produced English fricatives: (1) what (if any) systematic changes are made in clear fricative productions, (2) whether clear-speech modifications are dependent on inventory-level and/or more local context provided by “listener” feedback, and (3) in what ways talkers vary in the production of clear fricatives. Using a novel elicitation method, a database of some 8,800 clear and conversational fricative productions by 20 speakers was collected and examined along 53 potentially informative dimensions in order to determine the types of adaptations that might be made in specific situations. Chapter 3 describes Experiments 2-3, which address (1) whether (and which) clearly produced fricatives are more intelligible than conversational fricatives for listeners with normal hearing in degraded conditions, (2) what acoustic modifications observed in Experiment 1 are related to any style-related intelligibility differences, and (3) how clear-speech intelligibility differences and identifiable acoustic correlates differ based on listener population, specifically for listeners with (simulated) sloping, recruiting hearing loss. Chapter 4 summarizes the results obtained from these experiments, and discusses likely immediate future directions for related research.

Chapter 2

Acoustic characteristics of clear fricatives

1. INTRODUCTION

1.1 Introduction

Language users can alter their speech productions in order to speak more or less 'clearly' in response to the communicative needs of different listeners in different situations. Deliberately clarified speech has been seen to yield intelligibility advantages of 3 to 38 percentage points relative to 'normal', conversational speech for hearing-impaired listeners in quiet (Picheny *et al*., 1985; Uchanski *et al.*, 1988) and in noise or reverberation (Payton *et al.*, 1994; Schum, 1996), normal-hearing listeners in noise or reverberation (Ferguson, 2002; Ferguson and Kewley-Port, 2002; Helfer, 1997; Krause and Braida, 2004; Payton *et al.*, 1994) or with simulated hearing loss or cochlear implants (Gagné *et al.*, 1994; Iverson and Bradlow, 2002; Liu *et al.*, 2004), elderly listeners with or without hearing loss (Helfer, 1998; Schum, 1996), cochlear-implant users (Iverson and Bradlow, 2002; Liu *et al.*, 2004), children with or without learning disabilities (Bradlow *et al.*, 2003) and (to a lesser extent) nonnative listeners (Bradlow and Bent, 2002).

Acoustic descriptions of clear speech have generally been dominated by global (sentence-level) patterns; reduced speaking rate, more and longer pauses, increased mean and range of fundamental frequency (f0), a shift in energy to higher frequency regions in long-term spectra, and deeper temporal amplitude modulations have been observed in clear speech (Bradlow *et al.*, 2003; Krause and Braida, 2004; Liu *et al.*, 2004; Picheny *et al.*, 1986; Smiljanić and Bradlow, 2005). At a phonological level, clear speech seems to involve less frequent vowel reduction, burst elimination, and alveolar flapping; and more frequent schwa insertion (Bradlow *et al.*, 2003; Krause and Braida, 2004; Picheny *et al.*, 1986). Previous study on fine-grained acoustic-phonetic characteristics of clear speech has mainly considered vowels, noting increases in vowel durations, expanded F1 × F2 space area, tighter within-category clustering, and more dynamic formant movements (Bradlow *et al.*, 2003; Chen, 1980; Ferguson, 2002; Ferguson and Kewley-Port, 2002; Johnson *et al.*, 1993; Moon and Lindblom, 1994; Picheny *et al.*, 1986; Smiljanić and Bradlow, 2005). Since clear speech is by definition produced in order to increase intelligibility, and since a vast majority of perceptual errors result from consonant confusions (e.g. Miller and Nicely, 1955) it is curious that clearly produced consonants have not been examined as thoroughly. Previous analyses have been limited to a few temporal and amplitudinal parameters including segmental duration, voice-onset-time (VOT), and consonant-to-vowel amplitude ratio (CVR) (Bradlow *et al.*, 2003; Chen, 1980; Krause and Braida, 2004; Picheny *et al.*, 1986). Chen (1980) and Picheny *et al.* (1986) found overall longer plosive, fricative, nasal and semivowel durations; longer VOT for voiceless plosives; and increased CVR for plosives and some fricatives. Larger word-initial CVR was also reported by Bradlow *et al.* (2003). Picheny *et al.* (1986) reported increased peak frequency and overall intensity at higher frequencies in /t/ and /s/ productions, although these changes were not consistently found for consonants produced clearly at faster rates (Krause and Braida, 2004).

This study attempts to identify the types and magnitude of adaptations in the clear productions of a class of consonant sounds, namely English fricatives, in order to characterize the strategies that speakers use when attempting to produce these sounds clearly and to provide a basis for determining what effect these strategies have on the intelligibility of their productions. Fricatives were selected for analysis since previous consonant confusion analyses have reported that fricatives, especially non-sibilants, are a large source of errors for hearing-impaired listeners and for normal-hearing listeners in noise (e.g. Bilger and Wang, 1976; Miller and Nicely, 1955; Wang and Bilger, 1973). To our knowledge, there have been no systematic investigations of acoustic-phonetic alternations related to clearly-produced fricatives. A few studies have considered vocal effort and rate modifications and hyperarticulation in describing fricative acoustics (Shadle and Mair, 1996; Jesus and Shadle, 2002; Perkell *et al.*, 2004) and perception (Feijoo *et al.*, 1998), but clear production was not the primary focus of these studies, which were therefore inconclusive with respect to specific clear-speech alternations.

1.2 Acoustic properties of English fricative sounds

Several studies have attempted to delineate stable acoustic correlates of fricative place of articulation and voicing. Parameters that seem to influence identification include gross spectral shapes and peak locations (Behrens and

Blumstein, 1988; Hughes and Halle, 1956; Jongman *et al.*, 2000a; Strevens, 1960), the first four moments of the spectral energy distribution (Forrest *et al.*, 1988; Jongman *et al.*, 2000a; Nissen and Fox, 2005; Nittrouer, 1995; Nittrouer *et al.*, 1989; Shadle and Mair, 1996), the slopes of lines fit to spectra in lower and higher frequency regions (Evers *et al.*, 1988; Jesus and Shadle, 2002), formant transition information (Jongman *et al.*, 2000a; McGowan and Nittrouer, 1988; Nittrouer et al., 1989; Soli, 1981), overall amplitude (Behrens and Blumstein, 1988; Jongman *et al.*, 2000a; Stevens, 1971; Strevens, 1960), amplitude relative to the neighboring vowel in specific frequency regions (Hedrick and Ohde, 1993; Jongman *et al.*, 2000a; Stevens, 1985), and duration (Baum and Blumstein, 1987; Crystal and House, 1988; Jongman, 1989; Jongman *et al*., 2000a). Briefly, alveolar fricatives are characterized by spectral energy and major peaks at higher frequencies compared to palato-alveolars, which display larger overall relative amplitudes. Dental and labio-dental fricatives show relatively flat spectra below 10 kHz with no dominating peaks, while alveolar and palato-alveolar fricative have well-defined peaks. Nonsibilants show higher standard deviations, lower overall amplitudes, and shorter durations than sibilants. Thus, these parameters clearly distinguish sibilants from nonsibilants and from each other, but are less effective at determining place of articulation for nonsibilants. Fewer studies have reported on the voicing distinction in fricatives (Baum and Blumstein, 1987; Crystal and House, 1988; Jesus and Shadle, 2002; Jongman *et al.* 2000a). These studies suggest that voiceless fricatives are characterized by higher spectral mean and peak values, more defined peaks, less variance, negative skewness, larger overall amplitude and longer duration than their voiced counterparts.

1.3 Contrastive effects of clear speech

Most previous studies characterizing clear speech have been fairly broad in terms of the sounds and contexts across which measurements were made. A secondary question of this study was whether talkers exhibit more specific context-dependent, segmental contrast-enhancing changes. It has been suggested that clear speech modifications are inventory-dependent and effectively increase the auditory distance between neighboring categories. For example, VOT for voiceless stops increases in clear speech but is unchanged for voiced stops (Chen, 1980; Krause and Braida, 2004; Ohala, 1994; Picheny *et al.*, 1986). Similarly, English tense vowels are lengthened to a greater extent than lax vowels to maximize the inherent duration difference between the two vowel categories (Ferguson and Kewley-Port, 2002; Picheny *et al.*, 1986; Uchanski *et al.*, 1996). Talkers also enlarge the distance between vowels in F1 × F2 space, producing more extreme, distinct categories (Bradlow *et al.*, 2003; Chen, 1980; Ferguson, 2002; Ferguson and Kewley-Port, 2002; Johnson *et al.*, 1993; Moon and Lindblom, 1994; Picheny *et al.*, 1986; Smiljanić and Bradlow, 2005). Thus, clear speech may reflect knowledge of the contrasts in an inventory and a general effort to maintain these contrasts. It is less clear whether talkers may attempt to preserve contrast at a more local level, adapting in response to perception errors that are likely to occur in specific contexts. According to Lindblom's H & H Theory (e.g., 1990, 1996), speakers constantly assess listeners' needs for explicit signal information and modulate their speech along a continuum from hypo- to hyper-speech in response to communicative constraints. Along these lines, a speaker's task and goals during clear speech production are quite variable depending on the information needs associated with perhaps each individual segment (depending on cues from the listener, knowledge of the language, etc.). Explicit feedback from the listener in particular might affect clear speech acoustics under these assumptions. For example, when a talker repeats a sequence containing some speech sound after it has been misapprehended for another, similar sound, is the talker likely to make specific adjustments that are not predictable based on general clear speech patterns or inventory-level contrast-enhancing manipulations? A few previous studies have touched on this issue. Ohala (1994) employed an elicitation method in which speakers received pseudo-misrecognitions as feedback to their productions and were asked to repeat target stimuli as clearly as possible. This method was designed to test whether speakers make effort not only to improve the intelligibility of the target stimulus but also to make it sound more unlike the sound it was mistaken for. Contrary to expectations, there were no differences in VOT, vowel duration, or the first three formants of vowels as a function of this feedback. There was no evidence of locally 'contrastive' variation in speech, so it was suggested that clear speech is 'stable' and guided more by general principles reflecting the phoneme inventory of a language, than by microscopic context information like anticipation of specific errors. Some caution is warranted, however, in interpreting Ohala's null result. Most notably, while the study considered some 70,000 measurements, this data set was used to account for a very large number of vowel and consonant contrasts and was therefore underpowered with respect to many of the critical comparisons. The present study replicates and extends Ohala's elicitation method for a much more targeted analysis of nearly 500,000 measurements relating to fricative voicing and place of articulation in a single vowel context.

1.4 Talker differences in clear speech production

Studies have shown that different talkers employ different techniques during clear speech production (Bradlow *et al.*, 2003; Chen, 1980; Ferguson, 2002; Krause and Braida, 2004; Liu, *et al.*, 2004; Picheny *et al*. 1986). For example, one speaker in Picheny et al.'s corpus increased VOTs for both voiceless and voiced plosives in clear speech while the other two increased VOT only for the voiceless one. This speaker also decreased intensity for fricatives in clear speech while the other two speakers showed the opposite pattern. The female talker from the Bradlow et al. (2003) study decreased her speaking rate in clear speech to a far greater degree than the male talker. These two talkers also differed noticeably in the pitch, vowel space, and CVR differences between clear and conversational speech. The female talker from Liu *et al.* (2004)'s database also increased the mean and variability of overall sentence durations more than the male talker. Chen's (1980) three talkers varied in complex ways in the degree to which the syllable, VOT, vowel, and formant transition durations changed. The speakers also differed in terms of within-vowel F1 ×F2 space variability and the magnitude of the increase in f0 mean in clear speech. Changes in f0 were also inconsistent across two talkers in the study by Krause and Braida (2004).

In short, the acoustics of clear speech are highly talker-dependent. However, most of the research that examined talker differences in acoustic modifications recorded small numbers of talkers (*n* = 2 for Bradlow *et al.*, 2003, Krause and Braida, 2004 and Liu *et al.*, 2004; *n* = 3 for Chen, 1980 and Picheny *et al.*, 1986); *cf n*=12 for Ferguson (2002)). With data only from a few speakers, it is unclear whether the patterns of variability observed across speakers and gender would maintain more generally, or if still other strategies would emerge. This study examined the productions of 20 speakers (10 female and 10 male) to address these questions more conclusively.

1.5 Hypotheses

This study was designed to answer three questions concerning the production of clear fricatives. First, what (if any) systematic changes are made in clear fricative productions? Based on previous findings, we hypothesized that clear fricatives would be (1) longer, (2) amplified relative to neighboring vowels, (3) higher in spectral content, including peak locations, spectral mean values, and related measures. Second, are clear-speech modifications dependent on inventory-level and/or more local context provided by "listener" feedback? We expected that, in general, clear productions would be influenced by the perceived likelihood of different misapprehension patterns. More specifically, we predicted that on average (1) fricative categories would differ more from minimally contrastive categories in clear than in conversational speech, and (2) fricatives repeated after misapprehension for similar sounds would be most different from the sounds they were mistaken for. For example, /s/, characterized by a high spectral peak frequency, was expected to increase overall in this value in clear speech, while /ʒ/, characterized by a very low peak, was expected to increase less or decrease. Likewise, /ʃ/ was expected to decrease in peak frequency after misapprehension as /s/, since this would increase the difference between the two sounds along this dimension, while it was expected to *increase* in frequency after misapprehension as /ʒ/, for the same reason. Finally, in what ways do talkers vary in the production of clear fricatives? We hypothesized that cross-talker differences would be seen both in the types of modifications that are made and in the extent of these changes.

2. Experiment 1: Acoustic data collection

2.1 Method

2.1.1 Participants

20 talkers (10 F, 10 M) aged between 19 and 34 were recruited from the University of California, Berkeley, and the University of Kansas, Lawrence, communities. Participants were native speakers of American English, without noticeable regional dialects. Talkers reported normal hearing and no history of speech or language disorders. Talkers had no professional speaking experience and their

experience with linguistics/phonetics varied. They volunteered for the experiment without monetary compensation.

2.1.2 Materials

The eight English fricatives /f, v, θ, ð, s, z, ʃ, ʒ/ and the vowel /ɑ/ were combined to form vowel-consonant-vowel (VCV) syllables. Production of each VCV token was recorded in isolation in conversational and clear speaking styles.

2.1.3 Procedures and apparatus

Participants' speech was recorded digitally at a 44.1 kHz sampling rate (16 bit resolution) in a sound-attenuating booth in the Phonology Lab, UC Berkeley, using a Marantz PMD670 recorder and Shure SM-10 A headset microphone. The microphone was placed 1 inch away from the corner of the talker's mouth at a 45-degree angle. Participants were seated at a comfortable distance from a visual display of prompt, instruction, and feedback on a computer screen. Before recording began, participants were provided with a list explaining the pronunciation of each sound. Items were written: 'afa', 'atha', 'asa', 'asha', 'ava', 'adha', 'aza', and 'azha'. Participants first read these syllables aloud a few times to become familiar with awkwardly-spelled syllables. A pronunciation key was available for reference throughout the session.

The recording session was divided into two parts: warm-up and experiment. Programs to provide prompts and feedback were designed using MATLAB 7.0.0.1 (The Mathworks, Inc., 2000). During warm-up, talkers produced five repetitions of each VCV in response to prompts appearing on the screen. At first, talkers read VCV syllables in a manner approximating the way they spoke in everyday conversation; later, they were instructed to speak more carefully, as if talking to a hearing-impaired or elderly person. This warm-up served to familiarize talkers with the interface and materials, allow them to rehearse the two styles, and provide 'baseline' recording of speech produced before talkers became aware of the rate and types of misperceptions that would be encountered during the experiment.

The elicitation method for the experimental session resembled the one used by Ohala (1994). Before the session, a subject was told that he/she would produce speech as part of an interaction with a computer program that would be recorded. They were instructed to speak first as naturally as possible, as if in casual conversation, when prompted by a VCV stimulus on the screen. This original input served as 'conversational speech' in acoustic analyses. Participants were told that the program would 'guess' which syllables were spoken and indicate its guess on the screen, and that it would frequently misperceive sounds, simulating a hearing-impaired listener. If a participant indicated that a guess was correct (by clicking a box on the screen), the trial terminated and the program moved on to the next stimulus. If a guess was marked incorrect, the speaker was given a chance to repeat the target stimulus, doing his or her best to deliver the stimulus as intelligibly as possible. "Hyperarticulate speech" was observed in the machine-human interaction in which speakers made global acoustic changes, after receiving the recognition error feedback from the computer program (Oviatt *et al.*, 1998a and b). These repeated productions served as ‘clear speech’ in acoustic analyses. ‘Guesses’ were unrelated to the speaker’s production pattern, and represented either: (1) the correct response, (2) voicing-matched but place-unmatched incorrect responses (e.g. /θ/, /s/, or /ʃ/ for /f/), (3) voicing-unmatched, place-matched incorrect responses (e.g. /v/ for /f/), and (4) ‘???’ (“don’t know”) responses. Each response occurred 5 times for each VCV during the experiment. Thus, there were 30 conversational and 25 (three place errors, one voicing error, one ??? × 5) clear productions of each fricative by each talker. The order of prompts was randomized separately for each talker, as was the pattern of pseudo-responses. After the second production, a second guess was displayed, which was correct 75% of the time and random otherwise; the participant scored this guess before continuing to the next trial. Recording sessions lasted 60-70 minutes, including the warm-up and a 10 minute break halfway through the main experiment.

2.1.4 Data processing and acoustic measurements

Recordings were hand-annotated into VCV segments using the PRAAT speech analysis software (Boersma and Weenink, 2000) and further segmented and analyzed using PRAAT and Matlab. Semi-automatic fricative segmentation was achieved following previous studies (Behrens and Blumstein, 1988; Jongman *et al.*, 2000a; Yeni-Komshian and Soli, 1981), in which the fricative was defined as a region of elevated zero-crossings due to the turbulent source, in the following manner. Each production was high-pass filtered at 300 Hz using a 2nd order Butterworth filter, to remove voicing and other low-frequency perturbations that might obscure zero crossings resulting from the turbulent source. The production was then converted into a time series in which each sample was labeled as either differing in sign from the previous sample [1] or not [0], and a zero-crossing envelope was created by low-pass filtering this series at 30 Hz. We found that good identification was achieved by taking the continuous region closest to the center of the production for which the log of this envelope was above half of its maximum value as corresponding to the fricative. Upon hand checking the segmentation based on visual inspection of the spectrogram and waveform, it was found that 91% of fricatives were accurately labeled; the remaining 9% were labeled by hand.

Acoustic analysis considered a 14 parameters that may work in combination to distinguish fricative voicing and place of articulation: spectral peak location (1), the first four spectral moments (2-5), F2 onset transitions (6), spectral slopes below (7) and above (8) peak locations, f0 of adjacent vowels (9), overall RMS amplitude (10), relative amplitude i.e., a change in amplitude of the frication relative to the vowel in F3 region for sibilants and F5 region for non-sibilants (11), harmonic-to-noise ratio (HNR; 12), energy below 500 Hz (13) and fricative duration (14). Except where noted, all analyses considered 40-ms Hamming windowed segments at five locations, centered over the fricative onset, 25, 50, and 75% points, and offset. Spectral peaks were defined as the frequency with the highest absolute magnitude in the FFT of a windowed segment. Moments 1-4 were also calculated from the absolute FFT spectrum. F2 values were calculated using the Burg algorithm as implemented in PRAAT, derived at fricative onset and offset and each vowel midpoint from an analysis that found at most five formants below 5000Hz (male speakers) or 5500Hz (females). Spectral slopes were computed following the procedures described by Evers *et al.* (1998) and Jesus and Shadle (2002) Lines were fit to FFT power spectra across two regions defined by the average peak frequency (across talkers and productions) for a fricative. Low-frequency slope (dB/kHz) was derived from the spectral values below this peak, and high-frequency slope from the peak to 22.05 kHz.

Fundamental frequency was derived using an autocorrelation-based algorithm (Boersma, 1993). F0 was averaged across the vowels preceding and following the target. Normalized amplitude was taken as the difference (dB) in RMS amplitude between the same five windowed fricative segments described above, and the averaged of the surrounding vowels. (The use of both vowels for F0 and amplitude analysis was necessary because some speakers tended to place emphasis on the first vowel and some on the second.). Relative amplitude was measured as described in Hedrick and Ohde (1993) and Jongman et al. (2000a). FFTs were taken of one 23.3-ms Hamming window centered on the fricative midpoint, and one with the left window skirt positioned at the onset of following vowel periodicity. For sibilants the peak in the region corresponding to F3 of the frication noise were compared to the peak of the vowel onset in the same frequency region; for non-sibilants the peak at F5 was used. Relative amplitude was then expressed as the difference (dB) between fricative and vowel amplitude. Mean HNR across the fricative was obtained using a cross-correlation algorithm. Intensity below 500 Hz was obtained similarly to normalized amplitude, except that the VCV was first low-pass filtered at 500 Hz.

2.1.5 Statistical analysis

For each metric at each window, hypotheses (1) and (2; inventory-based contrastive effects) were tested using a four-way (2 × 4 × 2 × 2), mixed-model analysis of variance (ANOVA) with speaking Style (clear vs. conversational), Place of articulation, and Voicing as within-subjects factors; and Gender as a between-subjects factor. Post-hoc pairwise comparisons for significant within-subject factors were done using Bonferroni corrected 95% confidence intervals. A stepwise linear discriminant analysis was conducted to test whether clear fricatives are more discriminable (in terms of place and voicing) than conversational fricatives. For peak location, moments, spectral slopes, normalized amplitude and energy below 500 Hz, 5 window locations served as predictors, and for F2, 4 window locations served as predictors, resulting in 53 total predictors. Classification scores developed by a "jack-knife" cross-validation procedure were compared between speaking styles. Hypothesis (2) was addressed further using a two-way repeated measures ANOVA for clear speech measurements with Misperception and Fricative as factors, to determine whether feedback affected production. A one-way repeated measures ANOVA was also conducted to examine whether differences in acoustic values between minimal pairs were larger in *contrastive* contexts (where a sound was produced after being misperceived as the other) than in non-contrastive. Again, a stepwise linear discriminant analysis was performed for all 16 +/- [voicing], [place] pairs and classification scores were compared to those obtained for conversational and non-contrastive clear speech. Hypothesis (3) was addressed using a two-way mixed-model ANOVA with Style as within-subject factor and Talker as between-subject factor.

2.2 Results

2.2.1 General clear speech alternations and talker differences

Complete results of the Style × Place × Voicing × Gender ANOVA for each measure are summarized in Table 2-1; specific findings are summarized below.

Table 2-1: Summary of repeated measures analyses of variance (ANOVA) for each acoustic measurement, with within-subject factors of Style (S: 2 levels), Place (P: 4 levels), Voicing (V: 2 levels), Fricative (F: 8 levels) and Misperception (M: 2 levels), and between-subject factors of Gender (G: 2levels) and Talker (T: 20 levels), (S=Style, P=Place, V=Voicing, F=Fricative, M=Misperception, G=Gender, T=Talker, ***p<.001, **p<.01, *p<.05, ·p<.1, -p>.1).

