Acoustic characteristics of clearly spoken English fricatives

Kazumi Maniwa[a] and Allard Jongman
Department of Linguistics, The University of Kansas, Lawrence, Kansas 66044

Travis Wade
Posit Science Corporation, 7th Floor, 225 Bush Street, San Francisco, California 94104

Speakers can adopt a speaking style that allows them to be understood more easily in difficult communication situations, but few studies have examined the acoustic properties of clearly produced consonants in detail. This study attempts to characterize the adaptations in the clear production of American English fricatives in a carefully controlled range of communication situations. Ten female and ten male talkers produced fricatives in vowel-fricative-vowel contexts in both a conversational 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 style effects, several of which (consonant duration, spectral peak frequency, and spectral moments) were consistent with previous findings and a few (notably consonant-to-vowel intensity ratio) of which were not; (2) specific acoustic modifications in clear productions of fricatives were influenced by the nature of the recognition errors that prompted the productions and were consistent with efforts to emphasize potentially misperceived contrasts both within the English fricative inventory and based on feedback from the simulated listener; and (3) talkers differed widely in the types and magnitude of all modifications.
© 2009 Acoustical Society of America. [DOI: 10.1121/1.2990715]

I. 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 3–38 percentage points relative to "normal" conversational speech for hearing-impaired listeners in quiet (Picheny *et al.*, 1985; Uchanski *et al.*, 1996) 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) non-native 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, alveolar flapping, and more frequent schwa insertion (Bradlow *et al.*, 2003; Krause and Braida, 2004; Picheny *et al.*, 1986). Previous study on the fine-grained acoustic-phonetic characteristics of clear speech has mainly considered vowels, noting increases in vowel durations, expanded $F1 \times F2$ 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., see Miller and Nicely (1955)], it is surprising 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).

---

[a]Present address: FB Sprachwissenschaft, Raum F521A, Universität Konstanz, 78457 Konstanz, Germany; electronic mail: kazumi.maniwa@uni-konstanz.de

© 2009 Acoustical Society of America

It is particularly unfortunate that clearly produced fricative consonants have not been the subject of more observation, since previous consonant confusion analyses have reported that fricatives, especially nonsibilants, contribute a large source of errors for hearing-impaired listeners and for normal-hearing listeners in noise [e.g., see Bilger and Wang (1976), Miller and Nicely (1955), and Wang and Bilger (1973)]. A few studies have considered vocal effort and rate modifications and hyperarticulation in describing fricative acoustics and perception, but clear production was not the primary focus of these studies, which were therefore inconclusive with respect to specific clear speech alterations. Jesus and Shadle (2002) reported that fricative amplitude and spectral slope increased with vocal effort but did not offer a statistical analysis of these results or consider other properties of the sounds. Perkell *et al.* (2004) observed that average (voiceless) sibilant duration increased from fast-rate speech to normal speech to "clear speech" (obtained by asking talkers to pronounce words carefully without increasing loudness). They also measured the spectral center of gravity but did not find differences related to speaking style. Feijóo *et al.* (1998) investigated the intelligibility of Spanish nonsibilant fricatives in "hyperarticulated" and "hypoarticulated" speech, but acoustic data were provided only for duration and energy level and without statistical analysis, and it was unclear how the two styles were elicited. In any case, hyperarticulated speech did not lead to better intelligibility. This study was designed to provide a comprehensive acoustic description of adaptations that occur in the clear production of English fricatives.

A. Acoustic properties of English fricative sounds

Several studies have attempted to delineate stable acoustic correlates of the fricative place of articulation and voicing. Parameters that seem to influence identification include gross spectral shapes and peak frequencies (Behrens and Blumstein, 1988; Hughes and Halle, 1956; Jongman *et al.*, 2000; Strevens, 1960), the first four moments of the spectral energy distribution (Forrest *et al.*, 1988; Jongman *et al.*, 2000; Nissen and Fox, 2005; Nittrouer, 1995; Nittrouer *et al.*, 1989; Shadle and Mair, 1996), the slopes of lines fitted to spectra in lower and higher frequency regions (Evers *et al.*, 1998; Jesus and Shadle, 2002), formant transition information (Jongman *et al.*, 2000; McGowan and Nittrouer, 1988; Nittrouer *et al.*, 1989; Soli, 1981), overall amplitude (Behrens and Blumstein, 1988; Jongman *et al.*, 2000; Stevens, 1971; Strevens, 1960), amplitude relative to the neighboring vowel in specific frequency regions (Hedrick and Ohde, 1993; Jongman *et al.*, 2000; Stevens, 1985), and duration (Baum and Blumstein, 1987; Crystal and House, 1988; Jongman, 1989; Jongman *et al.*, 2000). Briefly, alveolar fricatives (/s/, /z/) are characterized by spectral energy [above 4 kHz, Hughes and Halle (1956)] and major peaks [3.5–5 kHz, Behrens and Blumstein (1988); 6–8 kHz, Jongman *et al.* (2000)] at higher frequencies compared to palato-alveolars (/ʃ/, /ʒ/; 2–4 kHz; [Hughes and Halle (1956), Behrens and Blumstein (1988)], which display larger overall relative amplitudes. Dental (/θ/, /ð/) and labio-dental (/f/, /v/) fricatives

show relatively flat spectra below 10 kHz with no dominating peaks, while alveolar and palato-alveolar fricatives have well-defined peaks. Nonsibilants (/θ/, /ð/, /f/, /v/) show higher standard deviations (SDs), lower overall amplitudes, and shorter durations than sibilants (/s/, /ʃ/, /z/, /ʒ/). Thus, these parameters clearly distinguish sibilants from nonsibilants and from each other but are less effective at determining the place of articulation for nonsibilants. However, it was observed that the onset of F2 at the fricative-vowel boundary was significantly higher for dental fricatives than for labio-dental fricatives (Jongman *et al.*, 2000; Nittrouer, 2002) and higher for palato-alveolar fricatives than for alveolar fricatives (Jongman *et al.*, 2000; McGowan and Nittrouer, 1988; Nittrouer *et al.*, 1989). 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.* 2000). 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 compared to their voiced counterparts.

B. Contrastive effects in clear speech

A secondary question of this study was whether talkers exhibit more specific fricative-dependent, segmental contrast-enhancing changes. It has been suggested that clear speech modifications are inventory dependent (that is, related to the specific phonetic contrasts that must be maintained within a language) and effectively increase the auditory distance between neighboring sound categories. For example, VOT for voiceless stop consonants 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, maximizing 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 the 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 a phonetic inventory and a general effort to maintain these contrasts. We will refer to such efforts as *inventory-level* contrast-enhancing modifications.

