Published in final edited form as:
Speech Commun. 2012 January 1; 54(1): 147–160. doi:10.1016/j.specom.2011.07.008.

On the development of a frequency-lowering system that enhances place-of-articulation perception

Ying-Yee Kong[a),b),*] and **Ala Mullangi**[b)]

[a)]Department of Speech Language Pathology & Audiology, 106A Forsyth Building, Northeastern University, Boston, MA 02115, USA

[b)]Bioengineering Program, Northeastern University, Boston, MA 02115, USA

Abstract

Frequency lowering is a form of signal processing designed to deliver high-frequency speech cues to the residual hearing region of a listener with a high-frequency hearing loss. While this processing technique has been shown to improve the intelligibility of fricative and affricate consonants, perception of place of articulation has remained a challenge for hearing-impaired listeners, especially when the bandwidth of the speech signal is reduced during the frequency-lowering processing. This paper describes a modified vocoder-based frequency-lowering system similar to one reported by Posen, Reed, and Braida (1993), with the goal of improving place-of-articulation perception by enhancing the spectral differences of fricative consonants. In this system, frequency lowering is conditional; it suppresses the processing whenever the high-frequency portion (>400 Hz) of the speech signal is a periodic signal. In addition, the system separates non-sonorant consonants into three classes based on the spectral information (slope and peak location) of fricative consonants. Results from a group of normal-hearing listeners with our modified system show improved perception of frication and affrication features, as well as place-of-articulation distinction, without degrading the perception of nasals and semivowels compared to low-pass filtering and Posen et al.'s system.

Keywords

frequency lowering; speech perception; place of articulation; fricatives

1. Introduction

The term "frequency lowering" refers to the presentation of high-frequency speech information to the lower-frequency region. This form of processing is intended to improve speech intelligibility in individuals who have severe-to-profound high-frequency hearing loss but have usable hearing in the lower frequencies. Prior to the availability of the cochlear implant, a medical device surgically implanted in the cochlea to restore auditory function, the target clinical populations for frequency-lowering schemes were individuals with residual hearing in a very restricted low-frequency region (<1000 Hz). Recently, frequency

[*]Corresponding author: Department of Speech Language Pathology & Audiology, 106A Forsyth Building, Northeastern University, Boston, MA 02115, USA; Tel: +1 (617) 373-3704, Fax: +1 (617) 373-2239, yykong@neu.edu.

lowering has regained interest due to (1) findings on the importance of extended bandwidth of amplification on the development of speech and language in hearing-impaired children (Stelmachowicz et al., 2004); (2) the increasing number of cochlear implantees with residual low-frequency acoustic hearing; and (3) the lack of improvement in speech recognition for some individuals with high-frequency hearing loss greater than 50 dBHL or with dead region(s), after increasing the audibility of high-frequency speech information. (e.g., Hogan and Turner, 1998; Baer et al., 2002). There was also evidence suggesting that individuals with less severe hearing loss at high frequencies could benefit from frequency lowering (Wolfe et al., 2010; 2011).

Methods of frequency lowering for speech have been available since the 1930s (Dudley, 1939). The types of signal processing methods employed in frequency-lowering schemes include channel vocoding (e.g., Lippmann, 1980), slow playback (e.g., Beasley et al., 1976), frequency transposition (e.g., Velmans, 1974), and frequency shift with linear (e.g., Turner and Hurtig, 1999) or nonlinear frequency compression (e.g., Reed et al., 1983). Comprehensive reviews of each processing method are provided by Braida et al. (1979) and Simpson (2009). A Brief description of the commonly-used approaches – frequency transposition and nonlinear frequency compression, and a more detailed description of the channel-vocoding scheme (Posen et al., 1993) that is highly related to the present work are provided in the following paragraphs.

In the frequency transposition approach, high-frequency acoustic signals are shifted to the lower-frequency region and then the transposed signal is added to an original unprocessed low-frequency signal. As pointed out by Kuk et al. (2009) and Simpson (2009), the advantages of this approach include: minimization of the artifacts from frequency lowering such as pitch shift, preservation of frequency ratios of the high-frequency signal in the transposed sound, and preservation of natural sound quality. The main disadvantage is that the transposed signal which is superimposed onto the original unprocessed speech could potentially mask the useful low-frequency speech cues. In the nonlinear frequency compression approach, the higher-frequency signals are shifted to the lower-frequency region by reducing the bandwidth of the original speech signal in a manner in which the amount of frequency lowering is greater at higher frequencies compared to lower frequencies. As pointed out by Simpson (2009), the major advantages of this approach are that there is no overlap between the shifted high-frequency and lower-frequency signals, which overcomes the risk of masking, and that the low- and mid-frequency information is preserved if frequency compression is only applied to high frequencies. The main disadvantage is that the frequency ratios of the high-frequency signal are not preserved, which could potentially have a negative effect on sound quality as well as speech recognition, especially when frequency compression is extended to lower frequencies.