	S	P	V	G	S×P	S×V	S×G	P×V	S×P×V	P×G	V×G	P×V×G	S×P×G	S×V×G	S×P×V×G	T	S×T	M	M×F
Duration	***	***	***	-	-	***	-	*	-	-	-	-	-	-	-	***	***	.	-
FFTW1	***	***	***	*	**	-	-	***	***	***	-	**	-	-	-	-	***	-	.
FFTW2	***	***	***	-	***	.	.	***	***	**	-	-	-	-	-	-	-	***	*
FFTW3	***	***	***	-	***	**	-	***	***	**	-	-	-	-	-	-	-	***	.
FFTW4	***	***	***	-	***	***	-	***	***	**	-	-	-	-	-	-	-	*	.
FFTW5	-	***	*	-	-	***	-	.	.	.	-	-	-	-	-	-	-	-	-
M1W1	***	***	-	***	***	***	-	***	*	.	***	-	-	-	-	***	**	-	-
M1W2	***	***	***	***	***	**	-	***	***	***	.	-	.	-	-	-	**	***	*
M1W3	***	***	***	***	***	-	-	***	***	**	-	-	-	-	-	-	**	-	.
M1W4	***	***	***	***	***	-	-	***	***	**	-	-	-	-	-	-	***	***	.
M1W5	-	***	.	**	*	***	-	***	***	*	**	-	-	-	-	***	**	-	-
M2W1	**	***	-	***	***	***	-	***	***	-	***	-	-	-	-	***	***	*	-
M2W2	-	***	-	-	***	***	-	***	**	-	**	-	-	-	-	-	-	***	.
M2W3	-	***	-	-	**	***	-	***	*	-	*	-	-	-	-	-	.	***	.
M2W4	*	***	.	-	***	***	-	***	***	-	*	-	-	-	-	-	-	***	-
M2W5	.	***	-	.	***	***	-	***	***	-	*	-	-	-	-	***	**	*	-
M3W1	***	***	***	***	***	***	-	***	***	-	***	-	-	-	-	***	***	-	-
M3W2	***	***	***	**	***	***	-	***	***	*	*	-	-	-	-	-	-	*	.
M3W3	***	***	***	*	***	***	-	***	***	*	-	-	-	-	-	-	-	**	*
M3W4	***	***	***	*	***	***	-	***	***	*	-	-	-	-	-	-	-	**	*
M3W5	***	***	**	-	.	***	-	***	***	*	**	-	-	-	-	***	-	-	-
M4W1	***	***	***	**	***	***	-	***	***	-	**	-	-	-	-	***	***	-	-
M4W2	***	***	***	.	***	***	-	***	***	-	*	-	-	-	-	-	-	*	.
M4W3	***	***	***	-	***	***	-	***	**	-	.	-	-	-	-	-	-	**	*
M4W4	***	***	***	.	***	***	-	***	***	-	-	-	-	-	-	-	-	**	*
M4W5	**	***	*	-	**	***	-	***	***	.	*	-	-	-	-	***	-	-	-
F2W1	-	***	-	***	-	*	-	***	**	*	-	-	-	-	-	***	***	-	-
F2W2	***	***	***	***	-	-	-	***	-	.	-	-	-	-	-	***	***	-	-
F2W3	*	***	**	***	*	***	-	***	-	*	.	-	-	-	-	**	*	-	-
F2W4	***	***	-	***	***	-	-	-	-	-	-	-	-	-	-	***	***	-	-

2.2.1.1 Spectral peak locations.

Figure 2-1 shows mean peak location as a function of fricative, style and window location. Style effects were obtained at 4 locations, with peaks averaging 824 Hz higher in clear speech. Main effects of Place were observed at all 5 locations; Pairwise comparisons at the 3 central locations showed that alveolar fricatives had the highest peaks and palato-alveolars the lowest. Peak location differed significantly between all place pairs except for the dental - labiodental contrast. The Voicing effect was found at all 5 locations, with higher values for voiceless fricatives. A Style × Place interaction was found at 4 locations; Bonferroni *post hoc* tests revealed that peak locations increased significantly in clear speech for all fricatives except palato-alveolars. A Style × Voicing interaction was observed at 3 locations, with a bigger increase in clear speech for voiced fricatives ($p<.001$). A Gender effect was found at window W1 [F(1, 18)=6.386, $p<.05$]; female speakers had higher peak frequencies (2859 Hz) than male speakers (2290 Hz). No interactions involving Gender and Style were observed (F<1). There was no significant Talker effect; the Style × Talker interaction was seen only at W1 [F(19, 140)=2.672, $p=.001$].

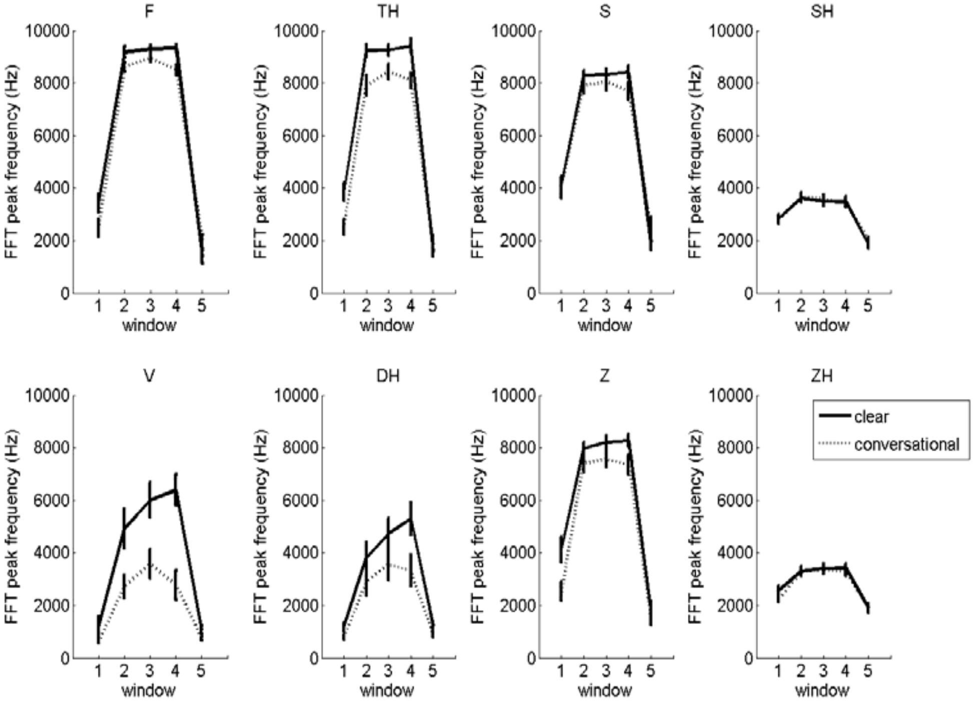


Figure 2-1: Mean FFT peak frequency (Hz) for each fricative at 5 window locations as a function of style

2.2.1.2 Spectral moments

Figures 2-2 to 2-5 summarize data for moments 1-4. For spectral mean (M1), the Style effect was seen at most locations, with clear speech means on average 1.16 times higher than conversational. The Place effect was significant at all locations. Pairwise comparisons showed higher values for alveolars and the lowest values for palato-alveolars, with no significant differences between the two non-sibilants. A Voicing effect was found at the 3 central locations; voiceless fricatives had higher M1 values. Style × Place interactions at all 5 locations, and *post hoc* tests, showed that labio-dentals, dentals, and alveolars increased in M1 in clear speech, while palato-alveolar fricatives did not. Style × Voicing interactions were found at 3 locations; M1 increased more for voiced fricatives at W1 and more for voiceless fricatives at W2, and decreased more for voiceless clear fricatives than voiced at W5. Main effects of Gender at all locations revealed that female speakers had higher mean frequencies than males. No Style × Gender interactions were seen, indicating that female speakers and male speakers did not differ in the extent to which they modified mean frequency values in clear speech. However, main effects of Talker were seen at 2 locations and Style × Talker interactions at all locations.

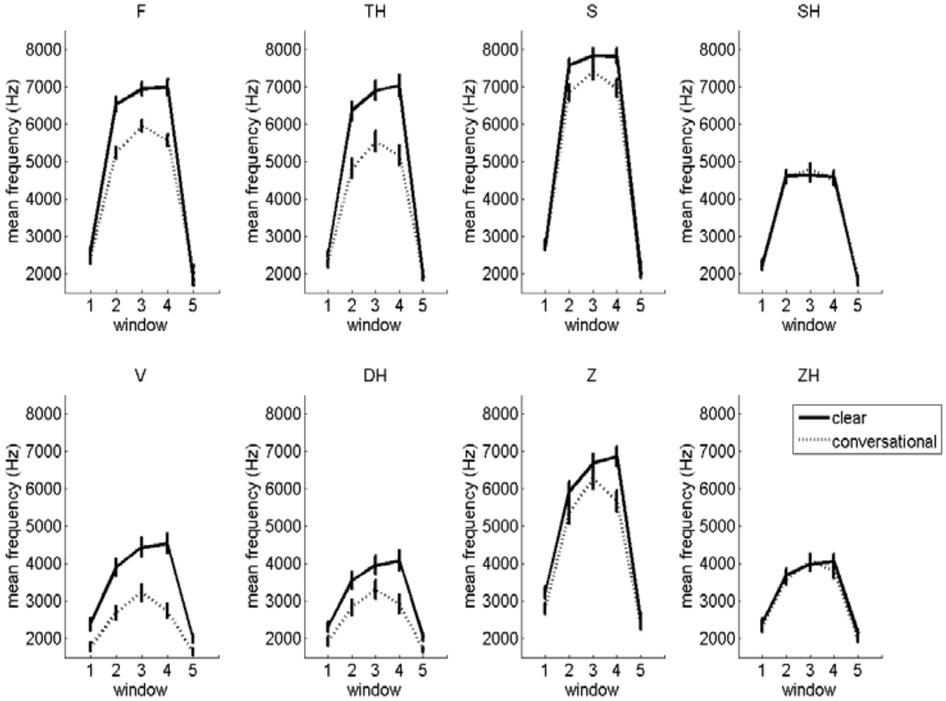


Figure 2-2: Mean moment 1 values (mean frequency; Hz) for each fricative at 5 window locations as a function of style

Main effects of Style for M2 (standard deviation) were found at two window locations, with (on average 1.4 times) higher values in clear speech. Main effects of Place were found at all locations; pairwise comparisons showed that non-sibilants had higher M2 than sibilants, and that palato-alveolars had the lowest overall values. M2 did not differ based on Voicing. Style × Place interactions at all 5 windows and *post hoc* tests showed that non-sibilants increased, and sibilants decreased slightly, in clear speech. Style × Voicing interactions were also seen at all 5 locations; M2 increased for voiced fricatives and decreased slightly for their voiceless counterparts. Style × Place × Voicing interactions were found at all 5 locations, deriving from these same two patterns. Female speakers showed higher standard deviation values than males at W1 [F(1, 18) = 14.446, *p* =.001] and tended higher at all locations. No Style × Gender interactions were observed. Talker effects and the Talker × Style interaction were seen at W1 and W5.

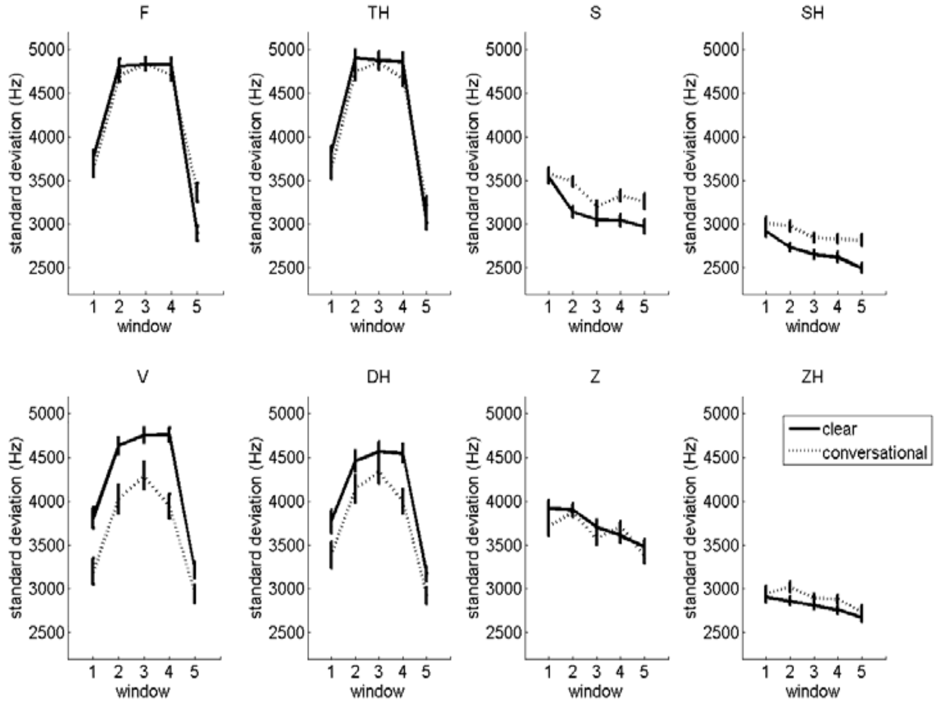


Figure 2-3: Mean moment 2 values (standard deviation; Hz) for each fricative at 5 window locations as a function of style

A main effect of Style for M3 (skewness) was seen at all 5 locations, with lower values in clear speech (mean 0.36 × conversational). This indicates that energy concentration was universally shifted to higher frequency regions in clear speech. The Place effect was also significant at all locations. Pairwise comparisons showed lowest values for alveolar fricatives, highest for palato-alveolars, and no differences between non-sibilants. There were effects of Voicing at all locations with lower skewness for voiceless fricatives. Style × Place interactions at all locations and *post hoc* comparisons revealed significant decreases for all places except palato-alveolars. Palato-alveolars either increased (n.s.) in skewness (at the central window locations) or did not decrease (in peripheral locations) as much as other fricatives. Style × Voicing interactions indicated larger decreases for voiced fricatives. Female speakers showed lower skewness values than males at 4 locations, indicating more energy concentration in higher frequency regions for females; no interactions involving Gender and Style were significant. The Talker effect was significant at the two peripheral locations, and the Talker × Style interaction was significant only at fricative onset.

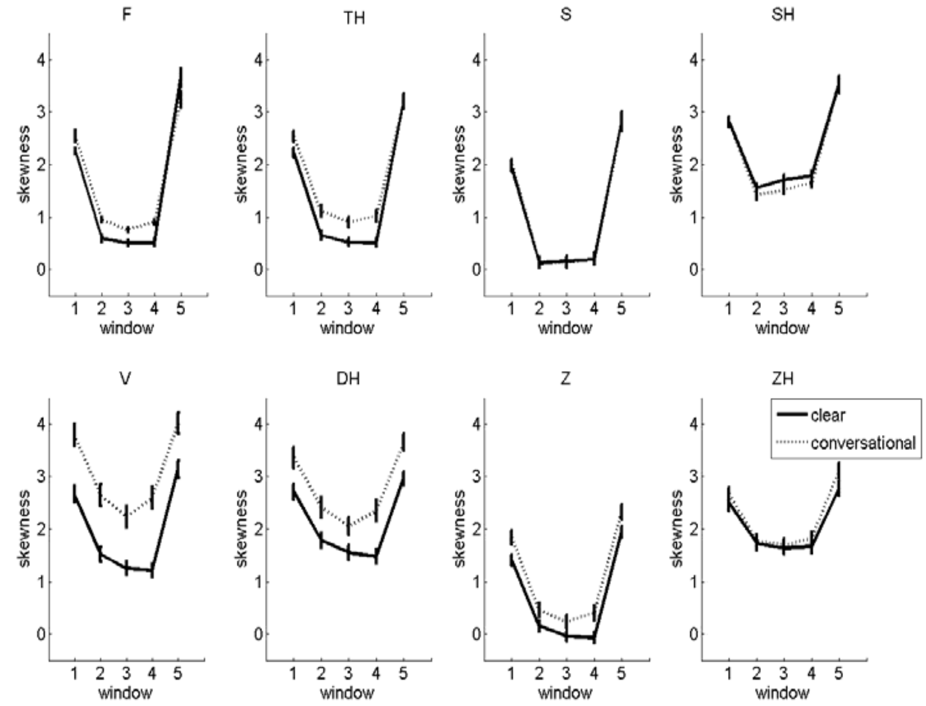


Figure 2-4: Mean moment 3 values (skewness) for each fricative at 5 window locations as a function of style

M4 (kurtosis) decreased in clear speech at all 5 locations (mean 0.7 × conversational value). Place effects at all locations were mainly derived from lower values for alveolar fricatives and highest values for palato-alveolars. Non-sibilants did not differ from each other except at W5. Voiced fricatives had significantly higher kurtosis than voiceless. Style × Place interactions at 5 locations and *post hoc* tests indicated that the decrease in M4 in clear speech was mostly due to non-sibililants; Style × Place × Voicing interactions indicated that voiceless alveolars and palato-alveolars *increased* in clear speech at central locations. Style × Voicing interactions at all locations showed that peakedness decreased for voiced fricatives while voiceless fricatives either (n.s.) increased or did not decrease in kurtosis as much as their voiced counterparts. Except for a main effect at W1 (higher for male speakers), there were no effects involving Gender. Talker effects were found at the transitional locations, and the Style × Talker interaction was found at W1.

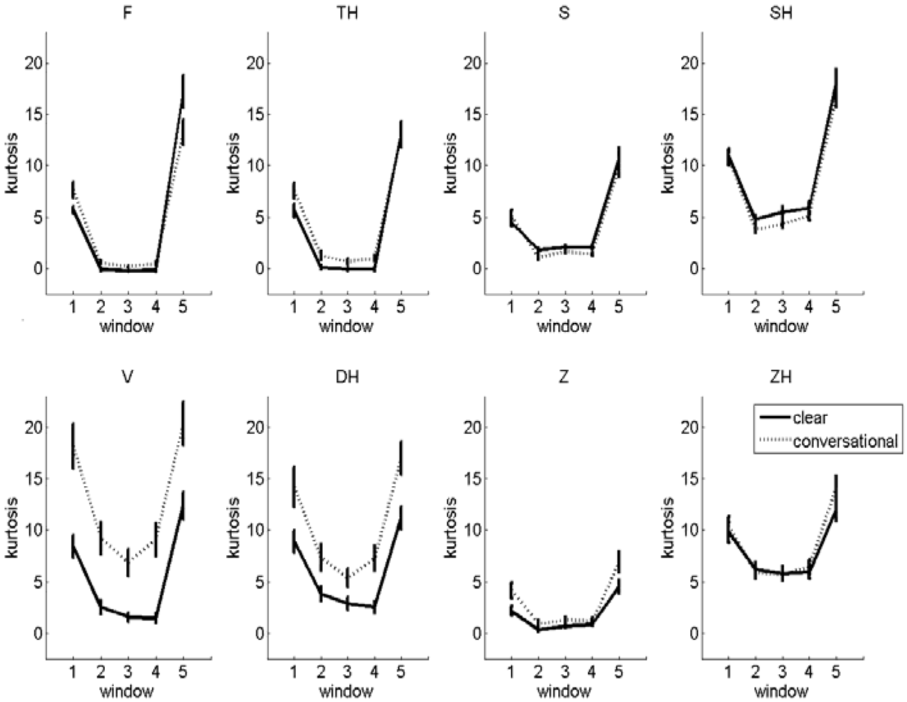


Figure 2-5: Mean moment 4 values (kurtosis) for each fricative at 5 window locations as a function of style

2.2.1.3 F2

Figure 2-6 summarizes F2 data. A Main effect of Style was found at 3 locations, with slightly higher values in clear speech at fricative onset and offset positions (1. 04 × conversational value) and lower values at the midpoint of the following vowel (0.97 × conversational). A Place effect at all 4 locations and pairwise comparisons revealed that (1) dentals showed higher F2 values than labio-dentals at both vowel midpoints, and (2) palato-alveolars had the highest overall values. A main effect of Voicing at the transition locations showed that voiceless fricatives had (1.07 times) higher F2 values. A Style × Place interaction at 2 locations resulted from increases in F2 for dental and palato-alveolar fricatives at W3, and decreases for dental, alveolar, and palato-alveolar fricatives at W4 in clear speech. A Style × Voicing interaction at 2 locations showed a decrease for voiceless fricatives in clear speech at W1, and a significant increase for voiced fricatives at W3. A main effect of Gender showed higher F2 values for female speakers at 4 locations; no Style × Gender interaction was seen. The Talker effect and Talker × Style interaction were seen at all locations.

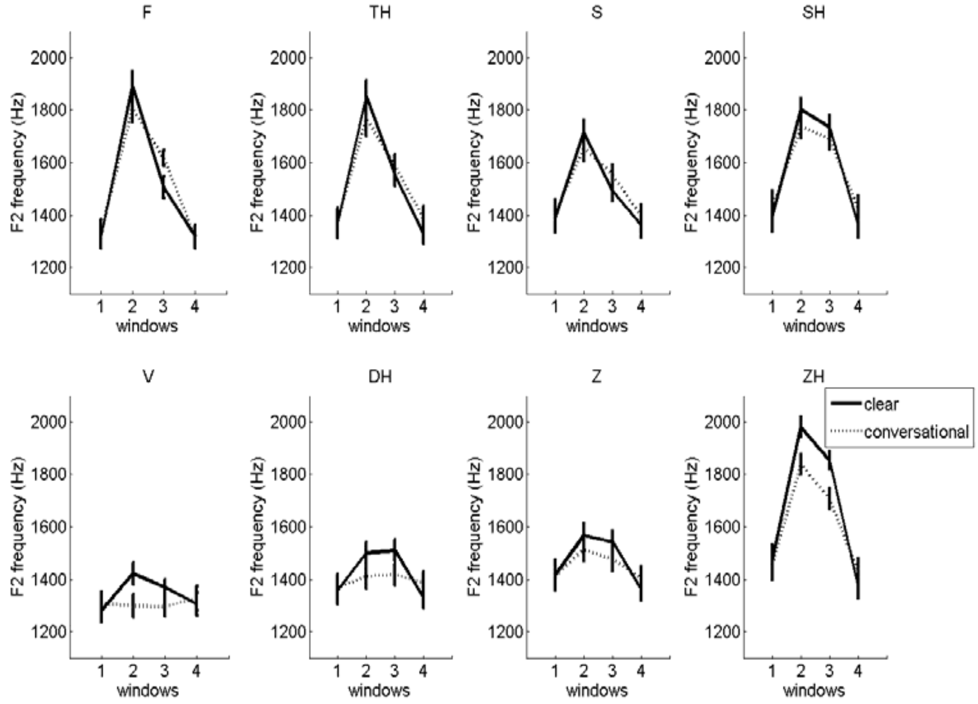


Figure 2-6: Mean F2 values (Hz) for each fricative at 4 window locations as a function of style

2.2.1.4 Spectral slopes

Figure 2-7 and 2-8 show spectral slopes before and after peak locations. For slope before the peak, the Syle effect was significant at all locations, with clear speech 0.751 dB/kHz steeper than conversational. The Place effect obtained at all 5 locations and pairwise comparisons showed that (1) sibilants had higher before-peak slopes than non-sibilants, (2) within sibilants, slopes were highest for palato-alveolars, and (3) non-sibilants did not significantly differ from each other except at W1 (higher for dentals). An effect of Voicing at 4 locations showed higher values for voiceless fricatives, in accordance with previous findings (Jesus and Shadle, 2002). Style × Place interactions were seen at 4 locations. Style × Voicing interactions were significant at 4 locations, showing that although both voiceless and voiced fricatives increased in clear speech, the increase was much larger for voiced fricatives. The Gender effect was found only at W1 with female speakers having higher values than male speakers. There were no Style × Gender interactions at any locations. A main effect of Talker was found at W1. Style × Talker interactions were significant at all 5 locations, indicating that talkers varied in the extent (and sometimes the direction) to which they modulated slope in clear speech.