It is less clear whether talkers may also attempt to preserve contrast at a more local level, adapting online in response to perception errors that are likely to occur in specific contexts. According to Lindblom's H and H theory (1990, 1996), speakers constantly assess the listeners' needs for explicit signal information and modulate their speech along a continuum from hypo- to hyperspeech 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 and lexicon, etc.). We pre-

dicted that explicit feedback from a 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 mistaken 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 pseudomisrecognitions from the experimenter 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 an effort not only to improve the intelligibility of a 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. Thus 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 (1994) 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 extends Ohala's (1994) 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, namely, /ɑ/. Naturally, including only this context does not allow us to examine differences in phonetic context influences on fricative production [or perception, e.g., Mann and Repp (1980), Mann and Soli (1991), Soli (1981), Yeni-Komshian and Soli (1981), and Whalen (1981)] as a function of speaking style or even to generalize our findings conclusively to other vowel contexts. However, it allows for maximal control in documenting fundamental aspects of (1) how clear speech influences the acoustics of fricatives in general and (2) how successfully listeners can enhance acoustic distance in terms of place of articulation, sibilance, and voicing and the distance between an intended target fricative and an anticipated misperception. As discussed above, we will refer to efforts of the latter type specifically as *local context-dependent* modifications.

An elicitation method somewhat similar to the one used in this study (Maniwa, 2006; Maniwa *et al.*, 2006) was also recently employed to examine the effects of linguistic focus on the production of fricatives in a contrastive context. Silbert and de Jong (2008) measured the duration, first four spectral moments, and power of fricatives produced in the carrier sentence "No, I said [target]," where the target was a fricative-/a/ syllable that the production was intended to disambiguate from a syllable differing in fricative voicing or place of articulation. As in the Ohala (1994) study, no specific disambiguation effects were observed, although the focused productions exhibited some general characteristics of clear speech (increased duration, for example) compared to

unfocused productions. Again, it is difficult to make generalizations based on this null result since only a small number of talkers (4), productions (576 total), fricatives (palato-alveolars were not included), and measurements were considered and since elicitation involved a somewhat explicit request for focus that might not have captured the speakers' ability to adapt spontaneously.

C. 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 the corpus of Picheny *et al.* increased VOTs for both voiceless and voiced plosives in clear speech, while the other two increased VOTs only for the voiceless one. This speaker also decreased the intensity for fricatives in clear speech, while the other two speakers showed the opposite pattern. The female talker from the study of Bradlow *et al.* (2003) decreased her speaking rate in clear speech to a far greater degree than the male talker. These two talkers also differed noticeably in the f0, vowel space, and CVR differences between clear and conversational speeches. The female talker from the database of Liu *et al.* (2004) 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 $F1 \times F2$ space variability and the magnitude of the increase in the 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 from only a few speakers, it is unclear whether the patterns of variability observed across speakers and gender would be maintained more generally, or if still other strategies would emerge. This study examined the productions of 20 speakers (10 females and 10 males) to address these questions more conclusively.

D. 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 predicted (henceforth hypothesis 1) that clear fricatives would be (i) longer, (ii) amplified relative to neighboring vowels, and (iii) higher in spectral content, including peak frequencies, spectral mean values, and related measures. Second, are clear speech modifications dependent on the nature of the contrasts in a fricative inventory and/or more local context information provided by "listener" feedback? We expected that, in general (hypothesis 2), clear productions would be influenced by the

perceived likelihood of different misidentification patterns. More specifically, we predicted that on average (i) fricative categories would differ more from minimally contrastive categories in clear than in conversational speech (inventory-level modifications) and (ii) fricatives repeated after misapprehension for similar sounds would be most different from the sounds they were mistaken for (local context-dependent modifications). Finally, in what ways do talkers vary in the production of clear fricatives? We predicted (hypothesis 3) that cross-talker differences would be seen both in the types of modifications that are made and in the extent of these changes.

II. METHOD

A. Participants

Twenty talkers (ten females and ten males) 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 Upper Midwest or Pacific Southwest American English. Talkers reported normal hearing and no history of speech or language disorders. They volunteered for the experiment without monetary compensation.

B. Materials

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

C. Procedures and apparatus

The participants' speech was recorded digitally at a 44.1 kHz sampling rate (16 bit resolution) in a sound-attenuating booth in the Phonology Laboratory, UC Berkeley, using a Marantz PMD670 recorder and a Shure SM-10 A headset microphone (frequency responses of 50–15 000 Hz). The microphone was placed 2.5 cm away from the corner of a talker's mouth at a 45° 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 as follows: "afa," "atha," "asa," "asha," "ava," "adha," "aza," and "azha." Participants first read these syllables aloud a few times to become familiar with uniquely spelled syllables. A pronunciation key was available for reference during the session.

The recording session was divided into two parts: warmup and experiment. Programs to provide prompts and feedback were designed using MATLAB 7.0.0.1 (The Mathworks, Inc., 2000). During warmup, talkers produced five repetitions of each VCV in each of two blocks, in response to prompts appearing on a monitor. In the first block, talkers read VCV syllables in a manner approximating the way they spoke in everyday conversation; in the second block, they were instructed to speak more carefully, as if talking to a hearing-impaired or elderly person. This warmup served to familiarize talkers with the interface and materials, to allow them to rehearse the two styles, and to provide a "baseline" recording of speech produced before talkers became aware of the rate and types of misperceptions that would be encountered during the experiment. Speakers were not explicitly instructed or coached on stress type or placement since this might have created a bias toward one style or the other or caused speakers to imitate the experimenter instead of producing clear speech modifications spontaneously.

The elicitation method for the experimental session resembled the one used by Ohala (1994). Before the session, a participant 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. Productions in response to these initial prompts served as the "conversational speech" in our acoustic analyses. Participants were told that the program would "guess" which syllables were spoken and would 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 was terminated and the program moved on to the next stimulus. If a guess was scored as incorrect, the speaker was given a chance to repeat the target stimulus, doing his or her best to deliver it as intelligibly as possible. These repeated productions served as clear speech in acoustic analyses. The program's guesses were, in fact, unrelated to the speaker's production pattern and represented either (1) the correct response, (2) the voicing-matched but place-unmatched incorrect responses (e.g., /θ/, /s/, and /ʃ/, for /f/), (3) the voicing-unmatched, place-matched incorrect responses (e.g., /v/ for /f/, and (4) the "???" ("don't know") responses. Each response occurred five times for each VCV during the experiment. Thus, there were 30 conversational (5×one following correct response, three place errors, one voicing error, one ???) and 25 total clear (5×three preceding place errors, one voicing error, one ???) 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 participant's second production, a second guess was displayed, which was correct 75% of the time and random otherwise; the participant scored this guess before finally continuing to the next trial. The purpose of this second guess was to encourage optimal effort in clear productions by giving the impression that (1) the program's guesses were actually based on a speaker's productions, (2) recognition performance improved for clear productions, and (3) this performance was actually being recorded for analysis instead of predetermined by the elicitation program. Recording sessions lasted 60–70 min, including the warmup and a 10 min break halfway through the main experiment.

D. Data processing and acoustic measurements

Recordings were hand annotated into VCV segments using the PRAAT speech analysis software (Boersma and Weenink, 2000), equalized for the total rms amplitude, and

further segmented and analyzed using PRAAT and MATLAB. Semiautomatic fricative segmentation was achieved following previous studies (Behrens and Blumstein, 1988; Jongman *et al.*, 2000; 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 second order Butterworth filter to minimize 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 (1) normalizing the log of this envelope to the range $[-1, 1]$ and (2) taking the single continuous region closest to the center of the production for which the resulting sequence was above zero 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.