Outcomes with the transposition and frequency shifting methods are generally positive; most of the studies showed a significant improvement for consonant, word, or sentence recognition in at least some hearing-impaired listeners (e.g., Simpson et al., 2005; Glista et al., 2009; Kuk et al., 2009). For consonant recognition, the improvement was mainly found on the perception of consonant classes of fricatives and affricates without compromising recognition of other consonant classes, such as stops, nasals, and semivowels (e.g., Robinson et al., 2007; Glista et al., 2009; Kuk et al., 2009). For individuals with high-frequency hearing loss, recognition of fricative and affricate consonants are particularly challenging because these consonants contain energy mainly at the high-frequency region (above 2000 Hz). Individuals with high-frequency hearing loss also often have great difficulty in place-of-articulation distinction for consonants (e.g., labiodental fricative/f/vs. alveolar fricative/s/vs. palatal fricative/ʃ/). An early report by Harris (1958) demonstrated that adult English-speaking listeners used primarily high-frequency spectral cues to

discriminate between alveolar/s/and palatal/ʃ/fricatives. The frequency of the frication spectral peak is higher (>4000 Hz) for alveolar fricatives than for palatal fricatives (<3000 Hz). This spectral contrast is likely to be reduced after frequency-lowering processing, especially for schemes that contain frequency compression.

Performance on place-of-articulation distinction in severe-to-profound high-frequency hearing-impaired listeners and in normal-hearing listeners tested with simulations of severe-to-profound high-frequency hearing loss usually did not improve by frequency-lowering processing compared to the control condition (e.g., conventional amplification or low-pass filtering) (e.g., Simpson et al., 2006; Robinson et al., 2007; Kuk et al., 2009; Füllgrabe et al., 2010). Simpson et al. (2005) tested a group of 17 hearing-impaired listeners with moderately-sloping hearing losses on phoneme recognition with CNC words. They found that as a group, their subjects did not show phoneme recognition improvement with their nonlinear frequency compression scheme compared to conventional amplification. About half of their subjects (8 out of 17), however, showed improvement for phoneme recognition with their frequency-lowering scheme. Subsequently, they performed an information transmission analysis on the eight subjects who demonstrated benefit with nonlinear frequency compression and reported an improved perception for consonant features of frication and place of articulation. Given that phoneme recognition scores and percent information transmission were obtained in an open-set task (i.e., word recognition) in Simpson et al.'s (2005) study, caution is needed when interpreting these results.

The primary goal of this paper is to describe a new frequency-lowering system which aims to improve the perception of consonant classes of fricatives and affricates, and more importantly, enhance the place-of-articulation distinction for fricative consonants. We chose a vocoder-based frequency-lowering system similar to that described in Posen et al. (1993). In this paper, we will first provide a detailed description of the vocoder-based frequency-lowering system developed by Posen et al. (1993), and discuss the advantages of this system over the frequency transposition and nonlinear frequency compression approaches described above, as well as its limitations. We will then provide a description of the development of our modified frequency-lowering system that aims to improve the Posen et al.'s system by enhancing the differences in spectral characteristics of fricative consonants differing in place of articulation. We will present results from our acoustical analyses on eight fricatives in American English. Results from acoustical analyses were used to determine the acoustic parameters and decision criteria in our system for classification of the transposed signals. The secondary goal is to provide evidence that the enhanced place-of-articulation feature in the frequency-lowered speech can improve speech perception. Particularly, we are interested to know if listeners can utilize the enhanced spectral cues in our modified system to distinguish among fricative consonants that differ in place of articulation. We will evaluate the effectiveness of our system on the classifications of speech sounds, and present perceptual data to demonstrate the benefit of our new system on the perception of consonant features of frication, affrication, and place of articulation.

2. A vocoder-based frequency-lowering system (Posen et al., 1993)

In a vocoder-based system described in Posen et al. (1993), online signal processing was performed in an experimental speech processor. High-frequency speech information was first analyzed by passing speech through a bank of eight contiguous one-third-octave analysis filters with standard center frequencies in the range of 1000 to 5000 Hz. The outputs of the adjacent filters were then combined to form four analysis bands. The output levels of these bands were measured with averaging times of 20 ms, and were then used to determine the output levels of low-frequency narrow-band noise signals. The noise signals were generated by passing wideband noise through four contiguous one-third-octave

synthesis filters whose center frequencies ranged from 397 to 794 Hz. The four high-frequency analysis bands and the four low-frequency synthesis filters were monotonically related in that the lowest analysis band controlled the lowest synthesis band, the second-lowest analysis band controlled the second-lowest synthesis band, and so on. The output level of a noise band was linearly related to the output level of its analysis band. That is, a 1-dB increase in the signal level in an analysis band caused a 1-dB increase in the level of the corresponding low-frequency noise-band signal. In addition, the level of each of the four low-frequency noise bands was attenuated to minimize any masking effect on the original speech signal. The four low-frequency narrow-band noise signals were then summed and added to the original speech signal. A block diagram of Posen et al.'s vocoder-based system is shown in Fig. 1. In this system, the low-frequency noise was added to the original speech signal only if the speech signal was dominated by high frequencies (i.e., fricatives, affricates, and stops). Frequency lowering only occurred when the power in the low-frequency region (summing from filters with center frequency of 125 to 1250 Hz) was less than that in the high-frequency region (summing from filters with center frequency of 1600 to 5000 Hz) plus 3 dB.