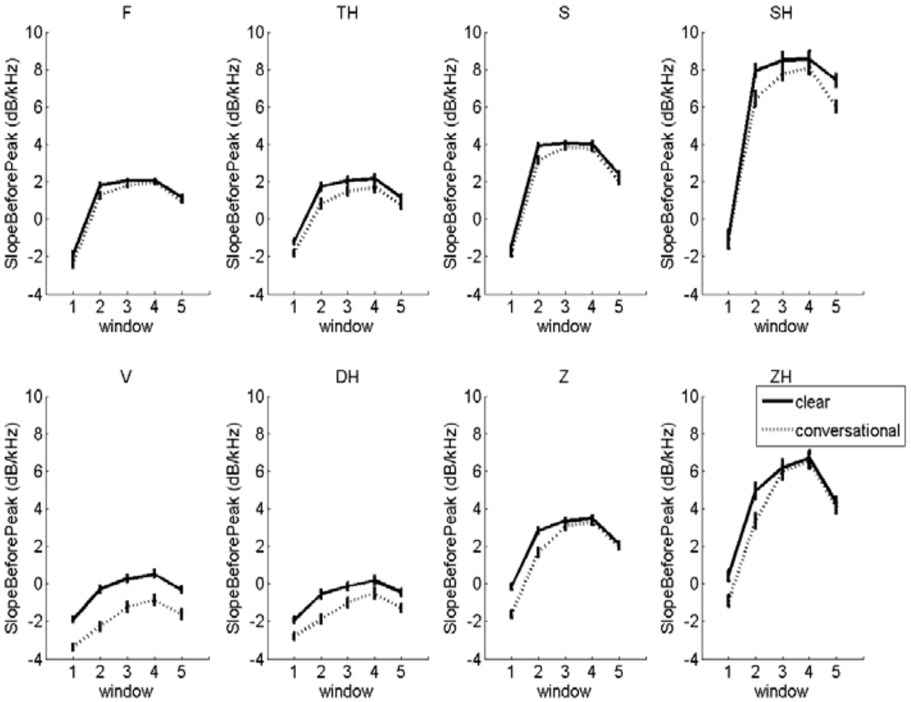


Figure 2-7: Mean slope values before the peak locations (dB/kHz) for each fricative at 5 window locations as a function of style

Slope *after* peak locations showed similar results. A main effect of style was seen at all 5 locations, with clear speech slopes on average 0.302 dB/kHz lower (steeper) than conversational. A main effect of Place at all locations and pairwise comparisons demonstrated greatest negative slopes for alveolars, followed by the palato-alveolars. Non-sibilants did not differ from each other. The Voicing effect was significant at the 3 central locations, with larger negative values for voiceless fricatives. Style × Place interactions were found at 4 locations; *post hoc* tests revealed that although all places of articulation decreased in clear speech, non-sibilants decreased to a much larger degree than sibilants. Style × Voicing interactions at 4 locations indicated voiced fricatives decreased more than voiceless. No effects or interactions involving Gender were observed. Talker main effects were found at the peripheral locations, and Style × Talker interactions were seen at three locations.

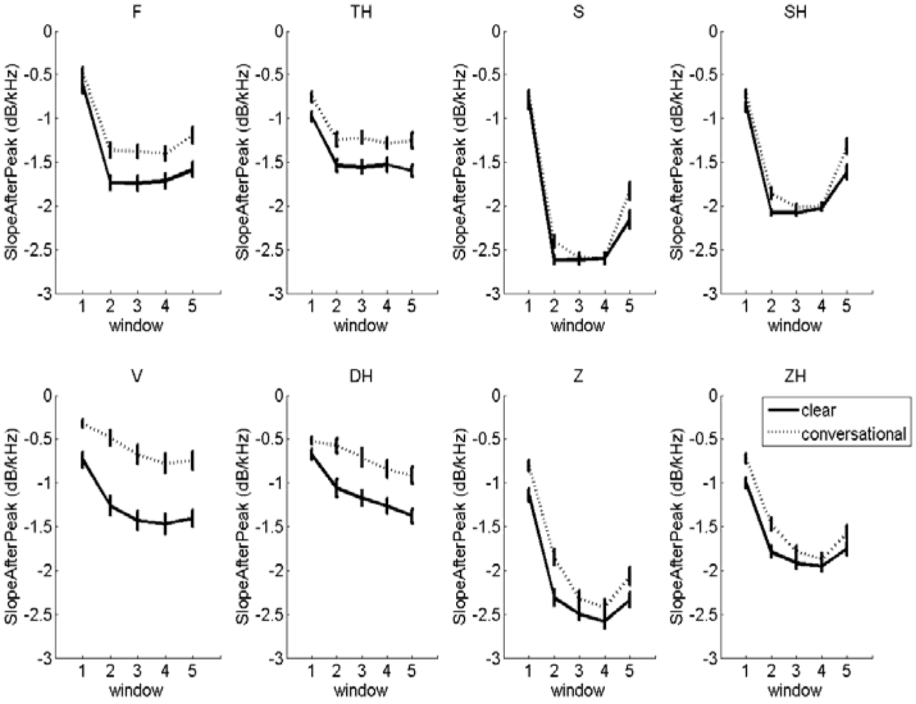


Figure 2-8: Mean spectral slope values after the peak locations (dB/kHz) for each fricative at 5 window locations as a function of style

2.2.1.5 Normalized RMS amplitude

Figure 2-9 shows normalized RMS amplitude for each fricative as a function of style. A main effect of Style was found at all 5 locations, with clear fricatives significantly lower in amplitude at the first four windows (mean 1.04 dB difference) and higher (0.565 dB) at fricative offset. The Place effect was seen at all locations, due to higher amplitude for sibilants than non-sibilants. At all locations, amplitude increased in the order of labio-dentals, dentals, alveolars, and palato-alveolars. Alveolar fricatives and palato-alveolar fricatives significantly differed from each other at 4 locations but labio-dentals and dentals significantly differed only at W5. A main effect of Voicing was found at W4, with lower amplitude for voiceless fricatives. Style × Place interactions were seen at all locations. A decrease in normalized amplitude was obtained only for non-sibilants at central locations. There were significant Style × Voicing interactions at 3 locations, due to a larger decrease for voiced fricatives at W1 and 4, and a significant increase for voiceless fricatives at W5. There was no Gender main effect or Style × Gender interaction. Talker effects and Style × Talker interactions were seen at all 5 locations. P*ost hoc* tests indicated that talkers significantly differed in both the extent and direction of changes; some talkers actually showed increased relative amplitude in clear speech.

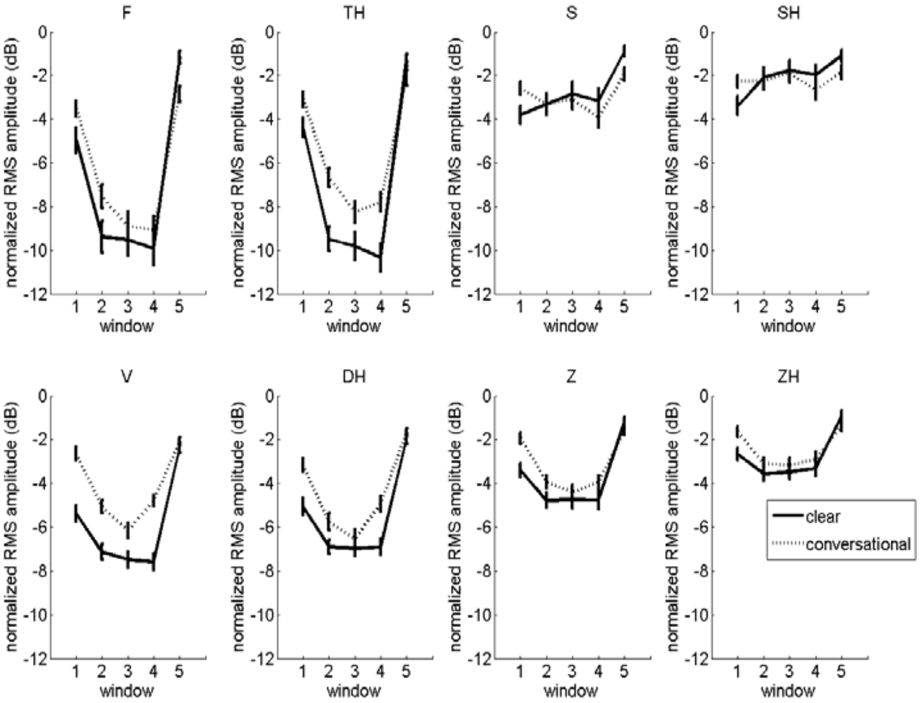


Figure 2-9: Mean normalized rms amplitude (dB) for each fricative at 5 window locations as a function of style

2.2.1.6 Relative amplitude

Figure 2-10 shows relative amplitude for each fricative and style. There was no overall effect of Style (F<1). A main effect of Place [F(3, 27.067)=102.650, *p*<.001] and *post-hoc* comparisons indicated that palato-alveolars had by far the highest relative amplitudes, followed by labio-dentals. There was a main effect of voicing, [F(1, 18)=69.543, *p*<.001], with greater values for voiceless fricatives. A Style × Place interaction [F(3, 54)=15.352, p<.001] resulted from a significant decrease for alveolar fricatives and a significant increase for palato-alveolar fricatives. Interactions involving Voicing were not significant, nor were the Gender effect or

Style × Gender interaction. A main effect of Talker [F(19, 140)=2.525, *p*=.001] and a Style × Talker interaction [F(19, 140)=3.582, *p*<.001] indicated talker variance in the extent and direction of production and modifications in relative amplitude.

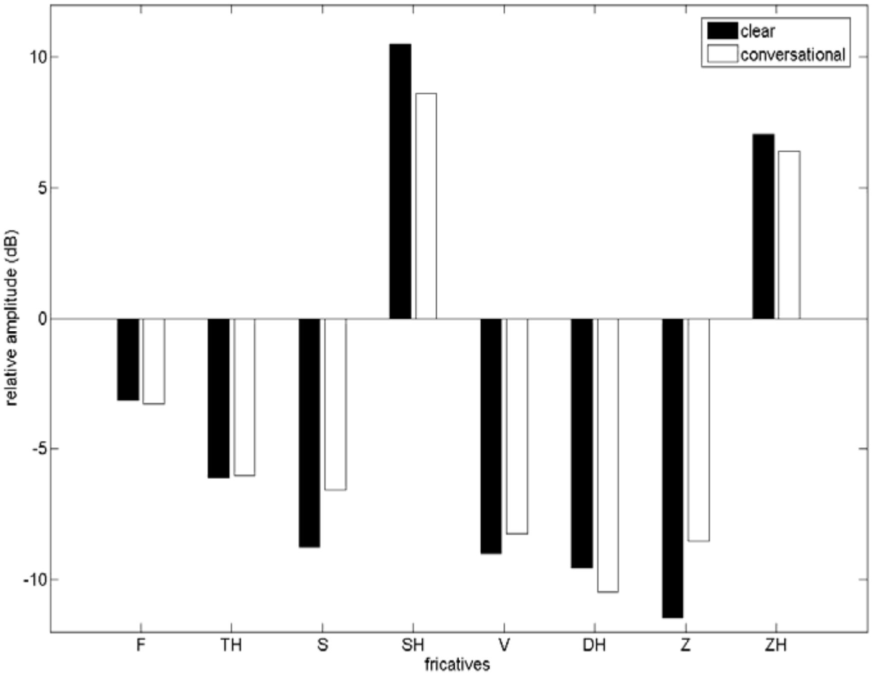


Figure 2-10: Mean normalized relative amplitude values (dB) as a function of fricative and style

2.2.1.7 Energy below 500 Hz

Figure 2-11 shows intensity below 500 Hz data. Clear speech significantly decreased in energy below 500 Hz (mean 1.49 dB difference) at the first 4 locations but significantly increased at fricative offset (1.019 dB). Main effects of Place were found only at peripheral locations, with higher values for sibilants; the Voicing effect was found at the 3 central locations, with greater values for voiced fricatives. Style × Place interactions were seen at all locations, derived from a smaller decrease for palato-alveolars than for other sounds. Style × Voicing interactions at all locations showed that voiced fricatives decreased less than voiceless fricatives, even increasing (n.s.) at the 3 central locations. There were no main effects of Gender or Style × Gender interactions (F<1). The main effect of Talker and the Style × Talker interaction were significant at all locations.

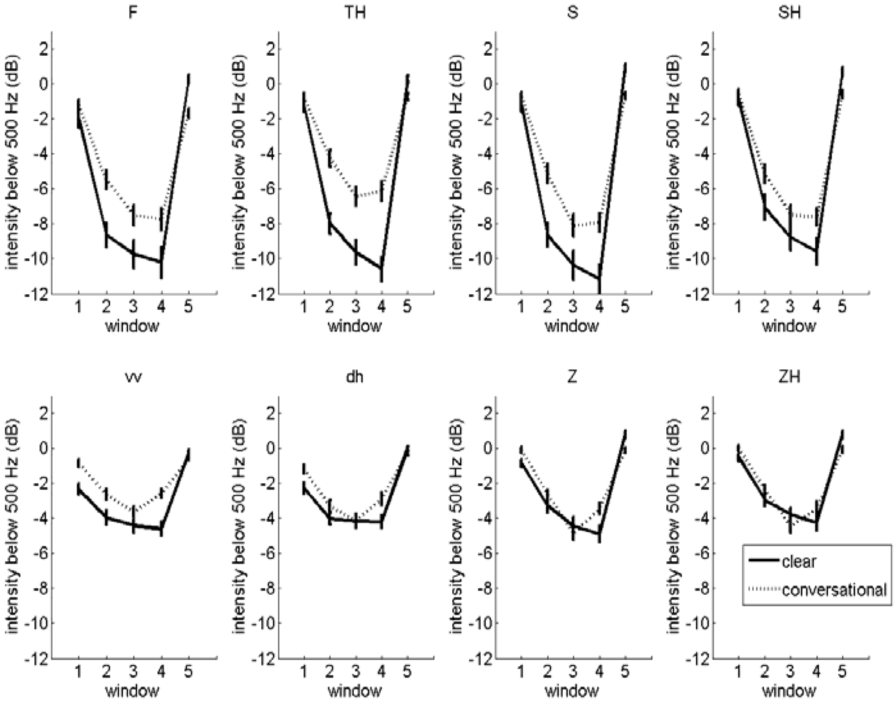


Figure 2-11: Mean normalized intensity below 500 Hz (dB) for each fricative at 5 window locations as a function of style

2.2.1.8 Noise duration

Figure 2-12 shows duration as a function of fricative and speaking style. A significant main effect of Style was seen [F(1, 18)=57.397, *p*<.001], with clear fricatives on average 2.9 times longer than conversational. Main effects of Place and Voicing were also seen [F(3, 46.622)=17.255, *p*<.001; F(1, 18)=190.859, *p*<.001]; sibilants and voiceless fricatives were longer than non-sibilants and voiced fricatives, consistent with previous studies (e.g. Jongman *et al.*, 2000a). The Style × Place interaction was not significant. A Style × Voicing [F(1, 18)=39.563, *p* < .001] interaction revealed that voiceless fricative durations increased more in clear speech. No effects or interactions involving Gender were observed (F<1; *cf* Bradlow *et al.*, (2003), Liu *et al.*, (2004)). The Talker effect and the Style× Talker interaction were both significant, [F(19, 140)=15.206, *p*<.001] and [F(19, 140)=48.590, *p*<.001], indicating that talkers differed in the extent of their modulation of fricative duration in clear speech.

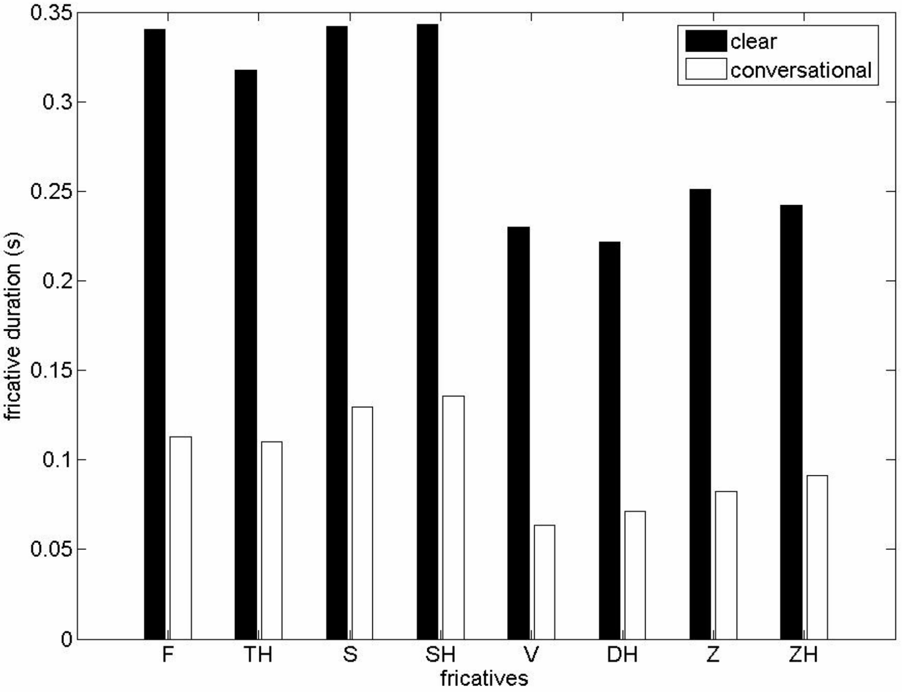


Figure 2-12: Mean noise duration (sec) as a function of fricative and style

2.2.1.9 HNR

Figure 2-13 shows HNR as a function of fricative and style. There was no overall effect of Style on HNR. A main effect of Place [F(3, 54)=57.692, *p*<.001] and pairwise comparisons indicated lower HNR for sibilants than non-sibilants, and palato-alveolars than alveolars. A Main effect of Voicing [F(1, 18)=185.729, *p*<.001] showed higher HNR for voiced fricatives. The Style × Place interaction was not significant, but Style × Voicing and Style × Place × Voicing interactions ([F(1, 18)=24.754, *p*<.001] and [F(3, 2.866)=15.196, *p*<.001], respectively) revealed that HNR significantly increased in clear speech for voiced fricatives and significantly decreased for the voiceless fricatives. AGender effect was found [F(1, 18)=6.424, *p*<.05] but not the Style × Gender interaction. Neither the Talker main effect nor the Style × Talker interaction was significant.

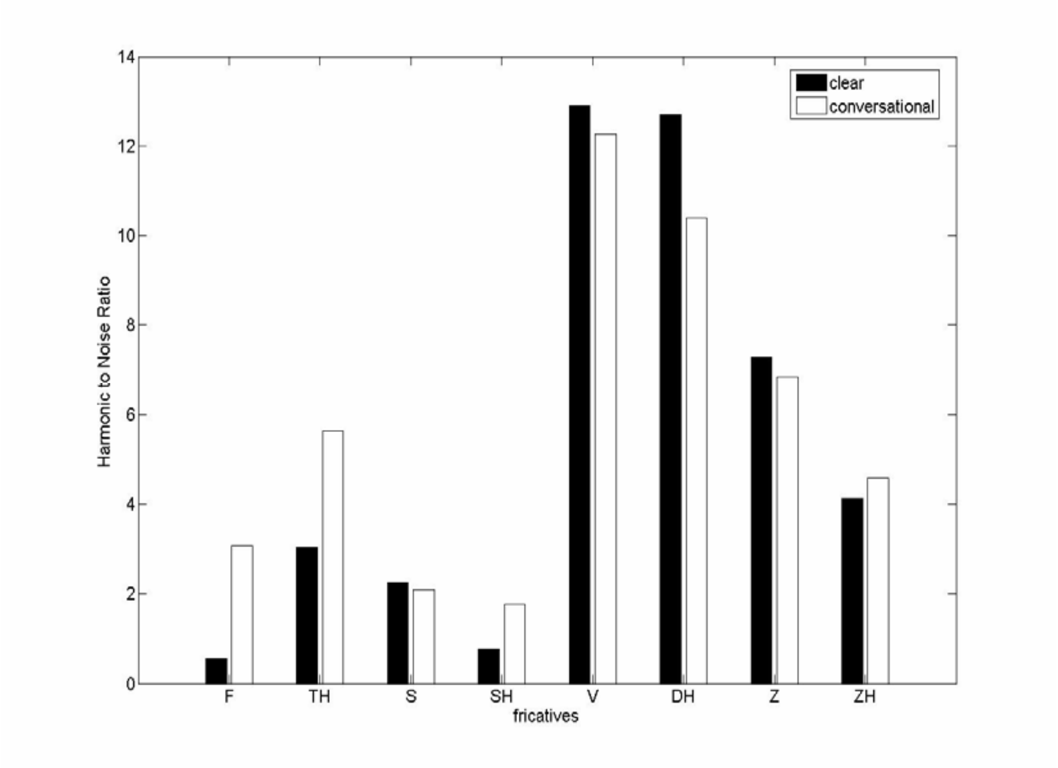


Figure 2-13: Mean HNR averaged across speakers as a function of fricative and style

2.2.1.10 F0

Figure 2-14 summarizes f0 results. No effect of Style or Place was found, with slightly higher fundamental frequency in clear speech (153 Hz for clear speech and 148 Hz for conversational speech); there was a marginal main effect of Voicing [F(1, 18)=4.248, *p*=.054], with higher f0 near voiceless fricatives (3 Hz higher). No Style × Place interaction was found, but there was a Style × Voicing interaction [F(1, 18)=4.454, p<.05], indicating a significant increase in fundamental frequency only for clear voiceless fricatives. The main effect of Gender [F(1, 18)=62.051, *p*<.001] resulted from significantly higher f0 values for female; however, there were no interactions involving Style and Gender. The Talker effect and the Style × Talker interaction were significant ([F(19, 140)=488.007, *p*<.001] and [F(19, 140)=15.482, *p*<.001], respectively).

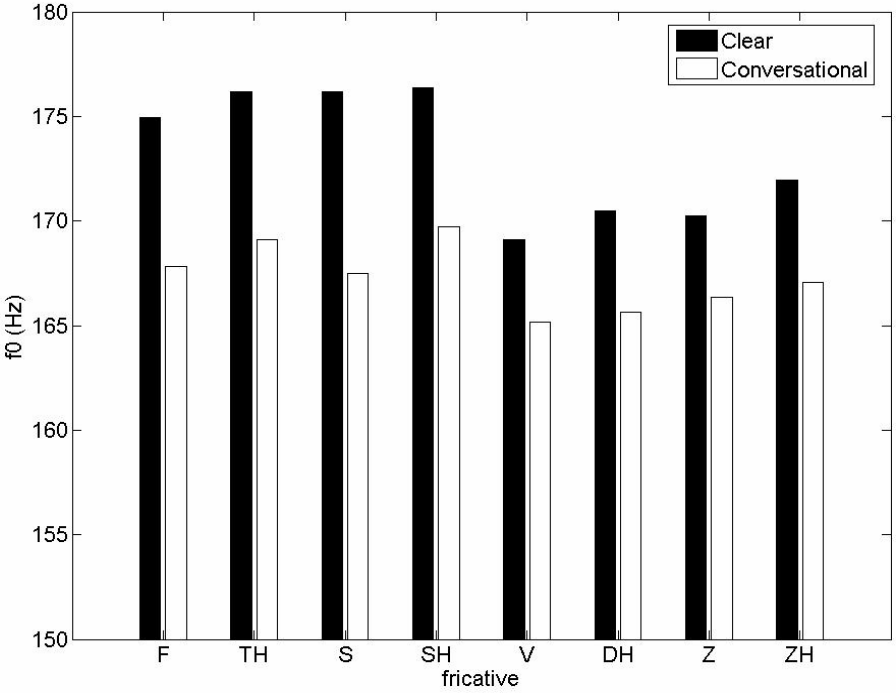


Figure 2-14: Mean f0 values (Hz) as a function of fricative and style

2.2.1.11 Discriminant analysis

As shown in Table 2-2, 84% of conversational fricative productions were classified correctly in terms of place of articulation by the discriminant analysis.

Classification accuracy was higher for sibilants than non-sibilants, and errors rarely involved the sibilant/non-sibilant distinction. Accuracy was slightly higher for clear productions. Specifically, (see Table 2-2), classification scores were higher for clear speech fricatives for 10 of 14 measures considered independently. Those measures that did not give better scores for clear speech included duration, intensity below 500 Hz, f0, and slope after the peak location, most of which are not generally considered to contribute to place of articulation identification.

For voicing, 94.1% of conversational and 96% of clear productions were classified correctly. As Table 2-2 shows, voicing classification was better for clear tokens for 9 of 14 independent measures. Measures that did not result in better performance for clear fricatives were in complimentary distribution with those that favored conversational productions for the place distinction, and are not generally considered to contribute to voicing identification.