The acoustic analysis considered 14 spectral, amplitudinal, and duration parameters that previous studies indicate may work in combination with signal fricative contrasts. Spectral measures included the discrete Fourier transform (DFT) spectral peak frequency (1), the first four spectral moments (M1–M4; 2–5), F2 onset transitions (6), spectral slopes below (7) and above (8) peak frequencies, and the average fundamental frequency (f0) of adjacent vowels (9). Amplitudinal measures included normalized rms amplitude (10) and a measure previously referred to [e.g., see Hedrick and Ohde, (1993)] as the "relative amplitude," the amplitude of a fricative relative to the following vowel in the F3 region for sibilants and the F5 region for nonsibilants (since these regions contain important prominences for the two fricative types). To distinguish this measure from the overall normalized amplitude (10), we will refer to it here as the *frequency-specific relative amplitude* (FSRA) (11). Other amplitude-related measures included harmonics-to-noise ratio (HNR) (12) and energy below 500 Hz (13). Finally, we considered the total fricative duration (14). As described in Sec. I A, these 14 measures seem to be the most important for distinguishing fricative place and voicing contrasts. A few measures that had been previously employed but either yielded inconsistent, contradictory, or unreliable results for these contrasts (e.g., F2 range, F3 transition, and locus equations) or are not yet fully understood with respect to fricative production and perception [i.e., fricative noise modulation and "dynamic amplitude"—e.g., see Jackson and Shadle (2000), Jesus and Shadle (2002), Pincas and Jackson (2006), Shadle and Mair (1996)] were not considered.

Except where noted otherwise, all analyses considered 20 ms Hamming windowed segments at five locations (W1–W5), centered over the fricative onset (25%, 50%, and 75% points;) and offset. All spectral measures were based on a 44 100-point DFT of this (zero padded to 1s) segment. Ensemble averaging across tokens within a given speaker/fricative/style condition was used to reduce error in spectral estimates; each spectrum $X(f)$ considered below, then, represents an average of 5–20 (depending on the comparison; see Sec. II E for analysis details) $|\text{DFT}|^2$ values at frequencies of 50–15 000 Hz (the frequency response of the microphone) in 1 Hz increments. In the case of spectral peak and slope measures, the windowed segment was first pre-emphasized with a factor of 0.98. The spectral peak was defined as the frequency bin corresponding to the largest value in $X(f)$. M1 was defined as the center of gravity of the spectrum (the mean frequency weighted by $X(f)$). The remaining three moments (M2–M4) were obtained by first calculating the sum $(\text{M}n = (\Sigma(f - \text{M1})^n X(f)) / \Sigma X(f)$ and then normalizing by the variance (M2) as follows. The SD a measure of the diffuseness of the spectrum around the center of gravity, was taken as the square root of the raw M2 measurement. Skewness, an indicator of spectral tilt, measuring asymmetry in the spectrum toward frequencies far above (positive values) or below (negative values) the center of gravity was obtained by dividing the raw M3 value by the 1.5 power of M2. Finally, kurtosis, a measure of the peakedness of the distribution, was obtained by dividing M4 by the square of M2 and subtracting 3. For space reasons, henceforth we use the notation M1–4 to refer to the normalized mean, SD, skewness, and kurtosis values.

F2 values were estimated using a linear prediction-based method [the Burg algorithm; Childers (1978), as implemented in PRAAT], derived at the fricative onset and offset and each vowel midpoint from an analysis that found at most five formants below 5000 Hz (male speakers) or 5500 Hz (female speakers).

Spectral slopes were computed following the procedures described by Evers *et al.* (1998) and Jesus and Shadle (2002). Lines were fit by least squared error to $\log(X(f))$ across two regions defined by the average peak frequency (across talkers and productions) for a place of articulation (8000 Hz for alveolars, 3300 Hz for alveo-palatals, and 6500 Hz for all nonsibilants). A low-frequency slope (dB/kHz) was derived from the spectral values below this peak, and a high-frequency slope was derived from the peak to 15 kHz.

The fundamental frequency was derived using an autocorrelation-based algorithm (Boersma, 1993). It was averaged across the vowels preceding and following the target fricative. The normalized amplitude was taken as the rms amplitude ratio (dB) between the same five windowed fricative segments described above and the average of the two surrounding vowels. The use of both vowels for the f0 and amplitude analysis was necessary because some speakers tended to place emphasis on the first vowel, some placed it on the second, and some placed emphasis inconsistently within and across speaking styles or produced ambiguous patterns with both or neither vowels appearing stressed. FSRA was measured as described in Hedrick and Ohde (1993) and Jongman *et al.* (2000). DFTs (ensemble averaged as described above) were taken of one 23.3 ms Hamming window centered on the fricative midpoint, and one beginning at the onset of the following vowel. For sibilants the peak in the region corresponding to F3 of the frication noise was compared to the peak of the vowel onset in the same

frequency region; for nonsibilants the peak at F5 was used. The relative amplitude was then expressed as the difference (dB) between fricative and vowel amplitudes. HNR was obtained by taking the amplitude difference (dB) between the periodic part of the fricative, estimated using a cross-correlation algorithm [Boersma (1993), as implemented in PRAAT], and the remaining (noise) part. An intensity below 500 Hz was obtained similarly to normalized amplitude, except that the VCV was first low-pass filtered at 500 Hz.

E. Statistical analysis

As discussed in Sec. II D, most acoustic measures were considered at several separate time points across fricatives. This was considered important in general because dynamic patterns and not absolute values seem to drive human perception of speech sounds and also because it seems possible that specific clear speech modifications might disproportionally affect different regions of the sounds or might be dynamic in nature. However, based on previous research, we were only able to make specific hypotheses regarding overall style/fricative differences for each of the 14 acoustic measures and not on time-dependent patterns. For this reason, statistical analyses considered only a single value for each measure. For the ten measures that were observed at five time points, this value was the mean of the measurements for the central three (25%, 50%, and 75%) window locations; this tended to reduce error further by time-averaging measurements over the more stable portion of the fricative. For F2, the value was the mean formant transition distance toward the fricative [i.e., ((onset$-$V1 midpoint)+(offset$-$V2 midpoint))/2].

For each metric then, a mixed-model analysis of variance (ANOVA) with speaking style (clear versus conversational), fricative as a within-subject factor, and gender as a between-subject factor was used to address hypothesis 1 (that clear fricatives would be longer, louder, and higher in frequency content) and hypothesis 2 (that inventory-level and contrast-dependent fricative-to-fricative distances in the 14-dimensional acoustic space would be larger in clear speech).

Two additional analyses addressed hypothesis (2) more directly. First, distances between each of the 16 targeted fricative pairs (pairs differing in place or voicing) were calculated for the 14 acoustic measures and were compared depending on whether the sounds were produced in (1) a clear fricative-to-fricative contrastive context, (2) a clear but noncontrastive context, or (3) conversationally. For example, /s/-/ʃ/ distances were considered (1) between productions of /s/ that were produced specifically in response to a "misidentification" as /ʃ/ / (we will represent this with the notation s|ʃ) and /ʃ/ productions produced after identification as /s/ (ʃ|s), (2) between clear /s/ productions that were produced in response to misidentifications of sounds other than /ʃ/ (represented s|~ʃ) and ʃ|~s productions, and (3) initial conversational productions of the two sounds (s|ø, ʃ|ø). As described above, it was predicted that distances would be generally greater in clear than conversational tokens and greatest in

contrastive contexts. A one-way ANOVA with style as a within-subject factor was used to compare the distances, averaged across the 16 targeted pairs.