There are several advantage in Posen et al.'s system compared to the transposition and nonlinear frequency compression approaches described above. First, Posen et al. (1993) modified a vocoder-based system by Lippmann (1980) to conditionally perform transposition for speech sounds that are dominated by high frequencies (i.e., stop, fricative, and affricate consonants) to reduce the risk of masking when high-frequency information is superimposed onto lower-frequency regions of the original signals. Furthermore, they transposed the high-frequency components (1000–5000 Hz) of the speech sounds to the low-frequency region (400–800 Hz), where there is essentially no energy in the original signal for fricative and affricate consonants (see discussion of spectral characteristics of fricative consonants below), further minimizing the problem of masking due to superimposition. Second, the low- and mid-frequency signals remain unprocessed and the frequency lowering generally does not involve the vowels and sonorant consonants (e.g., nasals and semivowels). Thus, the harmonic structure and frequency ratios between the high-frequency components of the vowels and sonorant consonants are unchanged, preserving the natural sound quality. Third, the high-frequency signal is lowered to a low-frequency region, a region where hearing is normal or near-normal for listeners with severe-to-profound hearing loss at higher frequencies, leading to potentially greater speech recognition benefit compared to other frequency-lowering approaches. This is because spectral resolution in the normal or near-normal hearing region is generally better than that in the hearing-impaired region. For example, in the nonlinear frequency compression approach, compression takes place mainly at high frequencies where hearing loss is greater compared to the lower-frequency region for sloping hearing loss (e.g., Simpson et al., 2005). Similar to nonlinear frequency compression, the major disadvantage of Posen et al.'s system is that the bandwidth is reduced in the transposed signal (center frequencies of analysis filters from 1000 to 5000 Hz; center frequencies of synthesis filters from 397 to 794 Hz) and that the ratios between high-frequency components are not preserved. This could have a negative impact on speech recognition, especially when listeners use primarily spectral cues to perceive place-of-articulation distinction for fricative consonants (Harris, 1958). To alleviate the detrimental effect of frequency compression in Posen et al.'s system, the primary aim of our modified vocoder-based frequency-lowering system (see description below) is to enhance the spectral differences of the transposed signals for fricatives differing in place of articulation. The second aim of our modified system is to find a conditional frequency-lowering rule that could separate the non-sonorant sounds (i.e., stops, fricatives, and affricates) from other sonorant sounds (i.e., vowels, nasals, and semivowels) with a higher level of accuracy compared to the rule that compared the high-frequency and low-frequency energy used in Posen et al. (1993).

3. System Development: A modified vocoder-based frequency-lowering system with place-of-articulation feature enhancement

Our modified system made two modifications to Posen et al.'s system. In our system, the frequency lowering underwent two processing stages. Stage 1 involved a decision rule that determined the consonants with aperiodic high-frequency energy (a conditional frequency-lowering rule). Stage 2 involved decision rules that classified high-frequency frication sounds into three groups based on the spectral information distinguishing fricative consonants differing in place of articulation. The system then enhanced the spectral differences of non-sonorant sounds, particularly fricatives that differ in place of articulation, based on the classification results. Similar to Posen et al.'s system, signal processing in our system was performed offline using a 20-ms window.