Table 2-2: Classification scores (%) in terms of place of articulation (first two rows) and voicing (second two rows) distinction for each acoustic measures as a function of style (CL=clear, CO=conversational; Bold, higher score)

Distinction	Style	All	Dur	FFT	M1	M2	M3	M4	F2	SlpBef	SlpAft	RMS	RelAmp	HNR	Int500	f0
Place	CL	**85.3**	27.3	**51.3**	**48.3**	**65.5**	**57.1**	**56.3**	**38.5**	**68.6**	55.4	**49.6**	**47.2**	**37.1**	36	25
Place	CO	84	**31.1**	42.8	46.1	64.7	54.7	48.8	31.5	65.9	**61.3**	44.8	42.5	36.2	**37.6**	**25.4**
Voicing	CL	**96**	**79.7**	64.7	**81.9**	**61.6**	**77.4**	**72**	61.4	**73**	65.3	61.2	55	**79.5**	**85.5**	**65.3**
Voicing	CO	94.1	65.7	**68**	77.3	57.9	73.8	68.2	**63**	72.2	**71.3**	**76.3**	**55.9**	71.9	81.6	52.9

Table 2-4: Classification scores (%) in terms of minimal pair distinction as a function of context (CO=conversational, NON=noncontrastive clear, CTR=contrastive clear; Bold, highest score)

Style	/f/-/θ/	/f/-/s/	/f/-/ʃ/	/θ/-/s/	/θ/-/ʃ/	/s/-/ʃ/	/v/-/δ/	/v/-/z/	/v/-/ʃ/	/δ/-/z/	/δ/-/ʃ/	/z/-/ʃ/	/f/-/v/	/θ/-/δ/	/s/-/z/	/ʃ/-/ʃ/
CO	65	75	80	75	80	90	53	72	75	68	75	88	68	63	80	78
NON	67	77	82	74	82	95	57	75	77	67	74	92	75	70	85	83
CTR	**80**	77	**85**	**85**	**83**	**99**	**65**	**77**	**80**	68	**77**	**99**	**80**	**77**	**90**	**90**

2.2.2 Contrastive effects

2.2.2.1 Overall feedback effects

First, to assess overall effect of ‘listener’ feedback on production of each fricative, a 9 × 8 repeated measures ANOVA with Misapprehension and Fricative as within-subject factors was conducted for each acoustic parameter at each window location in clear speech. Results are summarized in the rightmost columns of Table 2-1. There were main effects of Fricative for many. Both Misapprehension effects and Misapprehension × Fricative were seen in several places, suggesting not only that the types of misperception influenced production in repeated clear speech but also that acoustic values for certain sounds were differentially affected by different types of misperception. In interpreting these results, the next step was to determine whether the difference between acoustic values for sounds in instances where one sound was just confused for the other, i.e. “contrastive” contexts is larger than non-contrastive contexts. For example, is the /s|ʃ/ (/s/ produced after misapprehension as /ʃ/) - /ʃ|s/ peak location difference larger than the /s|~ʃ/ (/s/ produced after misapprehended as /f/, /θ/, /v/, and ‘???’) - /ʃ|~s/ difference? Larger differences in “contrastive” contexts would suggest that speakers attempted to produce a fricative more unlike the sound for which it was just mistaken, by increasing the acoustic distance between these two similar sounds. A one-way repeated measures ANOVA with one within-subject factor (Context, i.e. contrastive vs. non-contrastive) and difference in acoustic values between two sounds (e.g. $|\text{peak}_s - \text{peak}_ʃ|$) in a minimal pair as the dependent variable was performed for each pair (*n* =16) for each acoustic measurement at all locations(*n*=53). The results of this comparison are summarized in Fig. 2-15; as seen in the figure, differences were greater much more often in contrastive contexts, reaching significance in a fair number of cases despite the relatively low power of this comparison. Critically, examination of individual results revealed that (1) acoustic parameters that showed significantly larger differences in contrastive context for place-of-articulation pairs were those considered to contribute to these distinctions (e.g. FFT peak location, all moments, RMS amplitude, relative amplitude and spectral slopes), and parameters that had larger differences in contrastive context for voicing pairs are associated with the voicing distinction (e.g. energy below 500 Hz, HNR, pitch, and less consistently F2, FFT peak location, moment 1 and 3), and (2) the difference was generally found to be larger for the pairs where the misapprehension differed from the target in only place or voicing and not by +/- sibilant.

For example, mean M1 values at W2 for /s/ and /ʃ/ in conversational speech were 6837 Hz and 4566 Hz, respectively. In /s|ʃ/ and /ʃ|s/ contexts, the value increased for /s/ (7611 Hz) but decreased for /ʃ/ (4540 Hz) while in /s|~ʃ/ and /ʃ|~s/ contexts, the value not only increased less for /s/ (7544 Hz) relative to the contrastive context but also instead of decreasing for /ʃ/, it increased (4644 Hz) compared to that value in conversational speech. In particular, M1 at W2 for /ʃ/ increased in /ʃ|ʒ/ context (4665 Hz) relative to that value in conversational speech to enhance the voicing distinction. Relative amplitude for /s/ and /ʃ/ in conversational speech were -6.561

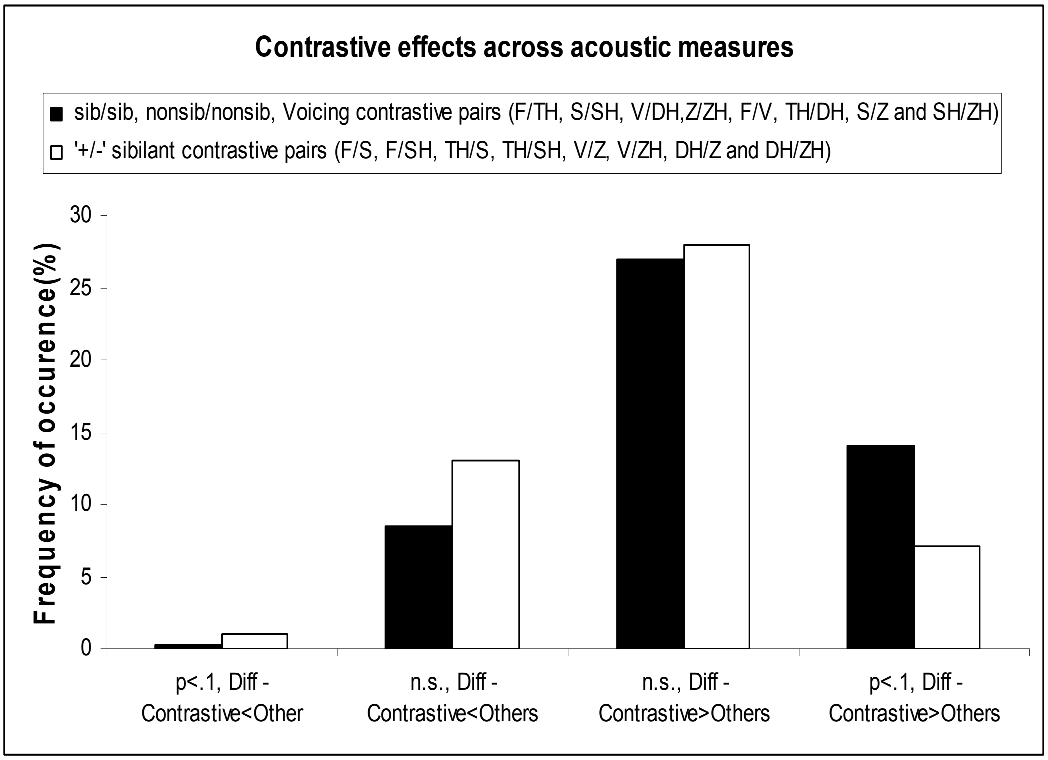


Figure 3-15: Frequency of occurrence (%) of differences in acoustic values between two sounds in a minimal pair as a function of pair type and context

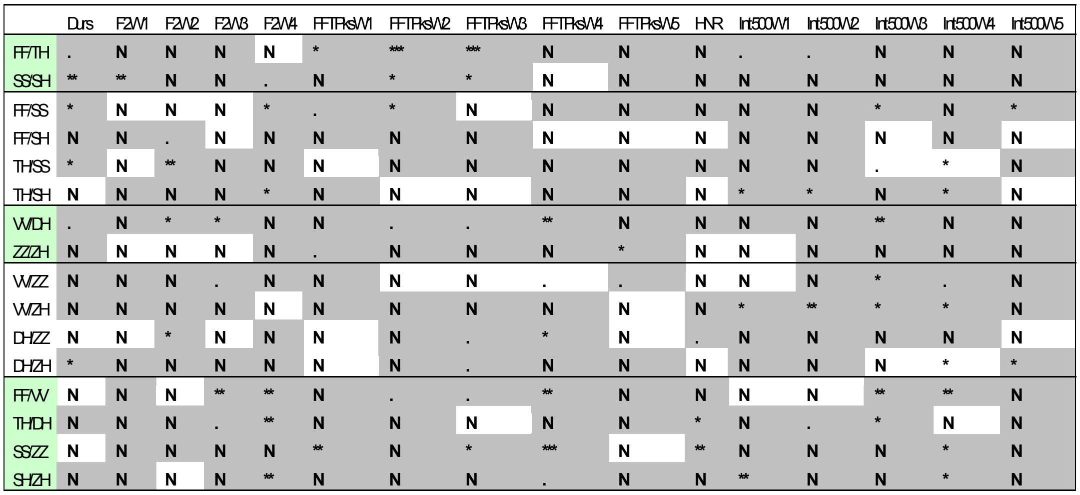
dB and 8.585 dB, respectively. Again, the value decreased to a larger extent for /s/ (-9.533 dB) and increased more for /ʃ/ (11.072 dB) in /s|ʃ/ and /ʃ|s/ contexts, relative to decrease for /s/ (-8.554 dB) and increase for /ʃ/ (10.353 dB) in /s|~ʃ/ and /ʃ|~s/ contexts. The same changes were found for the voiced pairs /z/ and /ʒ/. In addition, the decrease in relative amplitude for /s/ was smaller after misapprehension as /z/ (-8.275 dB) compared to the decrease for the same sound after misperception as /ʃ/. Similarly, intensity below 500 Hz for /f/ and /v/ in conversational speech were -

7.5185 dB and -3.644 dB, respectively. The value decreased for /f/ much more in /f|v/ context (-10.074 dB) than in /f|~v/ context (-9.45 dB) while the value for /v/ did not decrease in /v|f/ context (-3.6644 dB) as much as in /v|~f/ context (-4.4806 dB). Similar results were also found in many cases including /s/-/ʃ/ duration; /v/- /δ/ F2; FFT peaks for /f/-/θ/, /s/-/ʃ/, /v/-/δ/, /f/-/v/ and /s/-/z/; HNR for /θ/ and /δ/ and /s/-/z/; moment 1 for /f/-/θ/, /f/-/v/, and /ʃ/-/ʒ/ and moment 3 for /s/-/ʃ/, /z/-/ʒ/ and /f/-/v/ pairs. Summarization was found in Table 2-3.

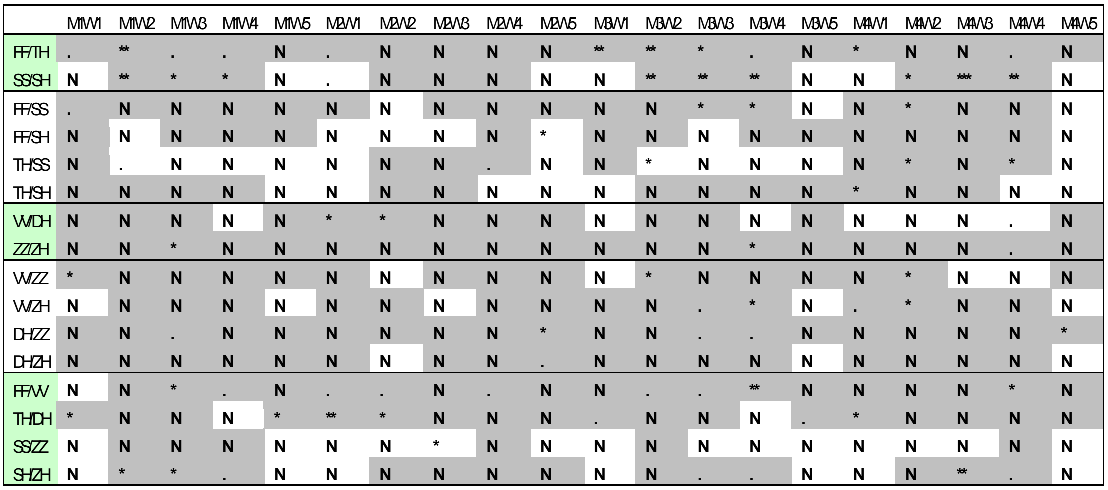
2.2.2.2 Discriminant analysis

As Table 2-4 shows, for 14 of 16 +/- [voicing], [place] pairs, classification scores in contrastive contexts were higher on average than other clear productions. In line with overall clear speech results (Table 2-2), classification scores were highest in contrastive contexts for most place pairs in parameters including F2, FFT peak location, spectral moments, amplitude, and spectral slope; classification scores were highest in contrastive contexts for voicing distinctions primarily in HNR, intensity below 500 Hz, f0, M1 and M3.

Table 2-3: Summary of results from the one-way repeated measures ANOVA with one within-subject factor (Context, i.e. contrastive vs. non-contrastive) and difference in acoustic values between two sounds as a dependent variable for each of 53 acoustic predictors, ****p*<.001, ***p*<.01, **p*<.05, ·*p*<.1, N *p*>.1, shaded = acoustic distance is larger in the 'contrastive' pair than in the 'non-contrastive' pair)



	Durs	F2W1	F2W2	F2W3	F2W4	FFTPksW1	FFTPksW2	FFTPksW3	FFTPksW4	FFTPksW5	HNR	Int500W1	Int500W2	Int500W3	Int500W4	Int500W5
FF/TH	.	N	N	N	N	*	***	***	N	N	N	.	.	N	N	N
SS/SH	**	**	N	N	.	N	*	*	N	N	N	N	N	N	N	N
FF/SS	*	N	N	N	*	.	*	N	N	N	N	N	N	*	N	*
FF/SH	N	N	.	N	N	N	N	N	N	N	N	N	N	N	N	N
TH/SS	*	N	**	N	N	N	N	N	N	N	N	N	N	.	*	N
TH/SH	N	N	N	N	*	N	N	N	N	N	N	*	*	N	*	N
W/DH	.	N	*	*	N	N	.	.	**	N	N	N	N	**	N	N
ZZ/ZH	N	N	N	N	N	.	N	N	N	*	N	N	N	N	N	N
W/ZZ	N	N	N	.	N	N	N	N	.	.	N	N	N	*	.	N
W/ZH	N	N	N	N	N	N	N	N	N	N	N	*	**	*	*	N
DH/ZZ	N	N	*	N	N	N	N	.	*	N	.	N	N	N	N	N
DH/ZH	*	N	N	N	N	N	N	.	N	N	N	N	N	N	*	*
FF/W	N	N	N	**	**	N	.	.	**	N	N	N	N	**	**	N
TH/DH	N	N	N	.	**	N	N	N	N	N	*	N	.	*	N	N
SS/ZZ	N	N	N	N	N	**	N	*	***	N	**	N	N	N	*	N
SH/ZH	N	N	N	N	**	N	N	N	.	N	N	**	N	N	*	N



	M1W1	M1W2	M1W3	M1W4	M1W5	M2W1	M2W2	M2W3	M2W4	M2W5	M3W1	M3W2	M3W3	M3W4	M3W5	M4W1	M4W2	M4W3	M4W4	M4W5
FF/TH	.	**	.	.	N	.	N	N	N	N	**	**	*	.	N	*	N	N	.	N
SS/SH	N	**	*	*	N	.	N	N	N	N	N	**	**	**	N	N	*	***	**	N
FF/SS	.	N	N	N	N	N	N	N	N	N	N	N	*	*	N	N	*	N	N	N
FF/SH	N	N	N	N	N	N	N	N	N	*	N	N	N	N	N	N	N	N	N	N
TH/SS	N	.	N	N	N	N	N	N	.	N	N	*	N	N	N	N	*	N	*	N
TH/SH	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	*	N	N	N	N
W/DH	N	N	N	N	N	*	*	N	N	N	N	N	N	N	N	N	N	N	.	N
ZZ/ZH	N	N	*	N	N	N	N	N	N	N	N	N	N	*	N	N	N	N	.	N
W/ZZ	*	N	N	N	N	N	N	N	N	N	N	*	N	N	N	N	*	N	N	N
W/ZH	N	N	N	N	N	N	N	N	N	N	N	N	.	*	N	.	*	N	N	N
DH/ZZ	N	N	.	N	N	N	N	N	N	*	N	N	.	.	N	N	N	N	N	*
DH/ZH	N	N	N	N	N	N	N	N	N	.	N	N	N	N	N	N	N	N	N	N
FF/W	N	N	*	.	N	.	.	N	.	N	N	.	.	**	N	N	N	N	*	N
TH/DH	*	N	N	N	*	**	*	N	N	N	.	N	N	N	.	*	N	N	N	N
SS/ZZ	N	N	N	N	N	N	N	*	N	N	N	N	N	N	N	N	N	N	N	N
SH/ZH	N	*	*	.	N	N	N	N	N	N	N	N	.	.	N	N	N	**	.	N

	F0	RelAmps	RMSW1	RMSW2	RMSW3	RMSW4	RMSW5	SlpAftW1	AftW2	AftW3	AftW4	AftW5	SlpBefW1	BefW2	BefW3	BefW4	BefW5
FF/TH	N	N	*	N	N	N	N	N	N	N	N	N	N	.	*	N	N
SS/SH	N	***	N	N	N	N	N	N	N	*	N	N	.	N	N	N	N
FF/SS	**	N	N	N	.	*	*	N	N	N	N	N	N	N	N	*	N
FF/SH	.	N	N	N	N	N	N	N	N	N	N	*	N	N	N	N	N
TH/SS	N	N	N	N	N	N	N	N	N	N	.	N	N	.	N	N	N
TH/SH	.	N	N	N	N	N	N	N	N	*	N	N	N	N	N	N	N
VV/DH	N	N	N	*	**	N	N	N	N	.	N	N	N	*	*	N	N
ZZ/ZH	N	*	.	*	N	.	N	.	N	N	N	N	N	N	N	N	N
VV/ZZ	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
VV/ZH	N	N	.	N	N	N	N	N	N	N	*	N	N	N	N	N	N
DH/ZZ	N	N	N	N	N	N	N	N	N	N	N	N	N	N	.	N	N
DH/ZH	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
FF/VV	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
TH/DH	.	N	*	N	N	N	N	N	N	N	N	N	.	N	N	N	N
SS/ZZ	N	N	N	N	N	N	N	N	N	*	N	.	N	N	N	N	N
SH/ZH	.	N	***	N	N	.	*	.	N	N	*	N	N	.	N	N	.

3. DISCUSSION

3.1 Overall clear speech modifications

The results of this study indicate that systematic modifications occur during the production of clear fricatives. Across speakers and fricatives, frication duration significantly increased, and spectral measures including peak location, mean and skewness of the energy distribution, and F2 frequency showed energy concentration in higher frequency regions, in clear speech. These results are in agreement with previous studies (*e.g.* Chen, 1980; Picheny *et al.*, 1986). Steeper spectral slopes also resulted in more defined peakedness and more positive before-peak slopes indicated greater noise source strength, consistent with reports on fricatives produced at elevated voice levels (Jesus and Shadle, 2002). Neighboring vowel F0 was variable but also tended higher in clear speech. Lower amplitude measures compared to neighboring vowels were somewhat unexpected considering numerous reports of increased CVR in clear speech (e.g. Bradlow *et al.*, 2003; Chen, 1980), but not completely surprising. Previous studies have not concentrated on fricatives, and in general have shown that changes in CVR are stimulus-, context-, and talker-dependent; decreases have even been seen for some fricatives (mostly non-sibilants) for some speakers (Picheny *et al.* 1986; Krause and Braida, 2004). The present results are probably best explained in terms of articulatory effort. Since the volume velocity required to increase the level of fricative sounds—particularly non-sibilants—is much greater than that required to increase vowel intensity by a similar amount, it is not surprising that, for a similar increase in effort across a word (or even slightly more effort on a fricative), intensity would increase more for vowels than for fricatives (especially non-sibilants).

3.2 Inventory-level contrastive patterns

Style × Place and Style × Voicing interactions for several measures were consistent with efforts to maintain contrasts within the fricative inventory. Inherently longer voiceless fricatives increased more in duration in clear speech than voiced fricatives, effectively increasing the distance between the two classes of sounds in terms of duration. Enhanced voicing contrasts were also seen: increased M2 values in voiced fricatives and decreased values in voiceless, a larger decrease in intensity below 500 Hz for voiceless fricatives, an increase in HNR for voiced fricatives and decrease for voiceless, and an increase in f0 only for voiceless fricatives.

Place of articulation contrasts were also enhanced in clear speech. For example, palato-alveolars are defined by energy concentration at low frequencies; FFT peaks and M1 for palato-alveolars increased much less or decreased in clear speech, and skewness increased or did not decrease as much as for other fricatives. Differences between sibilants and non-sibilants were also emphasized in clear speech. Non-sibilants with inherently more diffuse spectra increased in M2 while sibilants decreased in M2 in clear speech. Non-sibilants also decreased in kurtosis whereas voiceless sibilants increased in clear speech. Acoustic distance between sibilants and non-sibilants also increased in terms of amplitude; a significant decrease in normalized RMS amplitude in clear speech was seen only from non-sibilant fricatives. F2 increased most for palato-alveolars (with inherently higher F2 than alveolars), and dentals (with inherently higher F2 than labio-dentals); in fact, F2 was the only measure in which labio-dentals and dentals significantly differed in conversational speech, so this difference was important in contributing to the distinction. Distance between alveolar fricatives and palato-alveolars was also larger in terms of relative amplitude in clear speech; palato-alveolars increased and alveolars decreased. Thus, while it cannot be shown that changes were a direct result of knowledge of the fricative inventory and its critical contrasts, and while the actual effect of these modifications on the effectiveness of the contrasts must be evaluated through perceptual study, the pattern of results seen was consistent with the notion that clear speech acts to maximize contrast within a language (e.g. Bradlow *et al.*, 2003; Chen, 1980; Krause and Braida, 2004; Ohala, 1994; Picheny *et al.*, 1986; Smiljanić and Bradlow, 2005).

3.3 Local contrastive effects

Effects of Misapprehension and Misapprehension × Fricative interactions for several measures indicated that speakers were influenced by the online feedback provided by the interactive elicitation method. Moreover, comparison of acoustic distances between clear fricative pairs across measures and misapprehension patterns revealed that speakers tended to repeat sounds such that they differed maximally from the sounds for which they were initially mistaken. For example, after /ʃ/ was mistaken for /s/ it was repeated on average with a lower mean frequency and peak location; when it was mistaken for /ʒ/, however, it was repeated with a higher mean frequency and peak location, effectively enhancing the use of this dimension in signaling place of articulation and voicing, respectively. It should be noted that many of these effects were small in magnitude compared to the general and fricative-level clear speech modifications on which they were superimposed (see sections in 2.2 and table 2-1). They are important, however, in demonstrating the range of levels at which talkers are sensitive to the communicative demands of a speaking situation, and are consistent with the notion that talkers are able to adjust the details of productions based on relatively local, fine-grained information.

3.4 Discriminant analysis

Disriminant analyses showed that clearly produced fricatives resulted in higher classification accuracy in terms of both place of articulation and voicing distinctions. 10 out of 14 acoustic measures showed higher scores in clear speech for place distinctions, including all of those generally considered to contribute primarily to place identification (e.g. peak location, spectral moments, F2, slope before the peak). Likewise, 9 of 14 measures increased in clear speech for the voicing distinction, primarily those considered to contribute to voicing (e.g. duration, intensity below 500 Hz, HNR, f0). Again, this indicates that, on average, clear speech resulted in a more distinct, better separated inventory of fricative categories

(Hypothesis 2). Moreover, within the clear speech data set, minimal pair classification was better for every pair except two, i.e., /f/-/s/ and /ð/-/z/, in the contrastive context than in other contexts, indicating again that talkers were sensitive to local communicative demands.