Second, a linear discriminant analysis was used to measure whether fricatives were actually better separated along the measures that were considered. For each place or voicing pair, a set of 14 predictors was constructed, each consisting of 120 (20 speakers $\times$ 2 fricatives $\times$ 3 styles) possible training points. For each style (contrastive, noncontrastive, and conversational), a jack-knife verification method was used, in which the classification was run separately for each speaker and style (with the two relevant points for the speaker used as test data and the remaining 118 points as training data), and results were averaged within style conditions.

Finally, hypothesis 3 (which speakers would differ in type and/or extent of acoustic modifications) was addressed using a two-way mixed-model ANOVA with style as a within-subject factor and talker as a between-subject factor. Analyses made use of the R statistical package (v. 2.0.4) (Venables and Ripley, 2002; Balakrishnama and Ganapathiraju, 1998).

III. RESULTS AND DISCUSSION

All 20 participants seemed to have followed the instructions regarding speaking style and, in particular, were able to produce truly "conversational" tokens throughout the experiment despite the laboratory setting and the frequency of recognition errors. This was verified both informally by the first author during the experiment and by acoustically comparing clear and conversational tokens from late in the experimental session, with the samples produced during warmup and earlier in the experiment. For example, fricative duration (usually considered a robust indicator of speaking style) was compared with the sequential order of productions in the experiment (1–440) using Pearson's correlation. For clear fricatives, a small but reliable positive relationship was found ($r=0.095$, $p<0.001$), revealing a tendency for longer clear productions as the experiment progressed. For conversational productions, on the other hand, a small *negative* relationship was seen ($r=-0.090$, $p<0.001$), indicating that conversational productions became slightly *shorter* over the course of the experiment, in complying with the instructions (and possibly resulting from boredom or impatience) and despite the frequency of recognition errors.

Since the productions in response to recognition errors are being referred to as clear, it is also important to verify that they are actually more intelligible for human listeners and not just produced with greater effort. To date, we have observed significant intelligibility advantages for the clear over the conversational tokens discussed here for young normal-hearing listeners, listeners with simulated hearing impairment, and non-native listeners (Maniwa, 2006; Maniwa *et al.*, 2008; Kabak and Maniwa, 2007). By measuring babble thresholds for the same minimal pair distinctions targeted in the elicitation method described here (i.e., place and voicing contrasts), we have verified that each fricative category is more intelligible on average in clear speech. For

TABLE I. Summary of ANOVA results (see Sec. III E) for the style (S) ×fricative (F) ×Gender (G) comparison (left), the one-way comparison of mean fricative-to-fricative distance across styles (center), and the speaker (Sp) ×style comparison (right). F values are given; values in bold are statistically significant after FDR alpha correction.

	G	F	F×G	S	S×G	F×S	F×S×G	S (distance)	Sp	Sp×S
(df)	(1, 18)	(7, 126)	(7, 126)	(1, 18)	(1, 18)	(7, 126)	(7, 126)	(2, 56)	(19, 133)	(19, 133)
peak	0.224	**50.6**	**3.19**	**16.6**	1.72	**7.45**	0.878	3.77	**2.53**	1.75
M1	**12.3**	**128**	**4.77**	**65.6**	1.87	**11.6**	2.41	2.01	**4.25**	1.03
M2	0.169	**82.5**	**2.86**	**41.3**	1.38	**11.8**	1.35	2.11	**2.19**	0.927
M3	0.0212	**85.1**	**2.55**	**40.9**	0.704	**8.54**	0.673	1.32	**1.96**	1.27
M4	0.676	**29**	0.815	**20.1**	0.0016	**14.7**	0.879	3.33	1.58	1.91
FSRA	0.66	**40.5**	1.88	**11.3**	0.0827	**5.5**	0.269	4.21	**8.15**	**3.99**
Slope below	1.44	**269**	0.995	**27.2**	0.532	**3.77**	0.964	0.0392	**2.93**	**4.79**
Slope above	1.01	**234**	1.52	2.12	0.681	**5.07**	0.798	2.46	**6.49**	2.14
Duration	0.0188	**100**	1.35	**57.2**	0.0465	**18.4**	0.715	**12.3**	**88.6**	**92.2**
f0	**82.8**	**12.4**	**7.03**	**9.14**	3.26	1.31	2.06	1.23	**357**	**5.69**
amp	0.126	**107**	1.15	6.03	0.0014	**6.95**	0.533	**6.72**	**12**	**4.54**
amp500	1.73	**53.5**	1.68	**30.4**	0.0922	**14.8**	0.247	**13.7**	9.4	**7.08**
F2	5.83	**52.8**	1.82	**23.7**	0.0545	6.51	0.204	4.45	**10.3**	**4.61**
HNR	6.41	**88.2**	**6.69**	3.83	1.64	**11.7**	0.889	2.02	1.52	1.32

some populations (Maniwa *et al.*, 2008; also see Sec. III C), we have also been able to relate intelligibility differences to average acoustic modifications at the speaker level. While additional study is needed to determine, for example, the relative effects of different overall, inventory-level, and local contrast-enhancing strategies on intelligibility, it thus seems reasonable to refer to the present data as clear speech.

Figures 1–11 show the results of the analyses described in Sec. II D across fricatives, styles, and (where relevant) measurement locations. Table I summarizes the results of the analyses of variance described in Sec. II E. In the following sections, we describe in some detail the significance of these results with respect to our three hypotheses. A more comprehensive descriptive analysis of the data can be found in Maniwa (2006).

A. Overall clear speech modifications (hypothesis 1)

The leftmost columns in Table I summarize the results of the style × fricative × gender ANOVA. Because 14 separate analyses were conducted across measures that were in several cases highly correlated, a critical alpha level of 0.0073, based on the false discovery rate (FDR) estimate for a 5% false positive rate for the style main effect, was adopted.