For the development and the evaluation of this modified system, we used 22 consonant stimuli (stops/p, t, k, b, d, g/; fricatives/f, θ, s, ʃ, v, ð, z, ʒ/; affricates/tʃ, dʒ/; nasals/m, n/; and semivowels:/r, l, y, w/) in/VCV/utterances with three vowels (/a, i, u/), resulting in a total of 66 syllables. These stimuli were spoken three times (three repetitions) by each of 12 speakers (five male adults, five female adults, one male child age 11, and one female child age 11), resulting in a total of 2376 tokens. The adult speakers were taken from the recordings in Shannon et al. (1999) and the two child speakers were recorded in our laboratory. All stimuli were scaled to have equal root-mean-square (RMS) amplitude. We divided the stimuli into two sets: design set and test set. Parameters used in our frequency-lowering system were determined based on the acoustic properties in the design set and then verified in the test set to assess the generalization of the chosen parameters to other speakers. The design set contained speech stimuli from three adult males and three adult females. The test set contained the stimuli from the remaining speakers. The choice of the adult speakers for the test set was largely based on their fundamental frequency (F0) and the quality of the original recordings. Perceptual studies were later conducted on human listeners to evaluate our system, thus, speech tokens from the adult speakers that had the best recording quality were chosen for the test set. The test set contained stimuli with a wide range of F0s among the speakers in our entire stimulus set, from the lowest F0 of the male adult speakers to the higher F0s of the female adult and child speakers. Stimuli from the child speakers to the higher F0s of the female adult and child speakers. Stimuli from the child speakers were included in the test set to evaluate the generalization capability of the chosen parameters from adult speakers to children. Acoustical analyses of periodicity at high frequencies and spectral shapes for aperiodic high-frequency signals described below were performed on the design set to determine the acoustic parameters and decision criteria that consistently 1) identify speech signals that would undergo frequency-lowering processing, and 2) divide the high-frequency frication sounds into three groups based on the spectral characteristics of three classes of fricative consonants differing in place of articulation. We then processed the stimuli in the test set using these parameters and criteria. Subsequent analyses were performed on the test set, and results were compared to those obtained from the design set to evaluate the effectiveness of our acoustic parameters and decision criteria for the identification and classification of consonants, and their potential for real-life application.

Conditional frequency-lowering

Similar to Posen et al. (1993), frequency-lowering processing was not automatic for all speech signals in our system. Previous studies showed that unconditional frequency lowering could negatively affect the perception of vowels and semivowels (Lippmann, 1980; Posen et al., 1993). To avoid this detrimental effect, we only applied frequency lowering to consonants that are phonetically classified as non-sonorants (i.e., stop, fricative,

and affricate consonants). Posen et al. (1993) used a conditional rule that determined the predominance of the high-frequency energy. Our modified algorithm used a different conditional frequency-lowering rule which only added low-frequency noise signals to those consonants for which the high-frequency region (>400 Hz) contained aperiodic signals, as determined by the autocorrelation-based pitch-extraction algorithm in PRAAT (Boersma and Weenink, 2009). The rationale for periodicity detection above 400 Hz is because there is a strong periodicity at the low frequencies for voiced non-sonorant sounds, particularly voiced fricatives and affricates. For these voiced consonants, the higher-frequency signals are predominately aperiodic. Using this rule, voiced and voiceless stop, voiced and voiceless fricative, and voiced and voiceless affricate consonants from the design set were correctly identified by the system 95% of the time, and less than 2% of the vowels, nasals, and semivowels were mis-classified as containing aperiodic high-frequency energy. The overall accuracy was 97%. These results are significantly improved compared to the conditional rule used by Posen et al. (1993) whose overall accuracy was about 89%, based on our analysis of our speech samples.

Classification of high-frequency frication sounds

Once detected, consonants with aperioidc high-frequency energy will be further analyzed for their spectral shape. The purpose of Stage 2 analysis was to separate high-frequency frication sounds, determined by our conditional frequency-lowering rule described above, into three groups based on the spectral information that distinguished three groups of fricative consonants, differing in place of articulation. The acoustic features and decision criteria used for classification in our system were based on a series of acoustical analyses performed on fricative consonants in the design set. Description of the acoustical analysis and the results are presented as follows:

Acoustical characteristics of English fricatives have been studied extensively in the past (e.g., Hughes and Halle, 1956; Behrens and Blumstein, 1988a, 1988b; Nittrouer et al., 1989; Jongman et al., 2000; Onaka and Watson, 2000; Ali et al., 2001; Fox and Nissen, 2005; Nissen and Fox, 2005; Maniwa et al., 2009). Many static and dynamic acoustic properties have been identified that could potentially separate fricatives with different places of articulation. For the purpose of developing an algorithm for hearing devices which requires signal processing in real time, we did not consider any dynamic properties that involve analysis between neighboring sounds, such as relative amplitude of the frication noise that takes into account the amplitude of adjacent vowels. Among the static acoustic properties, combinations of spectral features including spectral slope, spectral peak location, and spectral mean (Jongman et al., 2000; Ali et al., 2001; Fox and Nissen, 2005; Nissen and Fox, 2005; Maniwa et al., 2009), were found to be the most robust features to separate fricatives into three groups: labio- and inter-dental/f, θ v, ð/vs. alveolar/s, z/vs. palatal/ʃ, ʒ/. It is noted that previous reports did not find any combinations of static properties that could classify fricatives into four groups with a high degree of accuracy (e.g., Onaka and Watson, 2000; Ali et al., 2001; Fox and Nissen, 2005).