3.5 Talker Effects

Style × Talker effects for most measures indicate that talkers varied significantly in the magnitude and sometimes direction of acoustic modifications in clear speech. For example, some speakers lengthened clear fricatives 4.5 – 5 times relative to conversational productions, while the smallest increase found was about 1.7 times. For relative amplitude, some speakers decreased precipitously in clear speech, while others increased substantially. Mean frequency (M1), F2, and slopes also showed high talker variance across styles.

3.6 Conclusions

This study demonstrates that there are systematic acoustic-phonetic modifications in the production of clear fricatives. Some overall clear speech effects were straightforwardly predictable based on previous findings (e.g. longer duration, energy at higher frequencies), and some were more surprising (esp. lower relative amplitude). Across a variety of measures, the acoustic distances between minimally contrasting sounds was enlarged in clear speech, indicating that talkers attempt to maintain contrast between category distributions across the inventory of English fricatives. In addition, talkers were sensitive to local listener feedback, adjusting repeated productions to be more unlike sounds that they had been misapprehended for. Individual talkers varied widely in the magnitude and sometimes direction of these changes; these differences were not related to talker gender. Questions left to future research include (1) whether 'repeated' clear fricatives are actually more, (2) whether differences in perception can be attributed to different acoustic strategies employed by talkers, and (3) how different listener populations including listeners with hearing loss, cochlear implant users, and nonnative listeners, perceive clear fricatives.

Chapter 3

Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners

1. INTRODUCTION

Previous research on consonant recognition and confusion indicates that fricative consonants, especially non-sibilants, present considerable identification difficulty for hearing-impaired listeners and for normal listeners under adverse conditions (Boothroyd, 1984; Dubno and Levitte, 1981; Dubno et al., 1982; Miller and Nicely, 1955; Owens, 1978; Owens et al., 1972; Sher and Owens, 1974; Singh and Black, 1966; Soli and Arabie, 1979; Wang and Bilger, 1973). This study was designed to measure whether, and how, speakers may be able to alleviate this difficulty by deliberately producing fricatives more clearly.

1.1. Clear speech intelligibility advantage

Lindblom (1990) maintains that speakers can adapt the phonetic details of their speech in response to on-line social and communicative demands of a speech situation, adopting an intelligibility-enhancing speaking style when they anticipate or sense perceptual difficulty or comprehension failure on the part of a listener (due to, e.g., background noise, reverberation, hearing impairment, lack of linguistic/world knowledge). "Clear speech" has been elicited in laboratory settings (e.g., Bradlow and Bent, 2002; Bradlow *et al.*, 2003; Ferguson and Kewley-Port, 2002; Gagné *et al.*, 1994, 1995, 2002; Helfer, 1997, 1998; Iverson and Bradlow, 2002; Krause and Braida, 2002; Liu *et al.*, 2004; Payton *et al.*, 1994; Picheny *et al.*, 1985; Schum, 1996; Uchanski *et al.*, 1996), and the intelligibility advantage seen for clear speech with sentence stimuli relative to "conversational" speech has ranged from 7 to 38 percentage points. Clearly spoken sentences have been shown to benefit young normal-hearing listeners in noise and/or reverberation (Bradlow and Bent, 2002; Gagné *et al.*, 1995; Krause and Braida, 2002; Payton *et al.*, 1994; Uchanski *et al.*, 1996) and with simulated hearing loss or cochlear implants (Iverson and Bradlow, 2002; Liu *et al.*, 2004), hearing-impaired listeners in quiet (Picheny *et al.*, 1985; Uchanski *et al.*, 1996) and in noise or reverberation (Payton *et al.*, 1994; Schum, 1996), elderly listeners with or without hearing loss (Helfer, 1998; Schum, 1996), cochlear-implant users (Iverson and Bradlow, 2002; Liu *et al.*, 2004), children with or without learning disabilities (Bradlow *et al.*, 2003) and (to a lesser extent) nonnative listeners (Bradlow and Bent, 2002). It has been suggested that clear speech intellibility benefits are roughly independent of presentation level, frequency-gain characteristics, and type of degradation, and may increase with increasingly adverse conditions (Bradlow and Bent, 2002; Iverson and Bradlow, 2002; Payton *et al.*,1994).

Recent results from Ferguson and Kewley-Port (2002) call into question the robustness of the "clear speech effect" and suggest that hyperarticulation strategies may interact in more complicated ways with specific types of signal degradation. While Ferguson and Kewley-Port saw intelligibility benefits for clearly produced vowels for young, normal-hearing listeners, they actually observed *negative* clear-speech intelligibility benefits (i.e., better recognition of conversational tokens) in elderly hearing-impaired listeners for the productions of one male talker. This pattern was mostly due to front vowels, for F2 frequency was a primary cue for the elderly listeners. A hallmark of clear speech is a greater concentration of energy in higher frequencies, in terms of both overall spectral distributions and individual formant frequencies (e.g. Krause and Braida, 2004; Picheny *et al.*, 1986); in this case, since the average F2 values for front vowels fell in a frequency region where these listeners had sloping hearing loss (above 2000 Hz), clear vowels' higher F2 resonances, on average, fell in regions of greater impairment than those of conversational vowels.

It is of course unclear whether the patterns observed for this talker are unique to him or whether they are typical of the production, and perception by hearing-impaired or older listeners, of clear front vowels. It is also unclear whether the absence of a clear speech advantage for elderly hearing-impaired listeners would hold for other sounds by even this same talker, since his sentence stimuli (and back vowels) did show a clear speech advantage. The present study was designed to determine whether clear speech advantages occur for another class of sounds with a preponderance of high-frequency energy (fricatives) over a wide range of talkers and for both normal young listeners and listeners with simulated hearing loss.

1.2 Talker-related acoustic correlates of clear speech intelligibility

A secondary goal of this study was to determine which aspects of clear fricative production influence intelligibility. Previous investigations of the intelligibility of clear and conversational speech that have included more than a single talker have revealed interactions between talker and speaking style; that is, there were significant differences in the magnitude of the clear speech effect across talkers (e.g. Bradlow *et al.*, 2003; Chen, 1980; Ferguson, 2002, 2004; Gagné *et al.*, 1994, 1995; Schum, 1996). For example, the intelligibility benefit for sentences produced by the female talker in Bradlow *et al*. (2003) was significantly greater than for sentences produced by the male talker. Consonants and vowels produced by two speakers in Chen's (1980) study showed similar clear speech advantages, while the intelligibility of a third speaker's productions was about the same in both styles. Similar variability was seen in several multi-talker studies by Gagné *et al.* (1994, 1995, 2002), using sentence, CV, and VCV stimuli; in one case, clear-speech benefits ranged from -17% to 39%.

A few studies have attempted to identify talker-specific acoustic-phonetic parameters that may be responsible for the enhanced speech perception, relating intelligibility differences to acoustic differences in clear and conversational speech. The female talker in Bradlow *et al.* (2003), who showed the greater intelligibility advantage for clear speech, substantially decreased her speaking rate with increased frequency and duration of pauses. This pattern led the investigators to infer that modifications to the temporal characteristics of the signal contributed most to intelligibility enhancement, although it is not clear whether the pattern would hold for more than two speakers. Ferguson (2002) attempted to relate intelligibility differences in clear speech to acoustic-phonetic differences on a larger scale. Ferguson compared ten vowel measurements (five steady-state metrics, four dynamic metrics, and duration) between the six speakers (of a total of 41) showing the largest clear-speech benefit in perception by normal-hearing listeners, and the six showing the smallest benefit, to determine which acoustic modifications contributed most to intelligibility enhancement. The amount of increased vowel duration and formant movement were similar for both groups while the "big benefit" talkers showed considerably greater increases in front vowel F2, F1 range, and the overall size of the vowel space. This suggests that increasing vowel duration and making vowels more dynamic were less important to improved vowel intelligibility, but that raised F2 for front vowels, expanded overall vowel space and F1 range led to effective enhancement.

While the primary goal of this study was to determine the overall effects of clear speech on the intelligibility of fricative contrasts in different conditions, an analysis along the lines of Ferguson's (2002) study was also employed. Productions from twenty talkers were used, for which 59 spectral, temporal, and amplitudinal measurements were previously reported (Maniwa *et al.*, submitted). Correlation analysis of acoustic and intelligibility differences across talkers was performed to assess the contributions of acoustic modifications to intelligibility.

1.3 Perception of English fricative sounds

Acoustic components that have been reported by previous studies to affect perception of English fricative place of articulation for listeners with normal hearing include frication duration (Hughes and Halle, 1956; Hedrick, 1997; Hedrick and

Carney, 1997; Hedrick and Younger, 2003; Jongman, 1989; Whalen, 1991), frication spectrum (Heinz and Stevens, 1961; Harris, 1958; Hedrick and Ohde, 1993; Hughes and Halle, 1956; Nittrouer, 1992; Nittrouer and Miller, 1997 a and b; Nittrouer, 2002; Zeng and Turner, 1990), and overall frication amplitude normalized to the neighboring vowel (Guerlekian, 1981; Heintz and Stevens, 1961; McCasland, 1979 a and b). Properties of or in combination with neighboring vowels that have been thought to influence perception of place of articulation include formant transitions (Harris, 1958; Hedrick, 1997; Hedrick and Carney, 1997; Hedrick and Younger, 2003; Heintz and Stevens, 1961; McCasland, 1978; LaRiviere *et al.*, 1975; Mann and Repp, 1980; Nittrouer, 1992, 2002; Zeng and Turner, 1990; Whalen, 1981), frequency-specific relative amplitude (Hedrick, 1997; Hedrick and Carney, 1997; Hedrick and Ohde, 1993; Hedrick and Younger, 2003; Stevens, 1985), and vowel quality (Harris, 1958; Mann and Repp, 1980; Mann and Soli, 1991; Yeni-Komshian and Soli, 1981, Whalen, 1981). Briefly, experiments using natural (Harris, 1958; Zeng and Turner, 1990), synthetic (Heintz and Stevens, 1961; Zeng and Turner, 1990), and hybrid (Nittrouer, 1992, 2002; Nittrouer and Miller, 1997 a and b) speech suggest that spectral cues are important for distinguishing sibilants, and seem to override transition cues for these sounds. On the other hand, formant transition cues may help to distinguish nonsibilants (Harris, 1958; Heintz and Stevens, 1961; Nittrouer, 2002) and take on more weight when spectral cues are ambiguous (Hedrick, 1997; Hedrick and Carney, 1997; Hedrick and Ohde, 1993; Hedrick and Younger, 2003; Whalen, 1981). Overall noise duration and amplitude seem to have less perceptual significance (Behrens and Blumstein, 1988; Hedrick, 1997; Hedrick and Carney, 1997; Hedrick and Ohde, 1993; Hughes and Halle, 1956; Jongman, 1989; *cf.* Guerlekiean, 1981; McCasland, 1979a and b); on the other hand, manipulation of frication amplitude relative to vowel amplitude in a particular frequency region does influence listeners' perception of place of articulation for /s/-/ʃ/ and /s/-/θ/ contrasts (Hedrick, 1997; Hedrick and Carney, 1997; Hedrick and Ohde, 1993; Hedrick and Younger, 2002; Stevens, 1985).

Somewhat fewer studies have investigated which acoustic components serve to distinguish voiced and voiceless fricatives, and most of them have concentrated on duration of noise in syllable-initial fricatives (Cole and Cooper, 1975) and preceding vowel duration for syllable-final fricatives (e.g. Denes, 1955; Raphael, 1972; Soli, 1982) as a cue. Stevens et al. (1992) showed that noise duration, the amplitude and duration of glottal vibration at the edge of the fricative, and the extent of F1 transitions appear to interact in determining listener judgments of voicing for intervocalic fricatives.

1.4 Perceptual cue weighting and fricative perception by normal and hearing impaired listeners

It seems that listeners do not process acoustic cues independently, but integrate information obtained from several dimensions in identifying fricatives. Furthermore, the perceptual weights assigned to acoustic properties may differ depending on contexts and listeners (Best and Strange, 1992; Christiansen and Humes, 1996, 1997; Crowther and Mann, 1992, 1994; Fledge, *et al.*, 1996; Hazan and Rosen, 1991; Nittrouer, 1992; Nittrouer and Miller, 1997a and b). Adult listeners with normal-hearing seem to make more use of spectral characteristics for place of articulation information (Heinz and Stevens, 1961; Harris, 1958; Hedrick and Ohde, 1993; Hughes and Halle, 1956; Nittrouer, 1992; Nittrouer and Miller, 1997 a and b; Nittrouer, 2002; Zeng and Turner, 1990), and temporal information for the voicing distinction (Cole and Cooper, 1975; Raphael, 1972; Soli, 1982). Hearing-impaired listeners may have difficulty integrating amplitude and spectral cues, and may generally place less weight on formant transitions than do listeners with normal hearing for labeling fricative place of articulation perception, even when formant transition information should be audible to them (Hedrick, 1997; Hedrick and Younger, 2003; Zeng and Turner, 1990). In addition, duration may play a more important role for hearing-impaired listeners; they may need a longer time period to process acoustic information. These issues are of course in addition to, and may apply differentially depending on, the effects of frequency-dependent hearing level. In particular, listeners with sloping hearing loss commonly have elevated thresholds, and reduced dynamic range, in regions relevant to fricative perception (e.g., Dubno *et al.*, 1982; Owens *et al.*, 1972; Sher and Owens, 1974). To further address these issues, this study examined the perception of clear and conversational fricatives both by normal-hearing listeners (Experiment 2) and listeners with simulated hearing impairment (Experiment 3). Of particular interest was whether effects of speaking style, and identifiable acoustic correlates, would differ depending on listener population.

1.5 Hypotheses

Two experiments were performed to address three questions. First, are clearly produced fricatives more intelligible than conversational fricatives for listeners with normal hearing in degraded conditions? Based on previous findings, we hypothesized that they would, although the effects might vary depending on fricatives (e.g. Ferguson and Kewley-Port, 2002). Second, what acoustic modifications are related to intelligibility? It was hypothesized that not all strategies employed by talkers serve to improve fricative identification. Third, do clear-speech intelligibility differences differ based on listener population, in particular for listeners with sloping hearing losses? We expected that hearing loss might interact with clear-speech strategies, perhaps resulting in reduced benefit where high-frequency information was critical.

2. Experiment 2: Effects of clear speech for fricative recognition by listeners with normal hearing

2.1 Method

2.1.1 Participants

Fourteen normal-hearing listeners (8 F, 6 M) aged between 19 and 32 were recruited from the University of California, Berkeley community. Participants were native speakers of American English, without noticeable regional dialects. Participants reported normal hearing and no history of speech or language disorders. Listeners were paid for their participation in the experiment.

2.1,2 Materials

Fricative intelligibility in clear and conversational speech was assessed using a database of 8800 VCV ([a]-fricative-[a]) stimuli produced by 20 speakers (10 F, 10 M) as part of a previous acoustic study of clearly produced fricatives (Maniwa *et al.*, submitted). Briefly, conversational and clear tokens were elicited using an interactive program that ostensibly attempted to identify the sequence of fricatives produced by a speaker. The program made frequent, systematic errors involving voicing and place alternations, after which the speaker repeated a sound more clearly, as if trying to disambiguate the production for an elderly or hearing-impaired listener. To eliminate amplitude differences among talkers and between the two speaking styles, all stimuli were normalized to the same long-term (word-level) RMS amplitude and presented at 60dB SPL using MATLAB (The Math Works, Inc., 2000). Test stimuli were delivered in a background of 12-talker (6 F, 6 M) babble recorded at a sampling rate of 44.1 kHz. A total of 60 s of babble was created for the purposes of the experiment; for each stimulus, a segment of babble was selected from a random location within this 60-s sample. The duration of this segment exceeded that of the test item by a total of 600 ms, with the test stimulus centered temporally in the babble. There were 5-ms and 100-ms linear on-off ramps for the target stimulus and the noise, respectively.

The goal of each test was to determine the signal-to-noise ratio (SNR) threshold at which a distinction can be made with 75% accuracy. An adaptive procedure was used to estimate this threshold and choose SNR values for the trials.

2.1.3 Procedures and apparatus

The perception test employed a two-alternative forced-choice identification task. The 8 fricatives were divided into 8 minimal pairs, depending on place of articulation and voicing: /f/-/θ/, /v/-/ð/, /s/-/ʃ/, /z/-/ʒ/, /f/-/v/, /θ/-/ð/, /s/-/z/, and /ʃ/-/ʒ/. Each pair was tested separately for clear and conversational styles, for a total of 16 sub-tests. Subjects listened to the stimuli presented via Koss headphones in sound-attenuated rooms, seated in front of a computer monitor and mouse. On each trial, a test syllable (VCV) and a segment of babble noise were selected at random from the appropriate production condition and the pre-recorded babble sample. The two waveforms were scaled based on the selected SNR and the constant target stimulus level, combined additively, and presented binaurally to the subjects, who were prompted to identify the fricative of the VCV stimulus from a minimal pair by using the mouse to click one of 2 letter combinations on the computer screen. Two response alternatives were displayed on the computer and written: “ff”, “th”, “ss”, “sh”, “vv”, “dh”, “zz”, and “zh”.

The experiment consisted of a 1-hour session, involving 16 tests (8 place or voicing pair × 2 styles). Test order was randomized across subjects. Each test included 80 stimuli, picked randomly from the productions of the 20 speakers.

Tokens were selected randomly across participants, so that individual productions would, on average, occur with equal frequency. First, listeners were oriented to the spelling of response alternatives on the screen with the written instruction. Each participant was familiarized with the test procedure prior to beginning the experimental conditions. For familiarization, a 10-trial block of fricative tokens at a high SNR (+10dB) was run with feedback before each test. Within experimental blocks, two 40-trial adaptive tracks were initiated at +3dB and -3dB and interleaved at random over the 80-trial block. SNR values for each track were selected using a Bayesian adaptive algorithm (ZEST; e.g., King-Smith *et al.*, 1994). The final threshold estimate was simply taken as the average (in dB) of the SNR values for each track on the final (40th) trial. While this approach may have resulted in less precise measurements of thresholds that were further from the initial guesses (since termination was not based on confidence criteria) it was considered more important that participants were exposed to equal numbers of stimuli from each contrast pair. +/-20dB were chosen as absolute maximum and minimum allowable SNR values.

2.1.4 Data analysis

Clear speech intelligibility effect was tested using a repeated measures analysis of variance (ANOVA) with two within-subject factors (Style; 2 levels, Pair; 8 levels) and threshold (dB SNR) as the dependent variable. In order to assess the effect of pair type more thoroughly, another repeated measures ANOVA was calculated with three within-subject factors. One of the factors was Style. The second factor was labeled depending on whether the pair consisted of sibilant fricatives or non-sibilant fricatives (e.g. /s/-/ʃ/ and /f/-/θ/, respectively). The third factor was labeled depending on whether the pair involved a place or voicing distinction. Pairwise comparisons for significant within-subject factors were done using Bonferroni corrected 95% confidence intervals.

In addition, as a first step in determining which acoustic modifications were related to intelligibility, correlation analyses were carried out across the 20 speakers included in the experiment, relating differences in their production strategies to differences in their clear-speech benefit. First, for each speaker, a single clear-minus-conversational difference value, averaged over all fricatives and productions, was calculated for each of the 59 acoustic measures reported in the Maniwa et al. (submitted) study. In the acoustic study, 14 parameters were measured: FFT spectral peak location (1), the first four spectral moments moments (2-5), F2 onset transitions (6), spectral slopes below (7) and above (8) typical peak locations, pitch of adjacent vowels (9), overall RMS amplitude (10), relative amplitude i.e., a change in amplitude of the frication relative to the vowel in F3 region for sibilants and F5 region for non-sibilants (11), harmonic-to-noise ratio (HNR; 12) energy below 500 Hz (13), and fricative duration (14). For (1)-(5), (7)-(8), (10) and (13), analyses considered 40-ms Hamming windowed segments at five locations, centered over the fricative onset, 25, 50, and 75% points, and offset (window (W) 1-5). For (6), acoustic values were derived at fricative onset and offset and each vowel midpoint from an analysis (W1-4). For (9), f0 was averaged across the vowels preceding and following the target. For (11), (12) and (14), the values were obtained over the entire fricative. In the present correlation analyses, LPC peaks (at the same 5 locations) were included as well, and f0 was considered separately preceding and following the fricative. Thus, the total number of acoustic values in the correlation analyses was 59.

Next, a similar overall clear-minus-conversational *intelligibility* difference had to be estimated for each speaker; this was of course less straightforward given the adaptive procedure used in the experiment. First, we verified that over the 32 total adaptive tracks that each listener in Experiment 1 heard, tokens from different speakers occurred, on average, with equal frequency and at equal signal-to-noise ratios. Then we simply took the clear-minus-conversational difference in accuracy (% correct), averaged across listeners, sub-tests, and SNR values, for each speaker as that speaker's approximate clear speech intelligibility advantage. While listener, sub-test, and SNR certainly all contributed to mean accuracy, we assumed that these contributions would essentially amount to random variability across speakers, making our measure of intelligibility advantage more conservative, and therefore no corrections were made based on these variables.

It is important to note here that this comparison was limited in the types of acoustic-perceptual relationships it could detect. As reported by Maniwa *et al.* (submitted), clear fricatives were characterized not only by overall differences in the 59 acoustic measures depending on speaking style, but by numerous and complex Style × Fricative interactions. Since correlation analysis capable of capturing these higher-order acoustic differences was not feasible given the constraints of the perception experiments described here (individual speakers were not represented well enough within subtests to ensure equalized average SNR), we did not consider these patterns in the present study.

2.2 Results and discussion

2.2.1 Fricative intelligibility for listeners with normal hearing

Figure 3-1 shows mean speech-to-noise ratio (SNR, dB) thresholds as a function of fricative pair and speaking style. The Style × Pair ANOVA showed an effect of Style [$F(1,13)=149.551$, $p<.001$], with lower thresholds for clear speech than for conversational speech. The Pair effect was also significant [$F(7,91)=113.830$, $p<.001$]; across speaking styles, thresholds were lowest for the voiceless sibilant place of articulation contrast /s/-/ʃ/, followed by /s/-/z/ and /ʃ/-/ʒ/. Non-sibilant place of articulation pairs /f/-/θ/ and /v/-/δ/ were the most difficult, in accordance with previous studies (e.g. Jongman, *et al.*, 2000b; Miller and Nicely, 1955; Wang and Bilger, 1973).. The Style × Pair interaction was marginally significant [$F(7, 91)=2.107$, $p=.051$], probably due to pairs /v/-/δ/ and /f/-/v/. *Post-hoc* comparisons revealed that the “clear speech effect” did not reach significance for these two pairs; all other pairs showed significant clear speech advantages.