As shown in Figs. 1–11, clear and conversational fricatives differed systematically along nearly every dimension we considered. Across speakers and fricatives, duration increased (Fig. 11, 187 ms longer in clear speech), and spectral measures including peak frequency (Fig. 1, on average 818 Hz higher in clear speech), mean (Fig. 2, 668 Hz higher in clear speech), and skewness (Fig. 4, 0.96 lower in clear speech) showed energy concentration in higher frequency re-
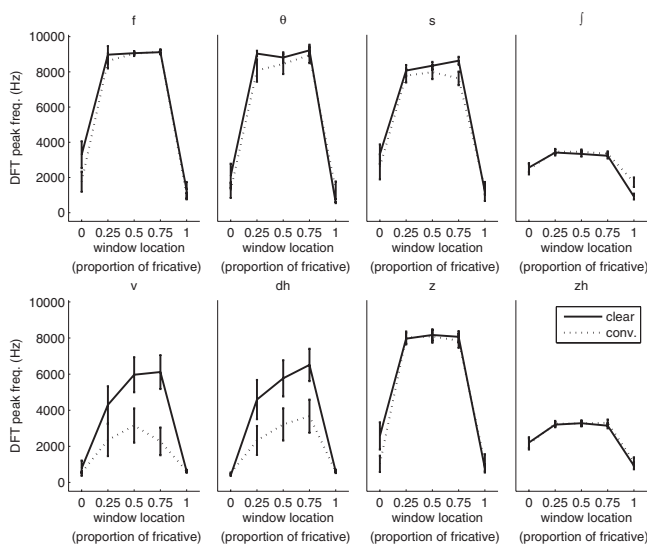

FIG. 1. Mean (and standard error) DFT peak frequency for each fricative as a function of speaking style. The horizontal axis indicates the location of the analysis window, ranging from 0 (fricative onset) to 1 (fricative offset). dh refers to the voiced interdental fricative and zh to the voiced palato-alveolar.
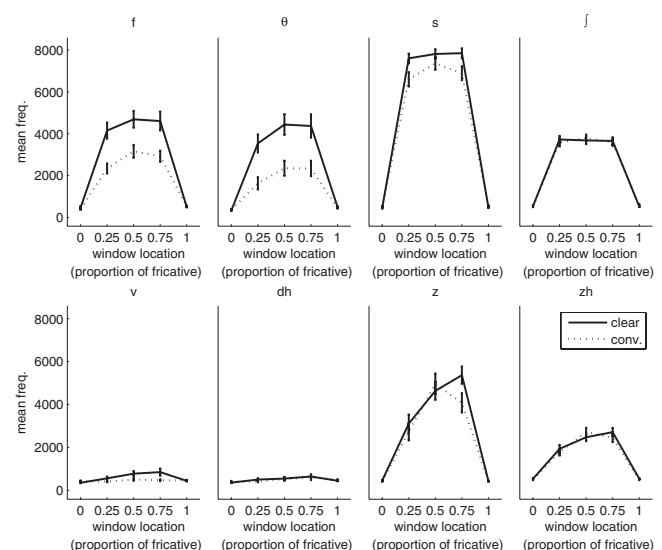

FIG. 2. Mean (and standard error) moment 1 values (center of gravity) for each fricative as a function of speaking style.
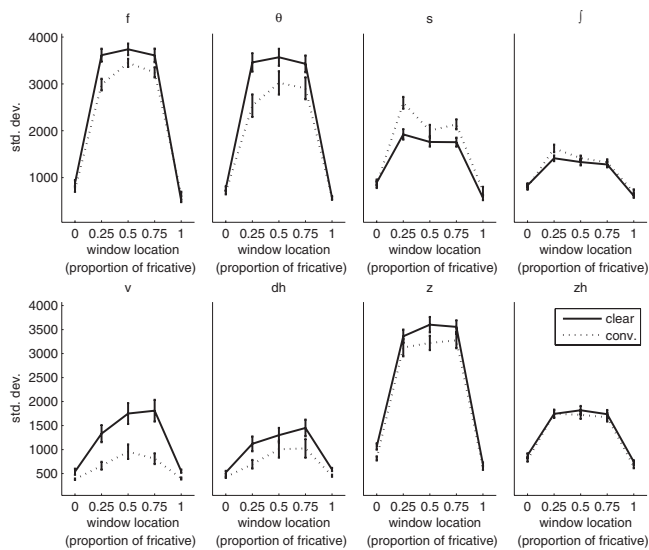
FIG. 3. Mean (and standard error) moment 2 values (SD) for each fricative as a function of speaking style.

FIG. 5. Mean (and standard error) moment 4 values (kurtosis) for each fricative as a function of speaking style.

gions in clear speech. In general, these effects were seen most clearly at central windows, where there was less variability relating to neighboring vowels or the fricative constriction release. F2 transitions also covered greater frequency ranges in clear speech (Fig. 6, 85 Hz difference on average). Steeper spectral slopes (Fig. 7, on average 1.8 dB/kHz steeper) below the peak frequency also suggest more defined peaks and greater noise source strength for clear speech, consistent with previous reports on fricatives produced with elevated vocal effort (Jesus and Shadle, 2002). Slopes above peak frequencies were more variable, with sibilants (with better defined peaks and steeper slopes overall) showing larger negative values (steeper slopes) but nonsibilants (with near-zero slopes overall) showing, on average, *smaller* values in clear compared to conversational speech. The averaged neighboring vowel f0 was also higher in clear speech (Fig. 11, 4.72 Hz higher on average). These results are in general agreement with previous studies [e.g.,
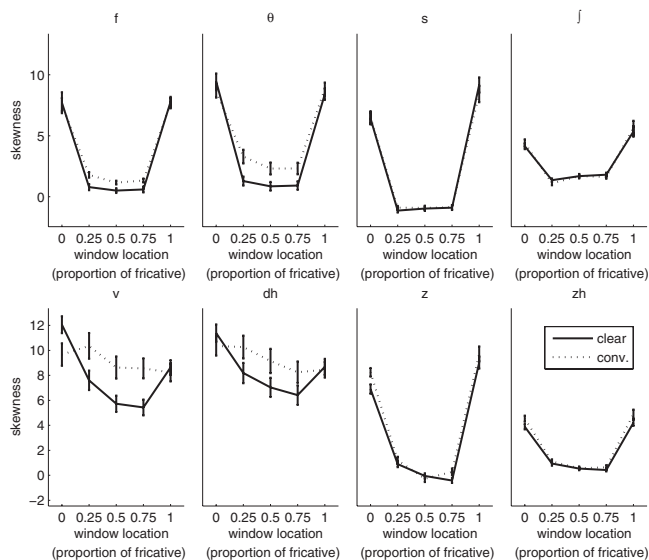
see Chen (1980) and Picheny *et al.* (1986)] and are consistent with increased vocal effort in clear speech.

On the other hand, a main effect of style for rms amplitude was found at all five locations, with clear fricatives significantly *lower* in amplitude (Fig. 8, on average 1.08 dB lower in clear speech; that it was higher in W5 is due to the onset of the following vowel). FSRA also decreased in clear speech (Fig. 10, 4.8 dB lower). Clear fricatives also had significantly less energy below 500 Hz (Fig. 9, on average 4.18 dB lower). Lower amplitude measures compared to neighboring vowels were somewhat unexpected considering reports of increased CVR in clear speech [e.g., see Bradlow *et al.* (2003) and 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 nonsibilants) for some speakers (Picheny *et al.*, 1986; Krause and Braida, 2004). The present

FIG. 4. Mean (and standard error) moment 3 values (skewness) for each fricative as a function of speaking style.