In our acoustical analyses, speech signals were first passed through 20 contiguous one-third-octave filters with standard center frequency in the range of 125 to 10079 Hz. For the purpose of determining the decision criteria that could be used for our frequency-lowering system, analyses were performed in the middle of the frication noise[1], as described in Jongman et al. (2000). Only the middle portion of the frication noise was examined because previous studies have shown that spectral properties are relatively stable throughout the

[1]Acoustical analyses for determining the features and decision criteria for fricative classification were performed using a 40-ms full Hamming window placed in the middle of the frication noise. As pointed out by Jongman et al. (2000), this larger window size yields better resolution in the frequency domain

frication noise (Behrens and Blumstein, 1988a; Jongman et al., 2000). The output level was calculated from each of the 20 analysis bands. Figure 2 (left panel) shows the output levels in dB from each analysis band averaged across tokens for each fricative consonant in our design set. It is clear that fricative consonants (including voiced and voiceless fricatives) have energy primarily in the high-frequency region above 1000 Hz, and essentially no energy between 400 and 1000 Hz. Voiced consonants, additionally, have high energy in the low frequencies that corresponds to the F0 and the low-frequency harmonics. With this observation and results from previous reports (Onaka and Watson, 2000; Ali et al., 2001), our spectral analyses focused in the high-frequency region (> 1000 Hz).

Measures of *spectral slope* were derived from a linear regression line fit to relative output levels in decibels (dB) from each of the analysis bands with a center frequency from 1260 Hz to 5040 Hz. *Spectral peak location* was defined here as the center frequency of the analysis band between 1260 Hz and 10079 Hz that contained the highest amplitude. To avoid the effect of sharp variations in the spectrum, smoothing was performed by averaging the output level of the analysis band with the output levels of its adjacent bands. *Spectral mean* was computed as frequency averaged across analysis bands with center frequencies from 1260 to 10079 Hz, weighted by the output level of its corresponding band, as described in previous studies (e.g., Ali et al., 2001). That is, the output of each analysis band was first multiplied by its corresponding center frequency; the sum of these values across frequency bands was then divided by the sum of the output levels of all frequency bands. Most of our effort was devoted to determining the parameter(s) and the thresholds of these parameters that could separate the fricative consonants into three groups with a high level of accuracy in our design set.

Previously reported results (e.g., Ali et al., 2001), as well as our own analyses, showed that no single parameter was able to separate three groups of fricatives with a high degree of accuracy. Thresholds for each parameter were statistically chosen using histogram analysis. The mean classification score was 59% correct with the spectral slope feature alone, and 67% correct with the spectral peak or spectral mean feature alone. We subsequently used a combination of acoustic parameters for the classifications, minimizing the number of parameters used as much as possible in order to reduce the amount of processing time and power consumption that would occur in a hearing aid circuit. With optimal threshold selections, a combination of spectral slope and spectral peak location or a combination of spectral slope and spectral mean produced similar classification accuracy of about 80%. We decided to use the combination of spectral slope and spectral peak location for fricative classification in our system. For spectral slope computed in a frequency range from 1260 to 5040 Hz, a threshold of 0.003 dB/Hz could reliably distinguish the sibilant fricatives (alveolar/s, z/and palatal/ʃ, ʒ/: slope >0.003) and non-sibilant fricatives (labio- and inter-dental/f, θ, v, ð/: slope <0.003). This criterion produced identification accuracy of 81% correct in our design set. Once classified as sibilants, stimuli underwent a second spectral analysis in the high-frequency region from 1260 to 10079 Hz to determine the spectral peak location to further separate alveolar/s, z/from palatal/ʃ, ʒ/fricatives. A threshold of 6000 Hz produced 92% correct distinction between the alveolar (peak > 6000 Hz) and palatal (peak < 6000 Hz) fricatives in our design set. Using these spectral features and decision criteria the overall percent correct for place-of-articulation fricative classification is 79%.

Enhancing place-of-articulation features

As described above in the Posen et al.'s system, frequency lowering was achieved by adding low-frequency noise signals (four bands of noise signals with center frequencies of 397, 500, 630, and 794 Hz) to the original speech signal. In their system, the high-frequency analysis filters and the low-frequency synthesis filters (center frequencies from 397 to 794 Hz) were monotonically related, and that the output level of a noise band was linearly related to the

output level of its analysis band. Unlike Posen et al.'s method, the spectral shapes of the added low-frequency noise signals for the non-sonorant sounds in our system were determined by the classification that employed the acoustic parameters and decision criteria described above. Three spectral patterns of low-frequency noise signals were used to signal the different groups:

1. For non-sonorant sounds that were classified as Group 1 (based on the spectral properties of labio- and inter-dental fricatives/f, θ, v, ð/), noise signals from the four low-frequency synthesis bands were summed before adding to the original speech. The output levels of eight analysis bands from the original speech with center frequencies from 2000 to 10079 Hz were first analyzed. Similar to Posen et al., the combined outputs of the two adjacent bands were then used to control the levels of the four low-frequency noise bands, and the high-frequency analysis filters and low-frequency synthesis filters (center frequencies from 397 to 794 Hz) were monotonically related. The output levels of the low-frequency synthesis bands were then set 10 dB lower than those at the high-frequency analysis bands.
2. For non-sonorant sounds that were classified as Group 2 (based on the spectral properties of palatal fricatives/ʃ, ʒ/), noise signals from only the lowest two frequency bands (397 and 500 Hz) were used. The noise level in the 397-Hz band corresponded to the output level of the original speech at the peak location (<6000 Hz), and the noise level in the 500-Hz band was set 10 dB lower than that in the 397-Hz band.
3. For non-sonorant sounds that were classified as Group 3 (based on the spectral properties of alveolar fricatives/s, z/), noise signals from the two highest frequency synthesis bands (630 and 794 Hz) were used. The level of the noise in the 794-Hz band corresponded to the output level of the original speech at the peak location (>6000 Hz), and the noise level in the 630-Hz band was set 10 dB lower than that in the 794-Hz band.

Figure 3 shows the noise levels averaged across speakers and tokens in the design set at each of the four low-frequency synthesis bands with center frequencies of 397, 500, 630, and 749 Hz for three groups of frication sounds.

Summary

A block diagram of the processing stages in our vocoder-based frequency-lowering system is shown in Fig. 4. This system consists of two stages of analysis, place-of-articulation enhancement processing, and frequency-lowering processing. The system first separates speech sounds into two classes – sonorants vs. non-sonorants, based on periodicity detection at the frequency region above 400 Hz. Only the non-sonorant sounds will be subjected to further analyses and frequency-lowering processing. The second stage of analysis separates non-sonorant sounds into three groups based on the spectral structure of the sounds at the high-frequency region. Low-frequency noise signals are spectrally shaped to enhance the spectral differences for different groups of frication sounds. Finally, the high-frequency components of non-sonorant speech signals are transposed to the lower-frequency region by the addition of the noise-vocoder output.

4. System Evaluation

The effectiveness of our frequency-lowering system was evaluated to determine (1) the accuracy of sound classification, and (2) the perceptual benefit of consonant identification in human listeners.

4.1. Classification of stimuli in the test set

Classification of speech signals was performed on a separate set of/VCV/stimuli (test set) (Figure 2, right panel). This stimulus set consists of 22 consonants in a/VCV/context with three vowels (/a, i, u/) spoken by six speakers (2 men, 2 women, 1 boy, and 1 girl) and repeated three times by each speaker, resulting in a total of 1188 tokens.

Periodicity detection—The same autocorrelation algorithm in PRAAT was used to determine the presence of periodicity in the speech signal above 400 Hz. Periodicity was detected for sonorant sounds 98% of the time, and the aperiodic signals were identified as non-sonorant consonants with 95% accuracy. This represents an overall accuracy of 97% for the conditional frequency-lowering rule. These results are very similar to those obtained in the design set.

Fricative classification—Spectral slope with a decision criterion of 0.003 dB/Hz was used to separate sibilant and non-sibilant fricatives in our test set. The overall accuracy for sibilant/non-sibilant distinction was 78% correct. Spectral peak location with decision criterion of 6000 Hz was used to separate alveolar and palatal fricatives, and the overall accuracy for alveolar and palatal distinction was 95% correct. The overall accuracy of fricative classification using both the spectral slope and spectral peak location parameters was 82% correct. Again, these results are very similar to those obtained in the design set. Percent correct for non-sibilant fricatives, alveolar sibilant fricatives, and palatal sibilant fricatives were 72%, 83%, and 90%, respectively (see confusion matrix in Table I).

4.2. Perceptual study

The purpose of the perceptual studies was to evaluate the perceptual benefit of our frequency-lowering system. Particularly, we wanted to evaluate the additional consonant identification benefit of the place-of-articulation enhancement method in our system compared to the frequency-lowering system described in Posen et al. (1993). The purpose of frequency lowering is to convey high-frequency speech cues to the lower frequency region that has normal or near-normal hearing for individuals who have severe-to-profound high-frequency hearing loss. For this clinical population, conventional amplification at high frequencies that compensates for loss of audibility has not shown improvement in speech intelligibility (Hogan and Turner, 1998; Baer et al., 2002). Posen et al. (1993) obtained consonant identification results from normal-hearing listeners listening to simulations of severe-to-profound high-frequency hearing loss above 800 Hz. Simulation of hearing loss with a low-pass (LP) filtering technique on normal-hearing listeners is commonly used to evaluate the effectiveness of frequency-lowering methods and parameters for hearing devices (e.g., Velmans, 1973; Reed et al., 1983, 1991; Posen et al., 1993; Korhonen and Kuk, 2008; Fullgrabe et al., 2010). More importantly, LP filtering to simulate the loss of audibility at high frequencies provides a good model for speech recognition in listeners with severe-to-profound high-frequency hearing loss, but with normal or near-normal hearing at low and mid frequencies. McDermott and Dean (2000) tested a group of hearing-impaired listeners with steeply-sloping hearing loss above 500-1500 Hz, but had normal or near-normal hearing at low-frequencies on CNC word recognition in steady-state speech-shaped noise. In one experiment, they compared the results obtained from these listeners to those from normal-hearing listeners presented with LP filtered speech reported in Henry et al. (1998). They concluded that their hearing-impaired subjects “obtained about the same information from the speech signal with a signal-to-noise ratio (SNR) of 6 dB as normally hearing subjects listening to speech filtered in a similar way.” (p.356). In a second experiment, McDermott and Dean (2000) tested a group of five normal-hearing listeners and a subset of hearing-impaired listeners as in their first experiment on CNC word recognition with speech stimuli presented with or without the phase-vocoder-based frequency