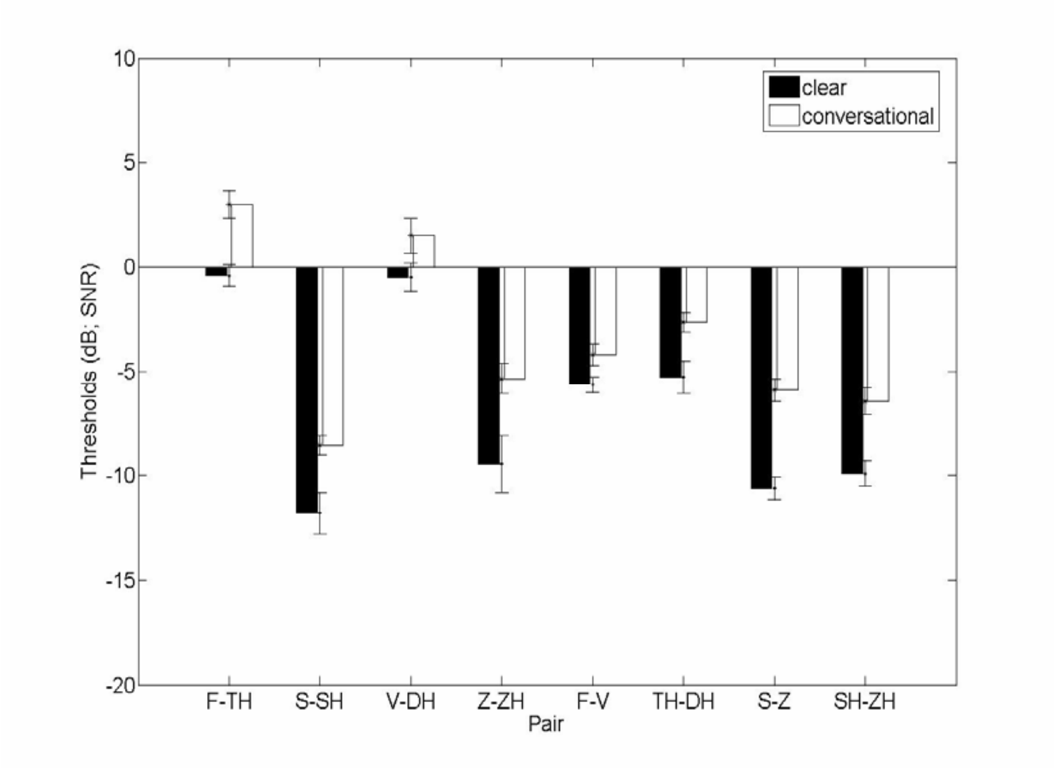


Figure 3-1: Speech-to-noise (SNR) thresholds (dB) as a function of style and fricative pair in Experiment 2

The Style × Sibilance × Distinction type ANOVA revealed a main effect of Sibilance [*F*(1, 27)=370.960, *p*<.001] with lower thresholds for sibilants than for non-sibilants. The main effect of Distinction was also significant [*F*(1, 27)=103.711, *p*<.001] with lower thresholds for voicing distinctions relative to place of articulation distinctions. A Style × Sibilance interaction [*F*(1, 27)=10.331, *p*<.01] showed that while both sibilants and non-sibilants were more intelligible in clear speech, the effect was larger for sibilant pairs. The Style × Distinction interaction was not significant [*F*<1]; clear speech resulted in similar benefits for place and voicing distinctions.

2.2.2 Talker-related acoustic-phonetic correlates of clear intelligibility advantage

On average, individual talkers appeared in 336 (std. 18.8) clear and 336 (20.01) conversational trials. Partly due to the adaptive procedure and the initial threshold guesses of +/- 3dB across styles and pairs, talkers appeared at -4.91 dB (std. 0.28) and -2.61 dB (0.22) SNR values, and were responded to with 81.9% (std. 4.58) and 77.3% (5.0) accuracy in clear and conversational conditions, respectively. Averaged across listeners, contrasts, and SNR values, the clear-minus-conversational difference in accuracy (% correct) varied considerably across speakers, from -4% to +11% (mean 4.6%, std. 3.9%), at least partly as a result of differences in the clear speech strategies that these talkers employed (i.e., this difference did not correlate well [*p*=0.34] with clear-minus-conversational SNR differences). As described above, then, individual speakers' previously reported average style-related differences in production were compared with their style-related intelligibility differences in an effort to relate clear speech benefits to specific acoustic modifications. Results from Pearson's correlations are summarized in Table 3-1. Overall, spectral measures at central window locations (window (W) 2, 3 and 4) appear to be the important predictors for improved intelligibility. Positive significant (or marginally significant) correlations were obtained between intelligibility advantages and acoustic modifications in FFT peak location at W2 and 3, LPC peak location at W2-4 , Spectral moment 1 (M1) at W2-4, and slope before peak location at W1-4; negative correlations were found for M3 at W2 and W5, and spectral slope after peak location at W1. These results suggest that a shift of spectral energy to higher frequency regions and greater source strength (Jesus and Shadle, 2002) in clear fricatives are most closely related to the overall intelligibility enhancement.

Table 3-1: Correlation coefficients (Pearson's *r*) showing the relation between the clear-minus-conversational differences in acoustic measures and the clear-minus-conversational differences in the intelligibility (percent identification correctness) for the overall effect (General), place-of-articulation distinction (Place), and voicing distinction (Voicing) in Exp.1. Significant values, $p<0.001$, $p<0.01$ and $p<0.05$ are starred as ***, **, and *, respectively. Moderate values, $p<0.1$ are marked as ., and no effect was given N. Negative correlation was marked as ξ.

Measures	General	p values	r values	Place	p values	r values	Voicing	p values	r values
Durs	N	0.19	0.3	N	0.66	0.11	N	0.29	0.25
F2W 1	Nξ	0.29	-0.25	Nξ	0.37	-0.21	Nξ	0.63	-0.12
F2W 2	N	0.15	0.34	*	0.04	0.45	N	0.77	0.07
F2W 3	N	0.49	0.16	N	0.66	0.1	N	0.52	0.15
F2W 4	N	0.31	0.24	Nξ	0.74	-0.08	N	0.14	0.34
FFTPksW 1	N	0.3	0.24	N	0.32	0.23	N	0.62	0.12
FFTPksW 2	*	0.02	0.53	*	0.01	0.55	N	0.37	0.21
FFTPksW 3	*	0.04	0.47	**	0.01	0.56	N	0.65	0.11
FFTPksW 4	N	0.1	0.38	**	0	0.64	Nξ	0.8	-0.06
FFTPksW 5	Nξ	0.62	-0.12	Nξ	0.11	-0.37	N	0.49	0.16
HNR	Nξ	0.86	-0.04	Nξ	0.42	-0.19	N	0.75	0.08
Int500W 1	N	0.95	0.01	Nξ	0.25	0.27	N	0.36	0.22
Int500W 2	Nξ	0.17	-0.32	Nξ	0.11	-0.37	Nξ	0.77	-0.07
Int500W 3	Nξ	0.17	-0.32	.ξ	0.07	-0.42	Nξ	0.92	-0.02
Int500W 4	Nξ	0.13	-0.35	.ξ	0.09	-0.39	Nξ	0.72	-0.09
Int500W 5	N	0.28	0.26	N	0.18	0.32	N	0.79	0.06
M1W 1	N	0.17	0.32	N	0.1	0.38	N	0.67	0.1
M1W 2	*	0.02	0.51	.	0.07	0.41	N	0.22	0.29
M1W 3	.	0.06	0.43	.	0.09	0.39	N	0.38	0.21
M1W 4	.	0.08	0.41	*	0.04	0.46	N	0.6	0.12
M1W 5	N	0.46	0.17	Nξ	0.92	-0.03	N	0.3	0.24
M2W 1	N	0.12	0.36	*	0.02	0.52	N	0.86	0.04
M2W 2	N	0.3	0.25	N	0.28	0.26	N	0.66	0.1
M2W 3	N	0.4	0.2	N	0.49	0.16	N	0.6	0.12
M2W 4	N	0.55	0.14	N	0.39	0.2	N	0.93	0.02
M2W 5	N	0.42	0.19	Nξ	0.63	-0.12	N	0.15	0.34
M3W 1	Nξ	0.11	-0.36	**ξ	0.01	-0.56	Nξ	0.97	-0.01
M3W 2	*ξ	0.04	-0.46	*ξ	0.01	-0.56	Nξ	0.65	-0.11
M3W 3	Nξ	0.12	-0.36	*ξ	0.03	-0.48	Nξ	0.85	-0.04
M3W 4	Nξ	0.23	-0.28	*ξ	0.01	-0.54	N	0.69	0.09
M3W 5	.ξ	0.1	-0.38	Nξ	0.28	-0.26	Nξ	0.27	-0.26
M4W 1	Nξ	0.19	-0.31	**ξ	0.01	-0.6	Nξ	0.72	-0.09
M4W 2	Nξ	0.19	-0.31	*ξ	0.01	-0.56	Nξ	0.77	-0.07
M4W 3	Nξ	0.32	-0.24	*ξ	0.04	-0.47	Nξ	0.7	-0.09
M4W 4	Nξ	0.48	-0.17	*ξ	0.03	-0.49	Nξ	0.43	-0.19
M4W 5	.ξ	0.09	-0.39	Nξ	0.24	-0.28	Nξ	0.29	-0.25
follPitch	N	0.32	0.24	**	0	0.66	Nξ	0.33	-0.23
prevPitch	Nξ	0.52	-0.15	Nξ	0.68	-0.1	Nξ	0.58	-0.13
RelAmps	N	0.88	0.04	Nξ	0.92	-0.02	N	0.68	0.1
RMSAmpsW 1	N	1	0	Nξ	0.14	-0.34	N	0.23	0.28
RMSAmpsW 2	Nξ	0.61	-0.12	Nξ	0.15	-0.33	N	0.52	0.15
RMSAmpsW 3	Nξ	0.98	-0.06	Nξ	0.29	-0.25	N	0.32	0.23
RMSAmpsW 4	Nξ	0.83	-0.05	Nξ	0.31	-0.24	N	0.47	0.17
RMSAmpsW 5	N	0.14	0.34	N	0.15	0.34	N	0.5	0.16
SlpAftW 1	.ξ	0.06	-0.42	Nξ	0.62	-0.12	.ξ	0.08	-0.4
SlpAftW 2	Nξ	0.22	-0.29	Nξ	0.12	-0.36	Nξ	0.88	-0.04
SlpAftW 3	Nξ	0.49	-0.16	Nξ	0.15	-0.33	N	0.7	0.09
SlpAftW 4	Nξ	0.94	-0.02	Nξ	0.31	-0.24	N	0.4	0.2
SlpAftW 5	Nξ	0.54	-0.15	Nξ	0.22	-0.29	N	0.72	0.09
SlpBefW 1	.	0.06	0.42	*	0.01	0.54	N	0.66	0.11
SlpBefW 2	.	0.09	0.39	*	0.01	0.54	N	0.86	0.04
SlpBefW 3	.	0.09	0.39	**	0	0.62	Nξ	0.96	-0.01
SlpBefW 4	*	0.05	0.45	**	0	0.62	N	0.81	0.06
SlpBefW 5	N	0.21	0.29	*	0.05	0.45	Nξ	0.98	-0.01
LPCPeakW 1	N	0.2	0.3	N	0.14	0.35	N	0.64	0.11
LPCPeakW 2	**	0.01	0.57	**	0	0.65	N	0.49	0.16
LPCPeakW 3	*	0.03	0.49	*	0.01	0.54	N	0.52	0.15
LPCPeakW 4	*	0.05	0.44	**	0	0.63	N	0.92	0.03
LPCPeakW 5	Nξ	0.93	-0.02	Nξ	0.47	-0.17	N	0.59	0.13

Since perception of place and voicing distinctions probably involve different acoustic cues, this analysis was repeated separately for the four place distinction subtests and the four voicing subtests in the experiment. Comparison of these analyses suggested that most of the effects mentioned above were due to place of articulation distinctions. As shown in Table I, considering only place distinctions, strong positive correlations between the intelligibility advantage of clear speech and clear-minus-conversational difference in acoustic measures were found for FFT peaks, LPC peaks, and M1 at central window locations, F2 at W2, and slope before the peak at all window locations whereas negative correlations were seen for M3 at W1-4, M4 at W1-4, and intensity below 500 Hz at W3 and 4. These results clearly suggest that shifts toward higher frequency regions, and greater source strength are likely to contribute to the better recognition of place of articulation for fricatives. In contrast, no strong correlations were obtained between any acoustic measures and intelligibility benefits for voicing distinctions.

An independent samples *t*-test showed no main effect of talker gender (*t*=-.720, *p*=.490); female and male talkers did not differ in the amount of intelligibility improvement in clear speech relative to conversational speech (*cf.* Bradlow *et al.*, 2003).

2.2.3 Discussion

Results show that clearly produced fricatives are more intelligible than casually produced fricatives for listeners with normal hearing in degraded listening conditions. In accordance with previous findings (e.g. Jongman, *et al.*, 2000b; Miller and Nicely, 1955; Wang and Bilger, 1973), sibilant pairs and voicing distinction pairs were easier to identify relative to non-sibilant and place of articulation pairs, respectively, regardless of speaking style. Clear speech effects were seen qualitatively for all minimal pairs tested, and reached significance for all pairs except /v/-/δ/ and /f/-/v/.

The 20 talkers included in the experiment varied considerably in the averaged clear-minus-conversational difference in the accuracy (% correct) to which they were responded. Correlation analyses suggested that in general, spectral peak location, mean frequency and spectral slope before the peak are positively correlated with improved intelligibility in clear speech while skewness is negatively correlated to intelligibility enhancement. For place of articulation distinctions, F2 and kurtosis also appear to influence clear speech intelligibility advantage. Overall, a shift of energy concentration toward higher frequency regions and greater energy source seem to contribute to the clear speech effect in fricatives, especially for place of articulation distinctions.

3. Experiment 3: Effects of clear speech for fricative recognition by listeners with simulated hearing impairment

3.1 Simulation method

3.1.1 Rationale

Experiment 2 results suggest that a clear speaking style can improve fricative intelligibility for listeners with normal hearing in background noise. It was also found that intelligibility advantages for place-of-articulation distinctions seemed to relate to spectral changes in clear speech; higher peak locations, higher mean frequency, lower skewness, and steeper spectral slopes before peak locations contributed to the higher correct identification scores in clear speech relative to conversational speech. Given these apparent relationships, it is important to ask whether the clear fricative advantages would hold for impaired listeners, particularly those who have poor hearing at higher frequencies. Listeners with sloping hearing losses have considerable difficulty recognizing sounds that have important acoustic information in higher frequency regions, such as fricatives (Dubno *et al.*, 1982; Owens *et al.*, 1972; Sher and Owens, 1974). These difficulties may be at least partially derived from suprathreshold abnormalities in the perceptual analysis of the speech signal, including reduced dynamic range (abnormal loudness recruitment), reduced frequency selectivity, and impaired temporal resolution. It has proven difficult to determine which changes in auditory abilities in individuals with hearing impairment are most important and relevant for altered perception of speech, since the elevation of absolute thresholds is usually intercorrelated with a variety of suprathreshold changes and thus the effects of these are difficult to separate from each other. A common strategy to control the confounding factors is to process sounds so as to simulate the effects of one specific aspect of hearing impairment, and to allow listeners with normal hearing to experience selected perceptual effects of hearing impairments. In this experiment, we were particularly interested in the influence of threshold elevation on recognition of fricative sounds, since important fricative information occurs at frequencies where many impaired listeners have elevated thresholds, and since Experiment 2 suggests that this may be *increasingly* so for clear fricatives. It is possible that listeners with sloping hearing loss cannot make use of the enhanced acoustic-phonetic information since it is less audible to them. To assess how this specific aspect of hearing impairment would affect the perception of clear fricative sounds, we repeated the perception experiment using stimuli processed to simulate sloping, recruiting hearing loss.

3.1.2 Implementation

Sloping, recruiting hearing loss was simulated in a manner similar to that described by Moore and Glasberg (1993), with some modifications due to a higher sampling rate (44.1 kHz) and the fact that all processing was done on-line during the experiment. Following the combination of signal and noise components (described below), stimuli were separated into 24 ERB-spaced bands, from 100 Hz to 22.05 kHz, using 4th order gammatone filters (Slaney, 1998). For each band, a smoothed envelope (E) was derived by low-pass filtering the full-wave rectified waveform at 100 Hz (4th order Butterworth filter, implemented in both forward and reverse directions to minimize phase distortions). The temporal fine structure for the band was then extracted by dividing the original waveform by this envelope. Loss simulation was accomplished by raising the envelope to a power related to the slope of the loudness growth function:

$$E_p = E^N$$

where N is frequency-dependent. Following Moore and Glasberg (1993), N was a constant 1.5 at bands up to 900 Hz, increased linearly to 3.0 at 4500 Hz, and remained at this value for all higher bands. Finally, the modified stimulus was obtained by multiplying E_p by the fine structure and summing the resulting band-limited waveforms. All processing was performed in Matlab. Processing on average took ~2s using the Intel® Pentium 4 processor used in testing; this resulted in an inter-trial interval that a few participants found slightly annoying but generally not distracting.

3.2 Experiment method

3.2.1 Participants

Fourteen normal-hearing listeners (9F, 5M) aged between 19 and 33 were recruited from the University of California, Berkeley community. Participants were native speakers of American English, without noticeable regional dialects. Participants reported normal hearing and no history of speech or language disorders. Listeners were paid for their participation in the experiment.

3.2.2 Materials

Test stimuli were identical to those of Experiment 2 except that (1) speech/ babble stimuli were processed as described above, and that (2) only the four place-of-articulation pairs /f/-/θ/, /v/-/ð/, /s/-/ʃ/, and /z/-/ʒ/ were tested. The same twelve-talker babble masker used in Experiment 2 was added to the speech stimuli prior to the simulation processing.

3.2.3 Procedures and apparatus

The procedure, task, presentation method, and adaptive procedure were identical to those of Experiment 2, except that since only four pairs were tested there was no mandatory break after the 8th sub-test. Testing took about 50 minutes.

3.2.4 Data analysis

As in Experiment 2, a repeated measure analysis of variance (ANOVA) with two within-subject factors (Style; 2 levels, Pair; 4 levels) and thresholds (dB SNR) as dependent variable was again performed. Pairwise comparisons for significant within-subject factors were done using Bonferroni corrected 95% confidence intervals.

3.3 Results and discussion

3.3.1 Fricative intelligibility for listeners with simulated hearing loss

Figure 3-2 shows average SNR thresholds as a function of pair type for clear and conversational fricative identification. For all place pairs except /f/-/θ/, clear speech showed lower SNR thresholds relative to conversational speech. The Style × Pair ANOVA showed an effect of Style [*F*(1,13)=13.892, *p*<.01] with 2.52 dB lower thresholds for clear speech . There was also a Pair effect [*F*(3,39)=149.551, *p*<.001], mostly derived from lower thresholds for sibilant pairs relative to non-sibilant pairs. The Style × Pair interaction was also significant [*F*(3,39)=5.989, *p*<.01]. Pairwise comparisons showed significant differences in thresholds as a function of style for /s/-/ʃ/ and /z/-/ʒ/ pairs, but not for non-sibilant pairs. In fact, for /f/-/θ/, clear speech resulted in higher (n.s.) thresholds compared to conversational speech. These results differed from Experiment 2 results in that (1) thresholds were on average much higher in simulated loss conditions, (2) there was an anti-clear speech effect for /f/-/θ/, while in Experiment 2 thresholds significantly decreased in clear speech for this pair, and (3) the /z/-/ʒ/ pair showed the biggest clear speech effect, followed by /s/-/ʃ/, /v/-/ð/,and /f/-/θ/, while in Experiment 2 the order was /z/-/ʒ/, /f/-/θ/, /s/-/ʃ/ and /v/-/ð/. On the other hand, the relative difficulty of fricative pairs was similar to Exp. 2; across speaking styles, the pair /s/-/ʃ/ resulted in the lowest thresholds, followed by /z/-/ʒ/, /v/-/ð/, and /f/-/θ/.

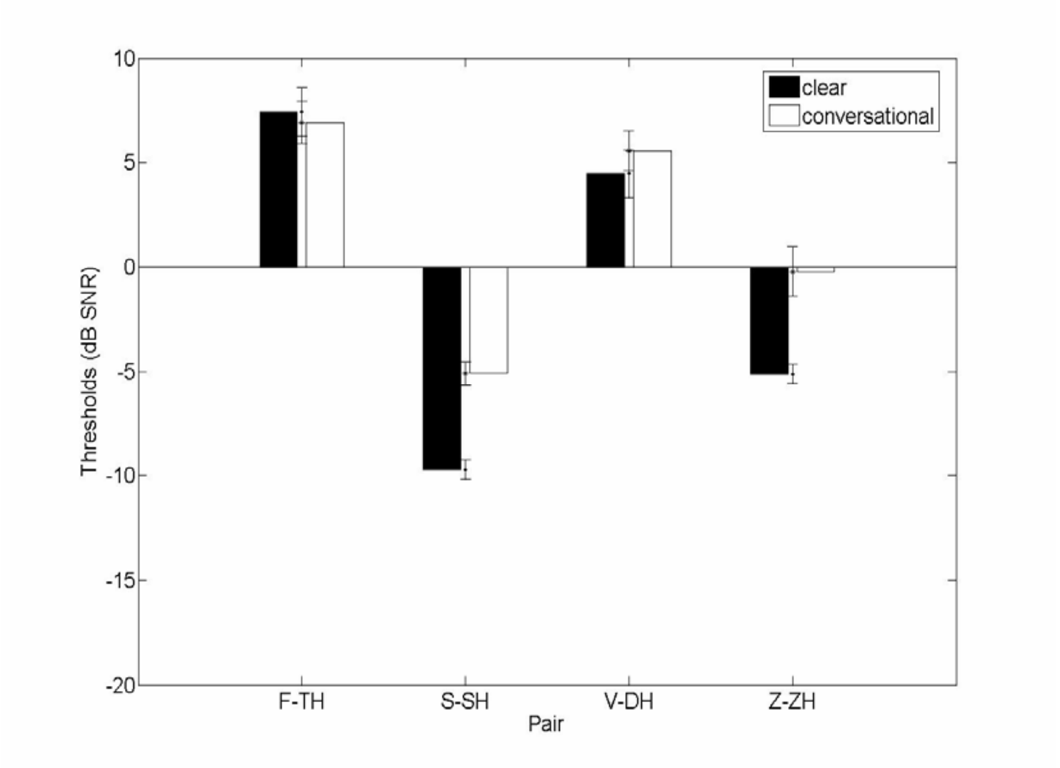


Figure 3-2: Speech-to-noise (SNR) thresholds (dB) as a function of style and fricative pair in Experiment 3

To determine how the loss simulation influenced the perception of fricatives in interaction with speaking style and contrastive pair, a three-way mixed model repeated measures of ANOVA was performed with two within-subject factors (Style, Pair) and listener group as a between-subject factor (2 levels; Exp. 2 and Exp. 3). Since the 4 voicing distinction pairs were not included in Experiment 3, only the 4 place-of-articulation distinction pairs from Experiment 2 were considered. This analysis showed a main effect of Group [*F*(1,26)=26.392, *p*<.001] with considerably (4.47 dB) higher thresholds for listeners with simulated hearing . A main effect of Style [*F*(1,26)=48.958, *p*<.001] indicated, again, an overall clear speech advantage across listener groups. There was no Style × Group interaction, suggesting that, on average, listeners with normal hearing and listeners with simulated impairment significantly benefited similarly from clear speech. The main effect of Pair was significant [$F(3,78)=212.756$, $p<.001$] but not the Pair × Group interaction, reflecting the common difficulty hierarchy mentioned above. Again, pairwise comparisons indicated that all 4 pairs were significantly different from each other, and that the effect was most notably derived from differences between sibilant and non-sibilant pairs. A Style × Pair interaction [$F(3,78)=212.756$, $p<.01$] indicated that, across listener groups, thresholds significantly decreased in clear speech for all fricative pairs except for /f/-/θ/. The Style × Pair × Group interaction was significant [$F(3,78)=2.9000$, $p<.05$]; *post-hoc* tests suggested that the interaction was related to an increase in the magnitude of the clear effect for sibilants, and a *decrease* for non-sibilants, in the simulated impairment condition. This finding is illustrated in figure 3-3, which shows the clear speech effect as a function of pair and listening condition.