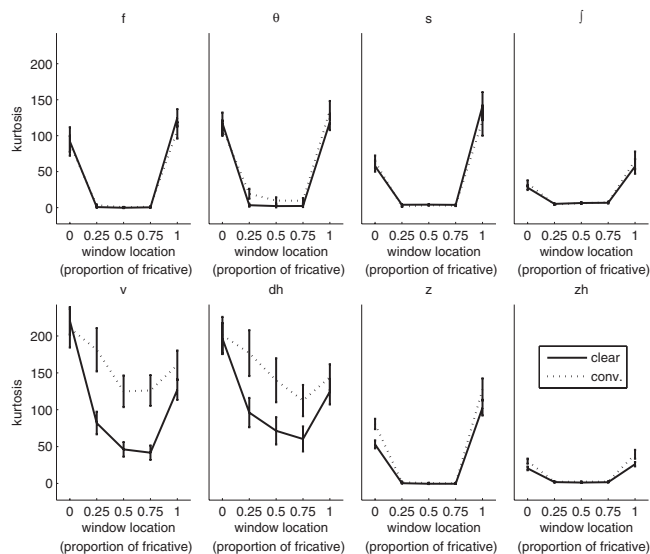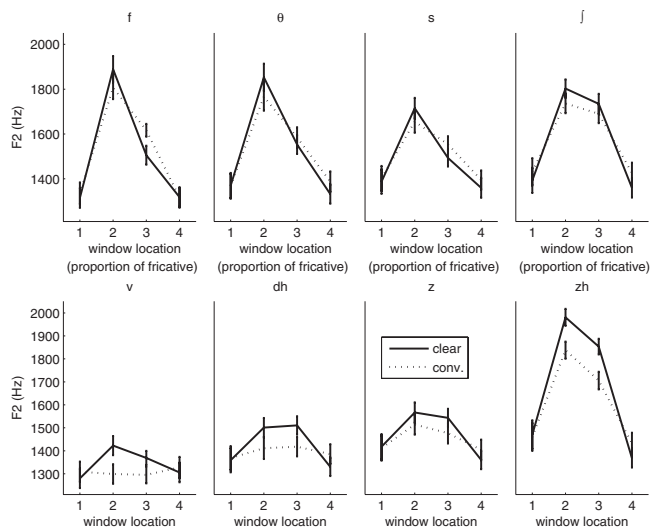FIG. 6. Mean (and standard error) F2 values (Hz) for each fricative at four window locations (W1=midpoint of the preceding vowel, W2=vowel-fricative onset, W3=fricative-vowel offset, and W4=midpoint of the following vowel) as a function of style.
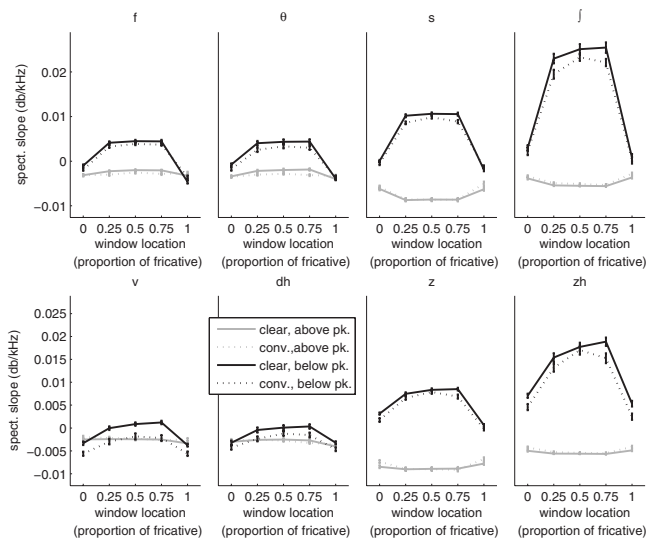
FIG. 7. Mean (and standard error) slope values below and above the peak frequencies for each fricative as a function of speaking style.

FIG. 9. Mean (and standard error) frequency-specific relative amplitude (FSRA) values as a function of fricative and style.

results are probably best explained in terms of articulatory effort. Since the volume velocity required to increase the level of fricative sounds—particularly nonsibilants—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 nonsibilants).

Thus, hypothesis 1 was clearly confirmed; robust overall changes were seen in the durations, spectra, and probably amplitude of clear fricatives that are consistent with increased vocal effort.

B. Inventory-level and local contrastive patterns (hypothesis 2)

1. Overall tendencies

Style × fricative interactions for several measures were consistent with efforts to maintain contrasts within the Eng-
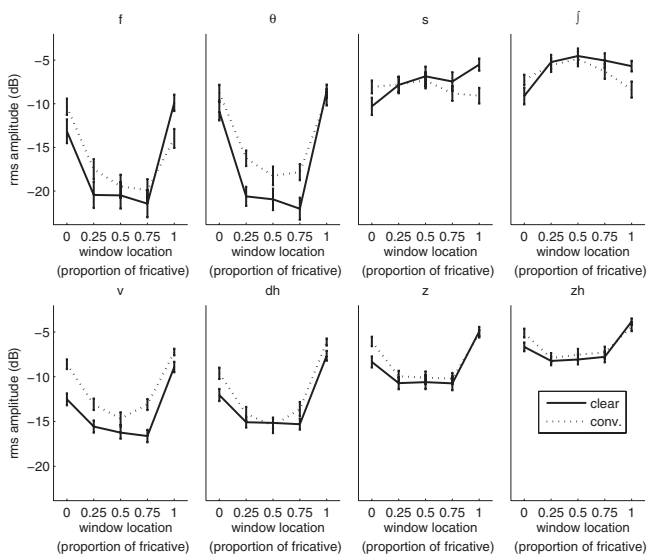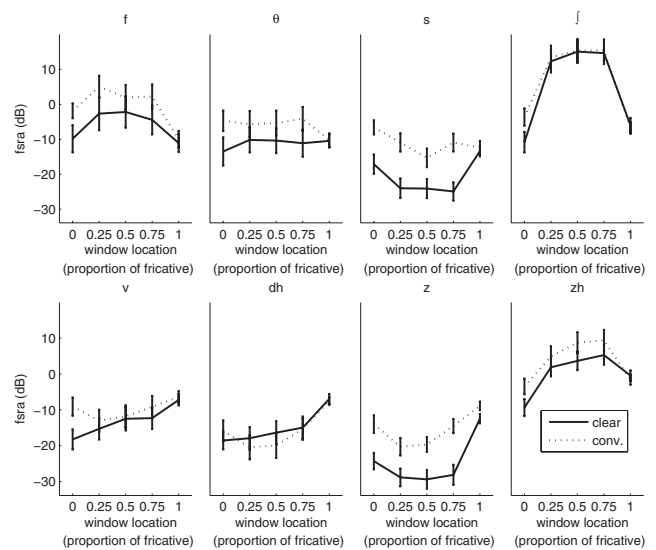
lish fricative inventory. The place of articulation contrasts, in particular, seemed to be enhanced in clear speech. For example, palato-alveolars are defined by energy concentration at low frequencies; DFT peaks and M1 for palato-alveolars increased much less than for other fricatives (even decreasing in some cases) in clear speech, and skewness generally decreased less (increasing in some cases) for other places of articulation. Differences between sibilants and nonsibilants were also emphasized in clear speech. Nonsibilants, with inherently more diffuse spectra, showed increases in M2 in clear speech, while sibilants decreased (+573 Hz versus −49 Hz). Nonsibilants also decreased in kurtosis in clear speech, whereas voiceless sibilants did not. Acoustic distances between sibilants and nonsibilants also increased in terms of amplitude (see comments in Sec. III A regarding the overall vocal effort for a complementary account of the differences); a significant decrease in the normalized rms amplitude in clear speech was seen only for nonsibilant fricatives; voiceless sibilants actually increased slightly. The F2