transposition described in Moore (1990). Normal-hearing listeners were presented with LP filtered speech at 1200 Hz to simulate the degree of steeply-sloping high-frequency hearing loss in the hearing-impaired group. They reported that patterns of results were similar between the hearing-impaired and normal-hearing groups; that is, both groups showed lack of benefit from frequency transposition. These results further strengthen the validity of using simulations of high-frequency hearing loss on normal-hearing listeners to evaluate the potential benefit of our modified frequency-lowering system.

4.2.1. Identification of fricative consonants—This study was designed to evaluate the perceptual benefit of our place-of-articulation feature-enhancement method in frequency-lowered fricative consonants.

A. Methods

Subject: The subjects were five female normal-hearing listeners in their early to mid 20s. They all had hearing thresholds of no more than 20 dB HL at audiometric frequencies from 250 to 8000 Hz and were native speakers of American English with standard dialect.

Stimuli: Eight fricative consonants/f, θ, s, ʃ, v, ð, z, ʒ/from the test set described above were used. The first two repetitions spoken by each of the speakers were used in the experiment, resulting in a total of 288 tokens. Each token underwent either simple LP filtering with a cutoff frequency of 800 Hz and attenuation rate of 50 dB/octave to simulate steeply-sloping high-frequency hearing loss, or one of three different types of frequency-lowering processing followed by LP filtering with the same filter parameters to simulate high-frequency hearing loss: (1) Posen et al.'s channel-vocoder processing: frequency-lowering system identical to Posen et al. (1993), in which the high-frequency analysis filters (1000–5000 Hz) and the low-frequency synthesis filters (400–800 Hz) were monotonically related; (2) Feature enhancement: our modified system that acoustically enhanced the place-of-articulation feature – i.e., using three different patterns of low-frequency noise signals to encode three different groups of fricatives which differed in place of articulation; (3) Extended bandwidth (BW): frequency-lowering system similar to the Posen et al.'s system, except that the frequency range of the high-frequency analysis filters expanded from 1000–5000 Hz to 1000–10079 Hz, and each of the analysis filters were half of an octave wide. This condition was created to evaluate if the benefit we saw with our modified system compared to that proposed by Posen et al. (1993) was indeed due to our feature-enhancement method or just due to the extension of the analysis frequency from 5000 to 10079 Hz. Given that the purpose of this experiment was to evaluate the extent to which listeners could use the enhanced spectral separation information at the low frequencies to distinguish different places of articulation, all fricative consonants were processed with an assumption of perfect (100% accuracy) periodicity detection (conditional rule) and perfect separation of different fricative groups (place-of-articulation separation rules) in our modified system. Performance in this condition represents the upper bound of perceptual ability with the feature-enhancement method in our frequency-lowering system. Results from this ideal condition will provide further information in regard to future system development (see discussion below). To facilitate comparisons across processing conditions, all fricative consonants were also processed with an assumption of perfect detection of fricatives for the Posen et al. and extended BW processing conditions; that is, all fricatives underwent the frequency-lowering processing.

Procedures: Stimuli were divided into two sets – a familiarization set and a test set. One repetition spoken by each speaker was used in the familiarization set and another repetition in the test set, thus each set consisted of a total of 144 tokens. The subjects were tested on the identification of frequency-lowered speech with different processing methods and simple