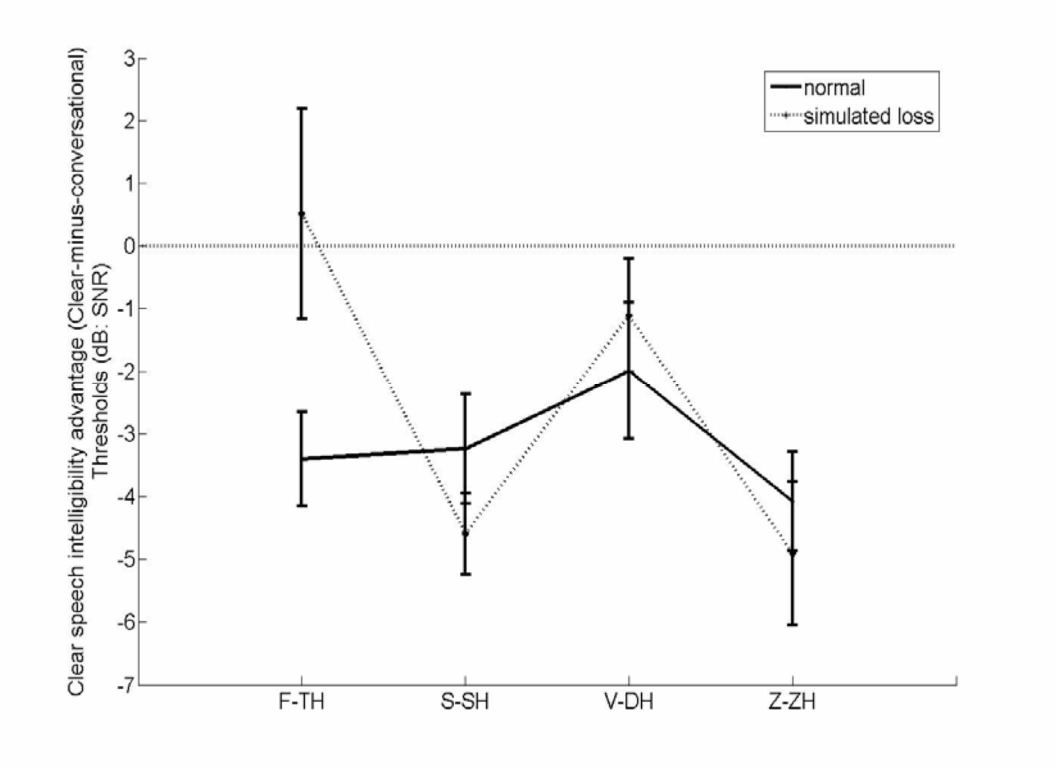


Figure 3-3: Clear speech intelligibility advantage (clear-minus-conversational thresholds) in dB SNR by listeners with normal hearing and listeners with simulated hearing impairment as a function of fricative pair

3.3.2 Acoustic correlates of intelligibility benefit for listeners with simulated hearing impairment

In Experiment 3, individual talkers appeared on average in 168 (std. 12.4) clear and 168 (11.63) conversational trials. Again, averaged across listeners, contrasts, and SNR values, the clear-minus-conversational difference in accuracy (% correct) varied considerably across speakers, from -6% to +18% (mean 3.9%, std. 6.6%), at least partly as a result of differences in the clear speech strategies that these talkers employed. As discussed in Experiment 2, individual speakers' previously reported average style-related differences in production were compared with their style-related intelligibility differences in a first effort to relate clear speech benefits to specific acoustic modifications.

The results of this comparison are shown in Table 3-2. Overall, correlations were much less consistent than for Experiment 2; in particular, conspicuously absent are the positive correlations with several spectral measures indicating shifts to higher frequency regions that were seen for place contrasts in Exp. 2. Since the perception of sibilant and non-sibilant pairs seemed to be affected differentially by the impairment, a final set of correlation analyses compared intelligibility differences across speakers with acoustic differences separately for each class of sounds. While this comparison was considerably less well powered than the others described above, the results were potentially interesting and are also included in Table II. For sibilant pairs, positive correlations were seen between intelligibility advantages and intensity below 500 Hz at W5 (p<.1), M3 at W4 (p<.1), M4 at W4 (p<.05), and RMS amplitude at W5 (p<.1), and *negative* correlations for FFT peaks at W3 and 4 (p<.1), LPC peaks at W3 and 4 (p<.05), M1 at W5 (p<.1), M2 at W4 (p<.05) and W5 (p<.1). For non-sibilant pairs, correlations were weaker and less straightforward, with only one significant positive correlate for f0 following the fricative (p<.05) and marginal correlate for M1 at W3 (p<.1), and negative correlates for F2 at W3 (p<.1), intensity below 500 Hz at W5 (p<.1), RMS amplitude at W3 (p<.1), slope after the peak locations at W5 (p<.1), and LPC peak at W4 (p<.1).

Table 3-2: Correlation coefficients (Pearson's *r*) showing the relation between the clear-minus-conversational differences in acoustic measures and the clear-minus-conversational differences in the intelligibility (percent identification correctness) for the overall effect (General), sibilant pair distinction (Sib), and non-sibilant pair distinction (NonSib). Significant values, $p<0.001$, $p<0.01$ and $p<0.05$ are starred as ***, **, and *, respectively. Moderate values, $p<0.1$ are marked as ., and no effect was given N. Negative correlation was marked as ξ

Measures	General	p values	r values	Sib	p values	r values	NopSib	p values	r values
Durs	Nξ	0.69	-0.09	Nξ	0.92	-0.02	Nξ	0.56	-0.14
F2W 1	N	0.16	0.32	N	0.29	0.25	N	0.56	0.14
F2W 2	Nξ	0.24	-0.28	Nξ	0.39	-0.2	Nξ	0.43	-0.19
F2W 3	Nξ	0.15	-0.33	Nξ	0.55	-0.14	.ξ	0.06	-0.43
F2W 4	N	0.82	0.05	Nξ	0.66	-0.1	N	0.67	0.1
FFTPksW 1	N	0.44	0.18	Nξ	0.64	-0.11	N	0.29	0.25
FFTPksW 2	N	0.69	0.1	Nξ	0.18	-0.31	N	0.27	0.26
FFTPksW 3	Nξ	0.9	-0.03	.ξ	0.08	-0.39	N	0.22	0.29
FFTPksW 4	Nξ	0.92	-0.03	.ξ	0.06	-0.43	N	0.13	0.35
FFTPksW 5	Nξ	0.73	-0.08	Nξ	0.46	-0.17	Nξ	0.79	-0.06
HNR	Nξ	0.37	-0.21	N	0.56	0.14	Nξ	0.2	-0.3
Int500W 1	Nξ	0.93	-0.02	N	0.61	0.12	Nξ	0.55	-0.14
Int500W 2	N	0.96	-0.01	Nξ	0.78	-0.07	Nξ	0.74	-0.08
Int500W 3	Nξ	0.49	-0.16	Nξ	0.69	-0.09	Nξ	0.2	-0.3
Int500W 4	Nξ	0.5	-0.16	Nξ	0.61	-0.12	Nξ	0.28	-0.25
Int500W 5	.	0.06	0.43	.	0.07	0.41	.ξ	0.05	0.44
M1W 1	N	0.51	0.16	Nξ	0.6	-0.13	N	0.31	0.24
M1W 2	N	0.48	0.17	N	0.77	0.07	N	0.25	0.27
M1W 3	N	0.34	0.22	Nξ	0.84	-0.05	.	0.09	0.39
M1W 4	N	0.56	0.14	Nξ	0.58	-0.13	N	0.15	0.33
M1W 5	Nξ	0.96	-0.01	.ξ	0.05	-0.44	Nξ	0.94	-0.02
M2W 1	N	0.94	0.02	Nξ	0.61	-0.12	N	0.75	0.08
M2W 2	N	0.81	0.06	Nξ	0.15	-0.34	N	0.49	-0.16
M2W 3	Nξ	0.81	-0.06	Nξ	0.14	-0.34	N	0.71	0.09
M2W 4	Nξ	0.33	-0.23	*ξ	0.02	-0.51	N	0.96	0.01
M2W 5	Nξ	0.42	-0.19	.ξ	0.06	-0.43	Nξ	0.29	-0.25
M3W 1	Nξ	0.52	-0.15	Nξ	0.96	-0.01	Nξ	0.51	-0.16
M3W 2	Nξ	0.62	-0.12	N	0.68	0.1	Nξ	0.55	-0.14
M3W 3	Nξ	0.7	-0.09	N	0.14	0.34	Nξ	0.39	-0.21
M3W 4	N	0.86	0.04	.	0.05	0.44	Nξ	0.56	-0.14
M3W 5	Nξ	0.54	-0.14	N	0.32	0.24	Nξ	0.74	-0.08
M4W 1	Nξ	0.65	-0.11	Nξ	0.83	-0.05	Nξ	0.8	-0.06
M4W 2	Nξ	0.98	-0.01	N	0.3	0.24	N	0.94	0.02
M4W 3	N	0.72	0.08	N	0.15	0.33	N	0.94	0.02
M4W 4	N	0.38	0.21	*	0.01	0.55	N	0.78	0.07
M4W 5	Nξ	0.65	-0.11	N	0.26	0.26	Nξ	0.93	-0.02
follPitch	.	0.09	0.39	N	0.53	0.15	*	0.03	0.49
prevPitch	Nξ	0.65	-0.11	Nξ	0.59	-0.13	N	0.44	0.18
RelAmps	Nξ	0.71	-0.09	Nξ	0.9	-0.03	N	0.65	0.11
RMSAmpsW 1	Nξ	0.46	-0.18	N	0.69	0.1	Nξ	0.14	-0.34
RMSAmpsW 2	Nξ	0.6	-0.13	N	0.58	0.13	Nξ	0.24	-0.27
RMSAmpsW 3	Nξ	0.37	-0.21	N	0.55	0.14	.ξ	0.07	-0.41
RMSAmpsW 4	Nξ	0.65	-0.11	N	0.25	0.27	Nξ	0.17	-0.32
RMSAmpsW 5	N	0.15	0.33	.	0.07	0.42	N	0.27	0.26
SlpAftW 1	Nξ	0.37	-0.21	Nξ	0.89	-0.03	Nξ	0.29	-0.25
SlpAftW 2	Nξ	0.63	-0.12	Nξ	0.69	-0.09	Nξ	0.51	-0.16
SlpAftW 3	Nξ	0.98	-0.01	N	0.36	0.22	Nξ	0.6	-0.13
SlpAftW 4	Nξ	0.67	-0.1	N	0.25	0.27	Nξ	0.39	-0.2
SlpAftW 5	Nξ	0.02	-0.27	Nξ	0.66	-0.11	.ξ	0.06	-0.42
SlpBefW 1	Nξ	0.7	0.09	Nξ	0.83	0.05	Nξ	0.79	-0.06
SlpBefW 2	Nξ	0.99	0	N	0.48	0.17	Nξ	0.61	-0.12
SlpBefW 3	N	0.98	0.01	N	0.24	0.28	Nξ	0.3	-0.24
SlpBefW 4	N	0.94	0.02	N	0.34	0.23	Nξ	0.52	-0.15
SlpBefW 5	N	0.62	0.12	N	0.14	0.34	Nξ	0.56	-0.14
LPCPeakW 1	N	0.52	0.15	N	0.98	0.01	N	0.47	0.17
LPCPeakW 2	N	0.51	0.16	Nξ	0.14	-0.34	N	0.18	0.31
LPCPeakW 3	N	0.7	0.09	*ξ	0.04	-0.45	N	0.11	0.37
LPCPeakW 4	N	0.8	0.02	*ξ	0.01	-0.46	.ξ	0.07	0.41
LPCPeakW 5	Nξ	0.45	-0.18	Nξ	0.44	-0.18	N	0.83	0.05

3.3.3 Discussion

On average, clear speech may benefit listeners with hearing impairment, as evidenced by the overall Style effect and lack of a Style × Group interaction seen here. The clear speech effect did not reach significance for non-sibilant pairs, and was actually in a negative direction for the voiceless non-sibilant pair; on the other hand, sibilant sounds showed robust clear speech advantages. It is not clear what caused this pattern (and the Style × Group × Pair interaction indicating that Experiment 3 listeners benefited less from clear speech for non-sibilants), but it seems likely that, since both voiceless non-sibilants are characterized by the highest peak values and F2 with diffuse spread of energy below 10 kHz, important spectral cues are less audible/available to listeners with sloping hearing loss the higher they are transposed. Sibilants, on the other hand, had both higher inherent consonant-to-vowel ratios (CVRs) and more potential cues (esp. palato-alveolar peak frequencies) involving energy in lower regions. These cues were generally better-preserved in stimuli with simulated sloping, recruiting losses, perhaps especially when clear speech involved more energy at lower frequencies.

Talkers varied considerably in estimated clear speech intelligibility benefit, although it was harder to relate this variability to overall differences in production than in the absence of loss simulation (Exp. 2). Again, there was no Style × Gender interaction (t=1.577, p=.149), indicating that female and male talkers did not differ in their clear speech intelligibility effect.

4. Discussion

4.1 Overall clear fricative intelligibility

Across two experiments, lower SNR identification thresholds for place of articulation identification were seen for clear relative to conversational fricatives, indicating that, on average, clearly produced fricatives are more intelligible for both young normal-hearing listeners and listeners with simulated sloping, recruiting hearing impairment. In addition, clear speech was beneficial to normal-hearing listeners in terms of the voicing distinction. However, these effects were not as uniform and robust across fricatives and listener groups as might have been expected. In Experiment 2, sibilant fricatives were easier to identify than non-sibilants for normal-hearing listeners overall, and clear speech provided slightly greater intelligibility benefits for sibilants than non-sibilants. Experiment 3 showed that these trends were exaggerated for simulated hearing-impaired listeners. In particular, a clear speech effect was seen *only* for sibilants, and clear speech may have even hurt intelligibility for voiceless non-sibilants, the worst-recognized sounds. These results are consistent with the notion (e.g. Ferguson and Kewly-Port, 2002) that the perceptual effects of clear speech acoustic modifications may be population-dependent, and may interact in complex ways with different types of hearing impairment. As discussed below, they probably derive from differences in the audibility and weighting of acoustic cues across fricatives and listening conditions.

4.2 Acoustic and talker-related correlates of clear speech intelligibility effect

Comparison of individual speakers' estimated clear-speech intelligibility advantages with their previously reported (Maniwa *et al*., submitted) clear-speech acoustic modifications revealed correlations that may be informative as to the acoustic sources of the "clear speech effect" in fricatives. Specifically, for place-of-articulation distinctions, strong positive correlations were found between acoustic and perceptual clear-vs.-conversational differences for spectral measures, especially at central locations, including peak locations, M1, and spectral slope before peak locations. In addition, there were negative correlations between intelligibility improvement and increases in M3 and M4 at all locations, and for intensity below 500 Hz at central locations. These results indicate that, overall, a shift of important spectral information toward higher frequency regions and higher source strength (produced by higher volume velocity) in clear speech contributed to the intelligibility enhancement for place distinctions. Of course, it is more likely that these "global" changes in conjunction with higher-order patterns specific to individual fricatives and contexts actually led to the intelligibility effects that were seen. The experiments described here could not address this possibility, since within individual subtests SNR values were not sufficiently equalized across speakers to make more specific comparisons. Probably for this reason, no strong correlations were seen relating acoustic measures and voicing intelligibility. In particular, acoustic results suggested that phonetic distance in terms of the voicing distinction was often "enhanced" in clear speech by increasing (or decreasing to a lesser degree) values for one class of fricatives while decreasing (or increasing to a lesser degree) values for the class. For example, intensity below 500 Hz decreased much less, and HNR significantly increased, for voiced fricatives whereas these values significantly decreased for voiceless fricatives. Similarly, noise duration and f0 increased for both voiceless and voiced fricatives in clear speech, but to a much greater extent for voiceless fricatives. These differences in clear speech manipulations, and their perceptual effects, would have mostly been obscured by the analysis described here.

Previous study (Maniwa *et al.*, submitted) indicates that voiceless non-sibilants have, in addition to very low amplitudes, very high peak frequencies (higher than /s/), mean frequency, and F2, across speaking styles, and that these values are even higher in clear speech. This is very likely a cause of the lack of clear speech benefits for (especially voiceless) non-sibilants, since the simulated impairment targeted higher frequencies (and low amplitudes). Sibilants, on the other hand, were characterized by more and lower energy, in some cases (esp. palato-alveolars) even more so in clear speech, so more potential cues for these sounds were preserved in the loss simulation. As a result of these differences, for listeners with simulated hearing impairment few overall correlations between acoustic and intelligibility differences in clear speech were apparent in Experiment 3. For identification of sibilant pairs specifically, contrary to Experiment 2 results, there were some *negative* correlations between acoustic changes in peak frequencies and enhanced intelligibility (and marginal *positive* correlation between M3 and intelligibility advantage). This suggests that the lower the spectral information moved for palato-alveolar fricatives in clear speech, the more intelligible these sounds were, because this information was better preserved in the impairment simulation. Fewer and less consistent patterns could be seen to relate non-sibilant acoustic modifications to intelligibility. In other words, elevated thresholds and loudness recruitment influenced listeners’ cue weighting for the perception of fricative sounds.

There were no Style × Gender interactions in either experiment, indicating that female and male talkers did not differ in terms of the effectiveness of their clear speech acoustic modifications for intelligibility (*cf.* Bradlow *et al.*, 2003).

4.3 Conclusion

This study showed that clear speech enhanced the intelligibility of fricatives for both listeners with normal hearing and listeners with simulated hearing impairment. However, the effect was fricative- and population-dependent; notably, compared to normal-hearing listeners, impaired listeners showed reduced clear speech effects for non-sibilant place of articulation distinctions. Likewise, apparent acoustic correlates of the clear speech benefit differed across populations. For normal-hearing listeners, intelligibility benefits seemed to correlate with moves toward higher frequency regions for important cues; these patterns were generally not seen for impaired listeners, and may even have been reversed for some sounds. These results are straightforwardly explained based on audibility of cues at different levels and frequencies. We leave for future study a more thorough investigation of potential higher-order acoustic correlates of the clear speech effect in fricatives; this could be accomplished straightforwardly by using the results of the adaptive design described here to inform blocked-design experiments that are optimally controlled (and powered) for the distribution of fricatives, styles, and SNR values across speakers and tokens.

Chapter 4

General discussion/conclusions

4.1 Summary of findings

The goal of this research was to extend our understanding of the acoustic and perceptual characteristics of speech that is produced when speakers attempt to enhance their intelligibility under adverse communicative conditions. Three experiments described here were designed to address this objective, focusing specifically on English fricatives. The first experiment aimed primarily at specifying reliable fine-grained acoustic-phonetic changes in fricative production related to speaking style. A secondary issue in this experiment involved measuring talkers' efforts at maintaining contrasts between similar fricative sounds in response to specific recognition errors by their communication partner(s). The second experiment was designed to determine whether talkers' strategies to alter the phonetic properties of fricatives are effective at enhancing the intelligibility of these sounds under noisy conditions; the goalof the third experiment was to examine how robust the "clear speech effect" would be for listeners with sloping, recruiting hearing losses, who have less audibility in higher frequency regions. A secondary goal of Experiments 2 and 3 was to get a preliminary indication of the relative contributions of different clear-versus-conversational acoustic changes in determining the clear speech intelligibility advantage for these two listener groups.

Acoustic results indicated that there are indeed systematic acoustic-phonetic modifications in the production of clear fricatives. Some overall clear speech effects were straightforwardly predictable based on previous findings (e.g. longer duration, energy at higher frequencies), and some were more surprising (esp. lower relative amplitude). For several important measures, the acoustic distance between minimally contrasting sounds was enlarged in clear speech, indicating that talkers attempt to maintain contrast between category distributions across the inventory of English fricatives. In addition, talkers were sensitive to more local listener feedback, adjusting repeated productions to be more unlike the specific sounds that they had just been misapprehended for. Individual talkers varied widely in the magnitude and sometimes direction of these changes; these differences were not related to talker gender. Perception experiments suggested that clear speech enhanced the overall intelligibility of fricatives for both listeners with normal hearing and listeners with simulated hearing impairment. However, the effect was fricative- and population-dependent. Compared to normal-hearing listeners, impaired listeners showed greatly reduced clear speech effects for non-sibilant place of articulation distinctions. Likewise, apparent acoustic correlates of the clear speech benefit differed across populations. For normal-hearing listeners, intelligibility benefits seemed to correlate with moves toward higher frequency regions for important cues; these patterns were generally not seen for impaired listeners, and may even have been reversed for some sounds. These results are straightforwardly explained based on audibility of cues at different levels and frequencies. Advantages that might have been derived from the transposition of acoustic information to higher frequency regions were confounded by increasing loss at these frequencies.

4.2 Theoretical implications

Experiment 1 results suggest that speakers tend to maintain important acoustic contrasts within the fricative inventory in attempting to make their productions more intelligible after receiving error feedback and recognizing a communicative difficulty. Speakers were even sensitive to trial-by-trial differences in the relevance of specific contrasts, adapting their speech patterns in response to specific perception errors provided on-line by the interactive program. These results lend considerable support and empirical evidence for H & H Theory (e.g., Lindblom, 1990) which argues that language users consistently assess listener's needs for explicit signal information and modify their speaking style by maximizing contrasts according to the demands of online communicative situations. The effects observed in Experiment 1 are important in demonstrating the range of levels at which talkers are sensitive to the communicative demands of a speaking situation; to the author's knowledge, no previous study has addressed the influence of listener demands on the production of a class of sounds in such a comprehensive way.

4.3 Practical implications

While this study was mostly empirical in nature, the results reported may have some immediate practical value with respect to clinical and technological applications.

4.3.1 Training speakers to communicate effectively with different listener populations

One potential clinical application is in training talkers to produce more intelligible speech when they interact with different listener populations, based on observed acoustic-phonetic correlates of speech intelligibility for these listeners. This and previous studies have demonstrated that clear speech is helpful for normal-hearing listeners in noisy hard-to-hear conditions, listeners with hearing impairment, aged listeners with or without hearing loss, and cochlear implant users. Schum (1996) found that, with instructions, both younger and older adults can be trained to alter their speaking habits to produce speech that is significantly easier for older adults to understand. It seems that some talkers may benefit from training on how to produce clear speech patterns that improve speech intelligibility, since not all individuals can instinctively produce clear speech that is highly intelligible. In order to make an effective training program, it is essential to understand what acoustic modifications are actually related to speech intelligibility benefits for the targeted listeners. Although intervention programs have been proposed to train communication partners of hearing-impaired listeners to use clear speech in order to optimize speech perception performance (Gagné and Jennings, 2000; Schum, 1997; Tye-Murray and Schum, 1994), no specific instructions/guidelines for speech patterns that improve fricative intelligibility have been provided. Since fricatives are hard to identify due to having significant information predominantly in higher frequency regions, and since the clear speech of untrained speakers seems to involve even higher-frequency cues, and since this transposition seems to make some sounds even harder to identify for these speakers, one obvious take-away from the present research is that some advantage may be gained from clear speech which involves increased CVR more than shifts to higher frequency regions.

4.3.2 Implications for digital signal-processing technique: prosthetic hearing devices

Analysis of naturally occurring acoustic-phonetic enhancements that are beneficial in speech perception could inform automatic signal-processing techniques that aim to enhance the intelligibility of (some aspect of) a speech signal for some listener population. Thus, one motivation for studying differences between conversational and clear speech is to design more intelligent algorithms for hearing aid devices that "transform the (incoming signal of) conversational speech into clear speech" (Helfer, 1998). Since the fricatives have considerable acoustic information in higher frequency regions where listeners with sloping loss have reduced audibility, it is important to consider how these problematic sounds might be best enhanced. Considering that frequency-dependent gain and dynamic amplitude compression techniques are common in prosthetic devices today, one conclusion that might be drawn from the present research is that, as far as processing limitations and recognition accuracy permit, classification of segments as specific fricatives or fricative classes, so that they can be processed differently depending on their identity, might provide considerable additional benefit (*cf* Hazan and Simpson, 1998, 2000).

4.3.3. Implications for digital signal-processing technique; auditory training program

Knowledge of the perceptual benefits associated with clear speech strategies can also inform processing techniques used in plasticity-based auditory training tasks. Research with language-impaired children and second-language learners (e.g. McClelland *et al.*, 2002; McCandlis *et al.* 2002; Merzenich et al., 1996; Nagarajan et al, 1998; Tallal, et al., 1996) suggests that perceptual learners might benefit most from exposure first to highly-differentiated, salient exemplars of phonetic categories that are likely to be robustly represented neurophysiologically, and then gradually, systematically introduced to more subtly differing tokens so as to drive the system to represent the relevant contrasts most efficiently. Since such training might simply involve categorization of sounds in minimal pairs, knowledge of Style × Fricative and Fricative × Misperception influences on production and perception like those described here can readily be incorporated into adaptive training procedures that are best suited for a listener or population.