FIG. 8. Mean (and standard error) normalized rms amplitude for each fricative as a function of speaking style.
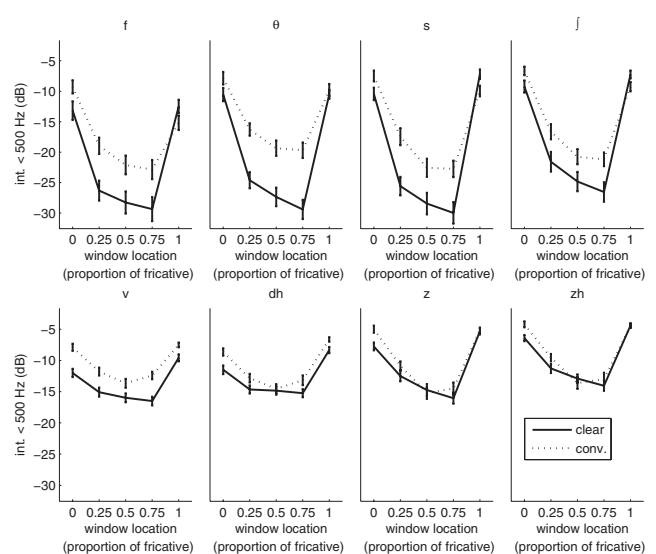

FIG. 10. Mean (and standard error) normalized intensity below 500 Hz for each fricative as a function of speaking style.
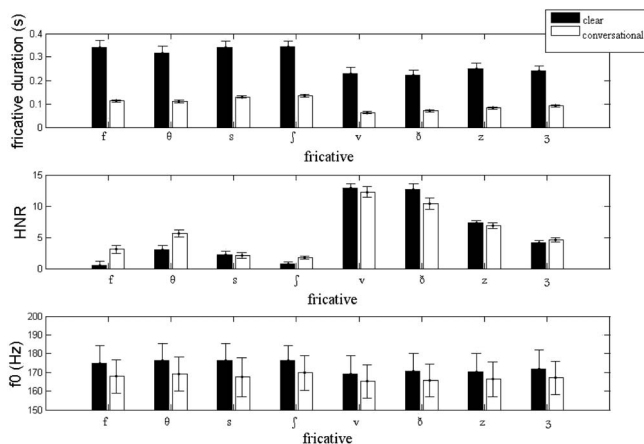
FIG. 11. Mean (and standard error) noise duration as a function of fricative and style (top), HNR averaged across speakers as a function of fricative and style (middle), and f0 values as a function of fricative and style (bottom).

transition distance increased more for palato-alveolars (with inherently higher F2) than alveolars, and dentals (with inherently higher F2) than labio-dentals in clear speech, increasing fricative-to-fricative distance in both cases. This is particularly important for the nonsibilants, for which F2 may be a critical cue (Jongman et al., 2000; Nittrouer, 2002).

Enhanced voicing contrasts were also seen. A style × fricative interaction for duration revealed that inherently longer voiceless fricatives increased more in length than voiced fricatives in clear speech (213 ms versus 159 ms), increasing the distance between the two classes of sounds in terms of duration. The style × fricative interaction was also seen for M2, showing greater increases for voiced than voiceless fricatives (and decreases for voiceless sibilants) in clear speech. These results are in accordance with those from Jongman et al. (2000), which showed that voiced fricatives had a significantly greater variance than voiceless ones and similarly increased the average M2 distance between voiced and voiceless sounds in clear speech. A larger decrease (5.24 dB larger) in intensity below 500 Hz for voiced fricatives, an increase in HNR for voiced fricatives (+0.48), and a decrease for voiceless (−0.73) and an increase in f0 in adjacent vowels only for voiceless fricatives (3.24 Hz larger increase for voiceless fricatives) were also consistent with efforts to maintain voicing contrasts in clear speech.

Thus, while it cannot be shown that these changes were a direct result of knowledge of the fricative inventory and its critical contrasts and while the actual results of the changes 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., see Bradlow et al. (2003), Chen (1980), Krause and Braida (2004), Ohala (1994), Picheny et al. (1986), Smiljanić and Bradlow (2005)]. These findings support the first (inventory-level contrast) part of hypothesis 2.

2. Distance comparisons

Differences between similar pairs of sounds and the acoustic characteristics of fricative productions, in general, were influenced not only by the speaking style overall but by

the specific misidentifications that prompted clear productions as well. An examination of clear productions as a function of misidentification seemed to suggest that context-dependent contrastive efforts (that is, attempts to make sounds more unlike the sounds they had been mistaken for) were responsible for some of the effects that were seen. For example, when speakers repeated the sound /ʃ/ in response to a misidentification of the sound as /s/ (/ʃ|s/), they produced the fricative with significantly lower DFT peak frequencies than when they produced the same sound in response to a misidentification as /ʒ/ (/ʃ|ʒ/; 3356 Hz versus 3504 Hz). This suggests that speakers tried to differentiate the sound /ʃ/ from neighboring sounds in clear speech since it has a typical peak frequency between /ʒ/ and /s/. Similarly, M1 was lower, and M3 higher, in /ʃ|s/ compared to /ʃ|ʒ/ productions.

The one-way ANOVA comparing fricative-to-fricative distances for each measure in contrastive (e.g., /ʃ|s/), noncontrastive (/ʃ|~s/), and conversational (/ʃ|∅/) contexts was designed to quantify these differences, as well as the inventory-level distance-enhancing manipulations discussed in Sec. III B. We first considered the mean distance across acoustic dimensions (after normalizing all measures to have a SD of 1.0 so that measures were weighted equally). A significant effect of style (F) $F(2, 56) = 4.02$; $p = 0.023$) revealed precisely what we predicted: distances were largest for contrastive productions (0.952 SD units), followed by noncontrastive clear productions (0.948), and smallest for conversational tokens (0.939). Results considering the 14 measures separately are summarized in the center columns of Table I. For nearly every dimension, the predicted order (contrastive > noncontrastive > conversational) was observed, although it reached significance only for duration, amplitude, and low-frequency amplitude ($\alpha = 0.0024$, based on the FDR analysis of observed p values). The robustness of the distance enhancement for these measures may be related to the fact that duration varied so much with style (Fig. 10) and that the amplitude measures were relevant to (and therefore may have been adjusted to emphasize) both place and voicing distinctions.

In summary, the comparison of acoustic distances between fricative pairs across measures and misidentification prompts revealed that speakers tended to repeat sounds such that they differed maximally from neighboring sounds and especially from those for which they were initially mistaken. This demonstrates the range of levels at which talkers are sensitive to the communicative demands of a speaking situation and is consistent with the notion that talkers are able to adjust the details of productions based on relatively local fine-grained information (hypothesis 2).