4.3.3 Implications for human-computer error resolution

Research has demonstrated that when talking to interactive systems and encountering recognition errors, language users tend to adopt a clarified style of speech, as they do in inter-personal speech (Oviatt *et al.*, 1996, 1998a and b). Unfortunately, clear speech introduces new sources of variability into the task of spoken language processing that can degrade the recognition performance since it differs from the productions used to train the recognizer, and it has been associated with elevated rates of system recognition failure (Shriberg *et al.*, 1992). Since the elicitation method introduced in the current project involved a human-machine interactive program that simulated recognition errors in order to elicit clear speech, knowledge of which acoustic modifications speakers made when attempting to resolve the errors in the present study may be useful in developing dynamic, adaptive, and user-centered approaches to speech recognition. Oviatt and her colleagues investigated changes in speech after recognition errors, but analyses were focused on very limited aspects of productions, i.e., duration, amplitude, and fundamental frequency, phonological alterations, and prosodic changes. The present results suggest that it may be necessary to take into account fine-grained acoustic-phonetic changes of individual phonemes or phoneme classes in order to resolve these issues.

5.4 Directions for further research

The present research focused primarily on (1) documenting the global, fricative-dependent, and feedback-dependent acoustic changes that occur in /a/-C-/a/ sequences containing clearly produced English fricatives and (2) evaluating the effects of these alternations on perception of these sounds in noisy conditions by normal and (simulated) hearing-impaired listeners. We leave to future study: (1) a more thorough investigation of potential higher-order acoustic correlates of the clear speech effect in fricatives. This could be accomplished straightforwardly by using the results of the adaptive design described here to inform blocked-design experiments that are better controlled (and powered) for the distribution of fricatives, styles, and

SNR values across speakers and tokens. (2) Examination of the extent to which the clear fricative effects observed here hold for other listener groups. Specifically, it will be of interest to determine whether clearly produced fricatives enhance the intelligibility for non-native listeners who have different fricative phonemes in their language, aged listeners who have been reported to have difficulties understanding consonants that cannot completely be accounted for by audibility concerns, and cochlear implant users, for whom the incoming signal is distorted. (3) The effects of clear production on memory of fricative sounds. Previous studies (*e.g.* Rabbitt, 1968, 1990; Surprenant, 1999, 2005) have suggested that signal salience influences auditory memory: for example, signal distortion, background noise, and hearing impairment could interfere with memory for speech sounds even when those sounds are identifiable. To the author's knowledge, there are no detailed studies that determine whether clear speech is easier to memorize and less susceptible to (more resistant to) memory decay than conversational speech.

Bibliography

Baum, S.R. and Blumstein, S.E. (**1987**). "Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English", J. Accoust. Soc. Am. **82**, 1073-1077.

Behrens, S.J. and Blumstein, S.E. (**1988**). "Acoustic characteristics of English voiceless fricatives: A descriptive analysis", J. Phonetics **16**, 295-298.

Best, C., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants", J. Phon. **20**, 305-330.

Bilger, R.C. and Wang, M.D. (**1976**). "Consonant confusions in patients with sensorineural hearing loss", J. Speech Hear. Res. **19**, 718-748.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound", IFA Proceedings **17**, 97-110.

Boersma, P. and Weenink, D. (**2000**). "Praat, a System for doing Phonetics by Computer, version 4.406" Institute of Phonetic Sciences, University of Amsterdam. Available: http://www.fon.hum.uva.nl/praat/.

Boothroyd, A. (**1984**). "Auditory perception of speech contrasts by subjects with sensorineural hearing loss", J. Speech Hear. Res. **27**, 134-144.

Bradlow, A.R. and Bent, T. (**2002**). "The clear speech effect for non-native listeners", J. Acoust. Soc. Am. **112**, 272-284.

Bradlow, A.R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for children with learning disabilities: Sentence perception in noise", J. Speech Lang. Hear. Res. **46**, 80-97.

Bybee, J. and Scheibman, J. (**1999**). "The effect of usage on degrees of constituency: the reduction of don't in English", Linguistics, **37**, 575-596.

Bybee, J. (**2000a**). "The phonology of the lexicon; evidence from lexical diffusion", in *Usage-Based Models of Language*, edited by M. Barlow and S. Kemmerer (Stanford: CSLI), pp. 65-85.

Bybee, J. (**2000b**). "Lexicalization of sound change and alternating environments", in *Laboratory Phonology V: Acquisition and the lexicon*, edited by M. Broe and J. Pierrehumbert (Cambridge: Cambridge University Press), pp. 250-268.

Bybee, J. and Hopper, P. (**2001**), "Introduction to frequency and the emergence of linguistic structure", in *Frequency and the emergence of linguistic structure,* edited by J. Bybee and P. Hopper (Amsterdom: John Benjamin), pp. 1-24.

Bybee, J. (**2002a**). "Phonological evidence for exemplar storage of multiword sequences", SSLA **24**, 215-221.

Bybee, J. (**2002b**). "Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change", Language variation and change **14**, 261-290.

Bybee, J. (**2006 to appear**). "From usage to grammar: The mind's response to repetition", Language **82**. Downloadable at: http://www.unm.edu/%7Ejbybee/Bybee%20plenary.pdf

Chen, F.R. (**1980**) "Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level", Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Cole, R. A. and Cooper, W. E. (**1975**). "Perception of voicing in English affricates and fricatives", J. Acoust. Soc. Am. **58**, 1280-1287.

Christensen, L. A. and Humes, L. E. (**1996**). "Identification of multidimensional complex sounds having parallel dimension structure", J. Acoust. Soc. Am. **99**, 2307-2315.

Christensen, L. A. and Humes, L. E. (**1997**). "Identification of multidimensional stimuli containing speech cues and the effects of training", J. Acoust. Soc. Am. **102**, 2297-2310.

Crowther, C. S. and Mann, V. A. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English", J. Acoust. Soc. Am. **92**, 711-722.

Crowther, C. S. and Mann, V. A. (**1994**). "Use of vocalic cues to consonant voicing and native language background: The influence of experimental design", Percept. Psychophys. **55**, 513-525.

Crystal, T.H. and House, A.S. (**1988**). "A note on the durations of fricatives in American English", J. Accoust. Soc. Am. **84**, 1932-1935.

Denes, P. (**1955**). "Effect of duration on the perception of voicing", J. Acoust. Soc. Am. **27**, 761-764.

Dubno, J. R., and Levitt, H. (**1981**). "Predicting consonant confusions from acoustic analysis", J. Accoust. Soc. Am. **69**, 249-261.

Dubno, J. R., Dirks, D. D., and Langhofer, L.R. (**1982**). “Evaluation of hearing-impaired listeners using a Nonsense-syllable Test. II. Syllable recognition and consonant confusion patterns”, J. Speech Hear. Res. **25**, 141-148.

Evers, V., Reetz, H., and Lahiri, A. (**1998**). “Crosslinguistic acoustic categorization of sibilants independent of phonological status”, J. Phonetics **26**, 345-370.

Feijóo, S., Fernández, S. and Balsa, Ramón. (**1998**). “Context effects in the auditory identification of Spanish fricatives /f/ and /θ/: Hyper and Hypospeech”, J. Acoust. Soc. Am. **103**, 2982.

Ferguson, S. H. (**2004**). “Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners”, J. Acoust. Soc. Am. **116**, 2365-2373.

Ferguson, S.H. (**2002**). “Vowels in clear and conversational speech: talker differences in acoustic features and intelligibility for normal-hearing listeners”, Doctoral dissertation, Indiana University, Bloomington, IN.

Ferguson, S.H. and Kewley-Port, D. (**2002**). “Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners”, J. Acoust. Soc. Am. **112**. 259-271.

Flege, J. E., Takagi, N., and Mann, V. (**1996**). “Lexical familiarity and English-language experience affect Japanese adults’ perception of /r/ and /l/”, J. Acoust. Soc. Am. **99**, 1161-1173.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R.N. (**1988**). “Statistical analysis of word-initial voiceless obstruents: Preliminary data”, J. Accoust. Soc. Am. **84**, 115-123.

Gagné, J.P., and Jennings, M. B. (**2000**). “Audiological rehabilitation: intervention services for adults with an acquired hearing loss”, in Audiologic Treatment Strategies edited by M. Valente, A. Hosford-Dunn, and R. J Roeser (Thieme Medical Publishers, New York), pp. 547-579.

Gagné, J.P., Masterson, V.M., Munhall, K.G., Bilida, N., and Querengesser, C. (**1994**). “Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech”, J. Acad. Rehabil. Audiol. **27**, 135-158.

Gagné, J. P., Querengesser, C., Folkeard, P., Munhall, K. G., and Masterson, V. M. (**1995**). "Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: an investigation of conversational and clear speech", Volta Rev. **97**, 33-51.

Gagné, J. P., Rochette, A-J., and Charest, M. (**2002**). "Auditory, visual and audiovisual clear speech", Speech Commun. **37**, 213-230.

Goldinger, S. D. (**1998**). "Echoes of Echoes? An Episodic Theory of Lexical Access", Psych. Review **105**, 251-279.

Guerlekian, J. A. (**1981**). "Recognition of the Spanish fricatives /s/ and /f/", J. Accoust. Soc. Am. **70**, 1624-1627.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables", Lang. Speech **1**, 1-7.

Hazan, V. and Rosen, S. (**1991**). "Individual variability in the perception of cues to place constraints in initial stops", Percept. Psychophys. **41**, 187-200.
Hazan, V. and Simpson, A. (1998). "The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise", Speech Commun. **24**, 211-226.

Hazan, V. and Simpson, A. (**2000**). "The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effect", Lang. Speech, **43**, 273-294.

Hedrick, M. S. (**1997**). "Effect of acoustic cues on labeling fricatives and affricates", J. Speech Lang. Hear. Res. **40**, 925-938.

Hedrick, M. S. and Carney, A. E. (**1997**). "Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants", J. Speech Lang. Hear. Res. **40**, 1445-1457.

Hedrick, M.S. and Ohde, R.N. (**1993**). "Effect of relative amplitude of frication on perception of place of articulation", J. Accoust. Soc. Am. **94**, 2005-2026.

Hedrick, M. S. and Younger, M. S. (**2003**). "Labeling of /s/ and /ʃ/ by listeners with normal and impaired hearing, revisited", J. Speech Lang. Hear. Res. **46**, 636-648.

Heinz, J.M. and Stevens, K.N. (**1961**). "On the properties of voiceless fricative consonants", J. Accoust. Soc. Am. **33**, 589-596.

Helfer, K. (**1997**). "Auditory and auditory-visual perception of clear and conversational speech", J. Speech Lang. Hear. Res. **40**, 432-443.

Helfer, K. (**1998**). “Auditory and auditory-visual recognition of clear and conversational speech by older adults”, J. Am. Acad. Audiol. **9**, 234-242.

Hintzman, D. L. (**1986**). ““Schema abstraction” in a multiple-trace memory model”, Pscyhol. Review **93**, 328-338.

Hughes, G.W. and Halle, M. (**1956**). “Spectral properties of fricative consonants”, J. Accoust. Soc. Am. **28**, 303-310.

Iverson, P. and Bradlow, A.R. (**2002**). “The recognition of clear speech by adult cochlear implant users”, in *Temporal Integration in the Perception of Speech*, edited by S. Hawkins and N. Nguyen (Cambridge: Center for Research in the Arts, Social Sciences, and Humanities). Downloadable at: http://www.phon.ucl.ac.uk/home/paul/IBTIPSPoster.pdf

Jesus, L.M.T. and Shadle, C. H. (**2002**). “A parametric study of the spectral characteristics of European Portuguese fricatives”, J. Phonetics. **30**, 437-464.

Johnson, K. (**1997**). “Speech perception without speaker normalization: An exemplar model”, in *Talker variability in speech processing* edited by K. Johnson and J. Mullenix (Academic press: San Diego), pp. 145-166.

Johnson, K., Flemming, E. and Wright, R. (**1993**). “The hyperspace effect: Phonetic targets are hyperarticulated”, Language **69**, 505-528.

Jongman, A. (**1989**). “Duration of frication noise required for identification of English fricatives”, J. Accoust. Soc. Am. **85**, 1718-1725.

Jongman, A., Wayland, R., and Wong, S. (**2000a**). “Acoustic characteristics of English fricatives”, J. Accoust. Soc. Am. **108**, 1252-1263.

Jongman, A., Wang, Y., and Sereno, J. (**2000b**). “Acoustic and perceptual properties of English fricatives”, *Proceedings of the 6th International Conference on Spoken Language Processing*, I, 536-539.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., and Supowit, A. (1994). “Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation”, Vision Res. **34**, 885-912.

Kruschke, J. K. (**1992**). “ALCOVE: An exemplar-based connectionist model of category learning”, Psych. Review. **99**, 22-44.

Krause, J. C. and Braida, L. D. (2002). “Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility”, J. Accoust. Soc. Am. 112, 2165-2172.

Krause, J.C. and Braida, L.D. (**2004**). “Acoustic properties of naturally produced clear speech at normal speaking rates”, J. Accoust. Soc. Am. **115**, 362-378.

LaRiviere, C., Winitz, H., and Herriman, E. (**1975**). “The distribution of perceptual cues in English prevocalic fricatives”, J. Speech Hear. Res. **18**, 613-622.

Lindblom, B. (**1990**). “Explaining phonetic variation: A sketch of the H & H theory”, in *Speech Production and Speech Modeling*, edited. by W.J. Hardcastle and A. Marchal (Kluwer Academic Publishers, Dordrecht),. pp. 403-439.

Lindblom, B., Brownlee, S., Davis, B., and Moon, S-J. (**1992**). “Speech transforms”, Speech Commun. **11**, 357-368.

Lindblom, B. (**1996**). “Role of articulation in speech perception: Clues from production”, J. Accoust. Soc. Am. **99**, 1683-1692.

Liu, S., Del Rio, E., Bradlow, A.R., and Zeng, F-G. (**2004**). “Clear speech perception in acoustic and electric hearing”, J. Accoust. Soc. Am. **116**, 2373-2383.

Maniwa, K., Jongman, A., and Wade, T. (submitted). “Acoustic characteristics of clear English fricatives”, J. Accoust. Soc. Am.

Maniwa, K., Jongman, A., and Wade, T. (In press). “Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners”, J. Accoust. Soc. Am.

Mann, V. A., and Repp, B. H. (**1980**). “Influence of vocalic context on perception of the [ʃ]-[s] distinction”, Percept. Psychophys. **28**, 213-228.

Mann, V. A. and Soli, S. D. (**1991**). “Perceptual order and the effect of vocalic context on fricative perception”, Percept. Psychophys. **49**, 399-411.

Mathworks, Inc., The. (**2000**). “MATLAB, The language of technical computing, version 7.0.0.19920”, http://www.mathworks.com/.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (**2002**), “Success and failure in teaching the [r]-[l] contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception”, Cogn. Affect. Behav. Neurosci. **2**, 89-108.

McCasland, G. P. (**1979a**). “Noise intensity and spectrum cues for spoken fricatives”, J. Accoust. Soc. Am.Suppl.1 **65**, S78-S79.

McCasland, G. P. (**1979b**). “Noise intensity cues for spoken fricatives”, J. Accoust. Soc. Am.Suppl.1 **66**, S88.

McClelland, J. L., Fiez, J. A., and McCandliss, B. D. (**2002**). “Teaching /r/-/l/ discrimination to Japanese adults: behavioral and neural aspects”, Physiol. Behav. **77**, 657-662.

McGowan, R., and Nittrouer, S. (**1988**). “Differences in fricative production between children and adults: evidence from an acoustic analysis of /ʃ/ and /s/”, J. Accoust. Soc. Am. **83**, 229-236.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., and Tallal, P. (**1996**), “Temporal Processing Deficits of Language-Learning Impaired Children Ameliorated by Training”, Science **5**, 77-81.

Miller, G.A. and Nicely, P.A. (1955). “An analysis of perceptual confusions among some English consonants”, J. Accoust. Soc. Am. **27**, 338-352.

Moon, S.-J. and Lindblom, B. (**1994**). “Interaction between duration, context, and speaking style in English stressed vowels”, J. Accoust. Soc. Am. **96**, 40-55.

Moore, B. C. J. and Glasberg, B. R. (**1993**). “Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech”, J. Accoust. Soc. Am. **94**, 2050-2062.

Nagarajan, S. S., Wang, X., Merzenich, M. M., Schreiner, C. E., Johnston, P., Jenkins, W. M., Miller, S., and Tallal, P. (**1998**). “Speech modifications algorithm used for training language learning-impaired children”, IEEE Trans. Rehabil. Eng. **6**, 257-268.
Nissen, S.L. and Fox, R.A. (**2005**). “Acoustic and spectral characteristics of young children’s fricative productions: A developmental perspective”, J. Accoust. Soc. Am. **118**, 2570-2578.

Nittrouer, S. (**1992**). “Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries”, J. Phon. **20**, 351-382.

Nittrouer, S. (**1995**). “Children learn separate aspects of speech production at different rates: Evidence from spectral moments”, J. Accoust. Soc. Am. **97**, 520-530.

Nittrouer, S. (**2002**). “Learning to perceive speech: How fricative perception changes, and how it stays the same”, J. Accoust. Soc. Am. **112**, 711-719.

Nittrouer, S. and Miller, M. E. (**1997a**). “Predicting developmental shifts in perceptual weighting schemes”, J. Accoust. Soc. Am. **101**, 2253-2266.

Nittrouer, S. and Miller, M. E. (**1997b**). “Developmental weighting shifts for noise components of fricative-vowel syllables”, J. Accoust. Soc. Am. **102**, 572-580.

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R.S. (**1989**). “The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults”, J. Speech Hear. Res. **32**, 120-132.

Nosofsky, R. M. (**1986**). “Attention, Similarity, and the Identification-Categorization Relationship”, J. Exp. Psychol. Gen. **115**, 39-57.

Ohala, J.J. (**1994**). “Acoustic study of clear speech: A test of the contrastive hypothesis”, Proceedings from the International Symposium on Prosody, Pacific Convention Plaza Yokohama, Japan, 75-89.

Oviatt, S., Levow, G-A., MacEachern, M., and Kuhn, K. (**1996**). “Modeling hyperarticulate speech during human-computer error resolution”, Proc. of the Int. Conf. on Spoken Language Processing, **2**, 801-804.

Oviatt, S., MacEachern, M., and Levow, G-A. (**1998b**). “Predicting hyperarticulate speech during human-computer error resolution”, Speech Commun. **24**, 1-23.

Oviatt, S., Levow, G-A., Moreton, E., and MacEachern, M. (**1998b**). “Modeling global and focal hyperarticulation during human-computer error resolution”, J. Accoust. Soc. Am. **104**, 3080-3098.

Owens, E. (**1978**). “Consonant errors and remediation in sensorineural hearing loss”, J. Speech. Hear. Disord. **43**, 331-347.

Owens, E., Benedict, M., and Schubert, E. D. (**1972**). “Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairments”, J. Speech Hear. Res. **15**, 308-322.

Payton, K.L., Uchanski, R.M., and Braida, L.D. (**1994**). “Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing”, J. Accoust. Soc. Am. **95**, 1581-1592.

Perkell, J.S., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F.H. (**2004**) “The distinctiveness of speakers’ /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect”, J. Speech Lang. Hear. Res. **47**, 1259-1269.

Picheny, M.A., Durlach, N.I., and Braida, L.D. (**1985**) “Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech”, J. Speech Hear. Res. **28**, 96-103.

Picheny, M.A., Durlach, N.I., and Braida, L.D. (**1986**). “Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech”, J. Speech Hear. Res. **29**, 434-446.

Picheny, M.A., Durlach, N.I., and Braida, L.D. (**1989**). “Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differentiate in intelligibility between clear and conversational speech”, J. Speech Hear. Res. **32**, 600-603.

Pierrehumbert, J. (**2001**). “Exemplar dynamics: Word frequency, lenition, and contrast”, in “*Frequency effects and the emergence of linguistic structure*” edited by J. Bybee and P. Hopper (Amsterdom: John Benjamin), pp. 137-157.

Pierrehumbert, J. (**2002**). “Word-specific phonetics”, in *Laboratory Phonology VII* edited by C. Gussenhoven and N. Warner (Mouton de Gruyter: Berlin), pp. 101-140.

Rabbitt, P.M.A. (**1990**). “Mild hearing loss can cause apparent memory failures which increases with age and reduce with IQ”, Acta. Otolaryngol Suppl. **476**, 167-175.

Rabitt, P.M. (**1968**). “Channel-capacity, intelligibility, and immediate memory”, Q J Exp. Pshychol. **20**, 241-248.

Raphael, L. (**1972**). “Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English”, J. Acoust. Soc. Am. **51**, 1296-1303.

Schum, D. (**1996**). “Intelligibility of clear and conversational speech of young and elderly talkers”, J. Am. Acad. Audiol. **7**, 212-218.

Schum, D. (**1997**). “Beyond hearing aids: The clear speech training technique”, The Hear. J. **50**, 36-40.

Shadle, C.H. and Mair, S.J. (**1996**). “Quantifying spectral characteristics of fricatives”, Proceedings from the International Conference on Spoken Language Processing (ICSLP), Philadelphia. pp.1521-1524.

Sher, A. E. and Owens, E. (**1974**). “Consonant confusions associated with hearing loss above 2000 Hz”, J. Speech Hear. Res. **17**, 669-681.

Shriberg, E., Wade, E., and Price, P. (**1992**). "Human-machine problem solving using spoken language systems (SLS): Factors affecting performance and user satisfaction", Proc. of the DARPA Speech and Natural Language Workshop, 49-54.

Singh, S., and Black, J. W. (**1966**). "Study of twenty-six intervocalic consonants as spoken and recognized by four language groups", J. Accoust. Soc. Am. **39**, 372-387.

Slaney, M. (**1998**). "Auditory Toolbox, version 2.0", http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/

Smiljanić, R. and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English", J. Accoust. Soc. Am. **118**, 1677-1688.

Soli, S.D. (**1981**). "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation", J. Accoust. Soc. Am. **70**, 976-984.

Surprenant, A. (**1999**). "The effects of noise on memory for spoken syllables", Int. Journal of Pshych. **34**, 328-333.

Surprenant, A. (**2005**). "Effects of noise on identification and serial recall of nonsense syllables in older and younger adults", *Aging, Neuropsychology, and Cognition.*

Stevens, K.N. (**1971**). "Airflow and turbulence for noise for fricative and stop consonants: Static consideration", J. Accoust. Soc. Am. **50**, 1182-1192.

Stevens, K.N. (**1985**). "Evidence for the role of acoustic boundaries in the perception of speech sounds", in *Phonetic linguistics: Essays in honor of Peter Ladefoged*, edited by V. Fromkin (New York: Academic Press). pp. 243-255.

Strevens, P. (**1960**). "Spectra of fricative noise in human speech", Lang. Speech **3**, 32-49.

Tallal, P., Miller, S. L., Bedi, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., and Merzenich, M. M. (**1996**). "Language Comprehension in Language-Learning Impaired Children Improved with Acoustically Modified Speech", Science **5**, 81-84.

Tye-Murray, N. and Schum, L. (**1994**). "Conversational training for frequent communication partners", in Research in *Audiological Rehabilitation: Current Trends and Future Directions* edited by J.-P.Gagné, and N. Tye-Murray (Monograph Supplement: J. Acad. Rehab. Aud., Vol. 27), pp. 209-222.

Uchanski, R.S., Choi, S.S., Braida, L.D., Reed, C.M., and Durlach, N.I. (**1996**). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate", J. Speech Hear. Res. **39**, 494-509.

Wang, M.D. and Bilger, R.C. (**1973**). "Consonant confusions in noise: a study of perceptual features", J. Accoust. Soc. Am. **54**, 1248-1266.

Whalen, D. H. (**1981**). "Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary", J. Accoust. Soc. Am. **69**, 275-282.

Whalen, D.H. (**1991**). "Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices", J. Accoust. Soc. Am. **90**, 1776-1785.

Yeni-Komshian, G.H. and Soli, S.D. (**1981**). "Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation", J. Accoust. Soc. Am. **70**, 966-975.

Zeng, F.-G. and Turner, C. W. (**1990**). "Recognition of voiceless fricatives by normal and hearing-impaired subjects", J. Speech Hear. Res. **33**, 440-449.