3. Discriminant analysis

Although fricative-to-fricative distances tended to be enhanced in clear speech and by local contrastive efforts, this does not necessarily mean that the speech manipulations introduced in these contexts actually made fricatives easier to identify. For example, increased variability in clear speech could have made the productions of individual speakers more confusable with one another even though mean values for each measure were further apart. The discriminant analy-

sis described in Sec. II E was designed to address this issue directly. Although there was some variability across styles from pair to pair, performance on average was, in accordance with overall distance measures, as predicted: noncontrastive clear productions were more discriminable than conversational ones (94.8% versus 93.9%), and classification was best in contrastive contexts (95.2%). In particular, difficult pairs such as /f/-/θ/ improved substantially in clear styles (0.65 conversation vs. 0.8 clear). Thus, again, consistent with hypothesis 2, clear and contrastive fricatives were more distinct from similar sounds when specific effort was made to reduce confusions.

C. Talker and gender effects (hypothesis 3)

Results of the style×talker ANOVA are shown in the rightmost columns of Table I. Significant talker effects (α =0.014) were seen for every measure except for kurtosis and HNR, and the style×talker interaction was seen for 8 of the 14 measures. This indicates that talkers varied significantly in the magnitude—and sometimes the direction—of acoustic modifications in clear speech. Spectral peak frequency, amplitude, slope above the peak, and FSRA, in particular, showed speaker variability in the direction of clear speech modifications, with SDs of (clear-minus-conversational) differences greater than the respective mean values. In short, hypothesis 3 was supported; talkers differed in their production strategies when they attempted to increase intelligibility; some increased duration more, while others shifted energy distributions toward higher frequency regions more or amplified frication noise relative to the neighboring vowels. Extensive intelligibility experiments will be necessary to determine exactly which of these combinations were most successful at enhancing fricative contrasts. Intelligibility results thus far (Maniwa *et al.*, 2008) seem to suggest that at least for normal-hearing native listeners, the greatest benefits were seen for speakers whose productions involved a relatively large increase in energy at higher frequencies (increased peak, M1, etc.).

One variable that did not seem to contribute to clear-to-conversational acoustic differences was speaker gender. No style×gender or fricative×style×gender interaction was observed for any of the measures (see Table I). This indicates that female and male speakers did not reliably differ in the extent or direction of any acoustic modifications in clear speech. This was somewhat unexpected considering previous reports that female speakers modified their speech to a greater extent than males [e.g., see Bradlow *et al.* (2003) and Liu *et al.* (2004)]. However, since these earlier studies considered a limited number of speakers [e.g., *n*=2 for Bradlow *et al.* (2003)], it was not clear whether the differences observed derived from gender differences or simply talker differences.

D. Dynamic patterns

Figures 1–10 show spectral and amplitude measures over the course of the fricative and not just at one point where the clearest prediction regarding style-related differences could be made. As discussed in Sec. II E, these data

were included partly because it was considered possible that differences in different measures might be more prominent at different locations or might be dynamic in nature. To the extent that this possibility can be addressed with the present data, it seems for the most part not to have been the case. Contours representing measures for clear and conversational tokens appear to be roughly parallel over the three central windows, with differences that were generally in the expected directions and that sometimes narrowed or changed direction at fricative-vowel boundaries.

IV. CONCLUSIONS

In sum, 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 and energy at higher frequencies), and some were more surprising (especially lower relative amplitude). Across a variety of measures, the acoustic distances between minimally contrasting sounds were 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 specific listener feedback, adjusting repeated productions to be more unlike sounds for which they had been misapprehended. Individual talkers varied widely in the magnitude—and sometimes the direction—of these changes; these differences were not related to talker gender.

Balakrishnama, S., and Ganapathiraju, A. (**1998**). "Linear discriminant analysis: A brief tutorial," available at http://www.zemris.fer.hr/predmeti/kdisc/bojana/Tutorial-LDA-Balakrishnama.pdf (Last viewed May 24, 2007).

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. Acoust. 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.

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, Vol. **17**, pp. 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 at http://www.fon.hum.uva.nl/praat/ (Last viewed May 24, 2007).

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.

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

Childers, D. G. (**1978**). *Modern Spectrum Analysis* (IEEE, New York), pp. 252–255.

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

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, R. (**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. (**2002**). "Vowels in clear and conversational speech: Talker differences in acoustic features and intelligibility for normal-hearing listeners," Ph.D. 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.

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

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

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

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.

Hughes, G. W., and Halle, M. (**1956**). "Spectral properties of fricative consonants," J. Acoust. Soc. Am. **28**, 303–310.

Iverson, P., and Bradlow, A. R. (**2002**). "The recognition of clear speech by adult cochlear implant users," ICSA Workshop*Temporal Integration in the Perception of Speech*, Aix-en Provence, France, 8–10 April.

Jackson, P. J. B., and Shadle, C. H. (**2000**). "Frication noise modulated by voicing, as revealed by pitch-scaled decomposition," J. Acoust. Soc. Am. **108**, 1421–1434.

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., 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. Acoust. Soc. Am. **85**, 1718–1725.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Kabak, B., and Maniwa, K. (**2007**). "L2 perception of English fricatives in clear and conversational speech: The role of phonetic similarity and L1 interference," ICPhS XVI, Saabrücken.

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

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, Dordrecht), pp. 403–439.

Lindblom, B. (**1996**). "Role of articulation in speech perception: Clues from production," J. Acoust. 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. Acoust. Soc. Am. **116**, 2373–2383.

Maniwa, K. (**2006**). "Acoustical and perceptual properties of clearly produced fricatives," Ph.D. dissertation, University of Kansas, Lawrence, KS.

Maniwa, K., Jongman, A., and Wade, T. (**2006**). "Acoustic characteristics of clearly produced fricatives," J. Acoust. Soc. Am. **119**, 3301.

Maniwa, K., Jongman, A., and Wade, T. (**2008**). "Perception of clear English fricatives by normal-hearing and simulated hearing-impaired listeners," J. Acoust. Soc. Am. **123**, 1114–1125.

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.

McGowan, R., and Nittrouer, S. (**1988**). "Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/," J. Acoust. Soc. Am. **83**, 229–236.

Miller, G. A., and Nicely, P. A. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

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

Nissen, S. L., and Fox, R. A. (**2005**). "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," J. Acoust. Soc. Am. **118**, 2570–2578.

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

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

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.

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

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. Acoust. Soc. Am. **95**, 1581–1592.

Perkell, J. S., Matthies, M. L., 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.

Pincas, J., and Jackson, P. J. B. (**2006**). "Amplitude modulation of turbulence noise by voicing in fricatives," J. Acoust. Soc. Am. **120**, 3966–3977.

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

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.

Silbert, N., and de Jong, K. (**2008**). "Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production," J. Acoust. Soc. Am. **123**, pp. 2769–2779.

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

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

Stevens, K. N. (**1971**). "Airflow and turbulence for noise for fricative and stop consonants: Static consideration," J. Acoust. 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 (Academic, New York), pp. 243–255.

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

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.

Venables, W. N., and Ripley, B. D. (**2002**). *Modern Applied Statistics*, edited by S. Fourth (Springer, New York).

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

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

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