LP filtered speech. The feature-enhancement processing condition was tested first, followed by the Posen et al.'s processing and the simple LP filtering condition, and lastly the extended BW processing condition. This fixed order of presentation was used to rule out the possibility of a learning effect if our feature-enhancement method or Posen et al.'s processing yielded greater fricative identification performance compared to the LP filtering and/or extended BW processing condition. For each processing condition, each subject first practiced with blocks of 144 trials until performance appeared to level off, defined as performance within 3 percentage points in three consecutive practice blocks, up to a maximum of 10 practice blocks. After practice, each subject was tested with three blocks of 144 testing trials. Stimuli within each block were presented in random order. A list of eight fricative consonants – "f, th (for/θ/), s, sh (for/ʃ/), v, dh (for/ð/, z, zh (for/ʒ)" was displayed on a computer screen and subjects responded by clicking a button corresponding to the fricative consonant they heard. During practice, subjects were given visual feedback first to indicate a correct/incorrect response for each trial, immediately followed by auditory feedback if the subject gave an incorrect response. During auditory feedback, a pair of stimuli that included the target stimulus and the stimulus corresponding to the incorrect response, was played twice for comparison. No feedback was provided during the test session. Three of the subjects were presented with processed stimuli to the left ear and the remaining two received the stimuli in the right ear. All stimuli were presented at an averaged RMS level of 72 dBA via headphones (Sennheiser HD 265), a comfortable listening level for band-limited speech (<1000 Hz) for normal-hearing listeners.

B. Results: Overall percent correct score and percent information transmission for the features of voicing and place were computed for individual subjects and for the group data (see group data in Fig. 5). The individual data was computed from confusion matrices combined across different runs for each subject, and the group data was computed from averaging the scores across subjects. A repeated-measures Analysis-of-Variance (ANOVA) showed a significant processing method main effect on the overall fricative identification performance [$F(3,12) = 55.2$, $p < 0.001$]. Three planned pairwise comparisons with Bonferroni adjustment ($p < 0.05/3$) were performed to examine the benefit of the feature-enhancement method over LP filtering, Posen et al.'s processing, and extended BW processing. Overall fricative identification was improved with the feature-enhancement method (76%) compared to the LP filtering (45%) [$t(4) = 52.1$, $p < 0.0001$], Posen et al.'s processing (57%) [$t(4) = 7.4$, $p < 0.005$], and extended BW processing (62%) [$t(4) = 4.6$, $p < 0.01$] conditions. In addition, two repeated-measures ANOVAs were performed for the processing method main effect on voicing and place-of-articulation features. As a group, while performance on voicing distinction (93–94%) was similar [$F(3,12) = 0.9$, $p > 0.05$] among the four types of processing, place-of-articulation distinction was significantly different across processing conditions [$F(3,12) = 91.2$, $p < 0.001$]. Again, three planned pairwise comparisons with Bonferroni adjustment ($p < 0.05/3$) were performed to examine the benefit of the feature-enhancement method over LP filtering, Posen et al.'s processing, and extended BW processing for the perception of place of articulation. Place-of-articulation distinction was significantly better for the feature-enhancement method (76%), compared to the other three processing methods (LP: 15% [$t(4) = 29.6$, $p < 0.001$]; Posen et al.: 39% [$t(4) = 11.6$, $p < 0.0001$]; extended BW: 44% [$t(4) = 7.9$, $p < 0.005$]), with the simple LP filtering producing the lowest place score. The superior place-of-articulation perception performance with the feature-enhancement method compared to the extended-BW method (by 32 percentage points) suggests that listeners were able to utilize the enhanced spectral difference among fricative consonants to improve performance.

Given these positive results in the ideal case (i.e, 100% correct periodicity detection and separation of fricatives), we further tested subjects' performance on fricative identification in a more realistic situation after the stimuli underwent the conditional frequency-lowering

decision rule in Stage 1 and place-of-articulation separation rule in Stage 2 (i.e., <100% accuracy). The same group of subjects participated for this study. Results on the group data showed a slight but significant decrease [$t(4) = 4.7$, $p < 0.01$] in overall performance (70% correct) compared to the results with ideal detection and separation (76% correct). While the voicing score was the same (93%) for both ideal and realistic cases [$t(4) = 0.2$, $p > 0.05$], the place score was 58% for the realistic case, 18 points lower than the ideal case [$t(4) = 7.8$, $p < 0.005$], but still remarkably higher than the LP [$t(4) = 17.5$, $p < 0.001$], Posen et al.'s processing [$t(4) = 6.1$, $p < 0.005$], and extended BW processing [$t(4) = 4.8$, $p < 0.01$] conditions by 43, 19, and 14 percentage points, respectively (Bonferroni adjusted $p < 0.05/4$ for four planned comparisons). A careful examination of the error patterns revealed that, in the realistic case, subjects' response errors corresponded to mis-classification during the processing in determining the peak location – i.e., confusion between/s, z/and/ʃ, ʒ/. For example, when the system mis-classified/s/as/ʃ/, subjects often mis-identified the target as/ʃ/. This suggests that subjects were using the spectral cues to distinguish between alveolar and palatal fricatives. However, when the mis-classification was made in determining the spectral slope – i.e., confusion between sibilant and non-sibilant fricatives, subjects' performance was generally unaffected by this mistake. For example, when the system mis-classified/f/as/s/, subjects could correctly identify the target as/f/. The differences in results between the ideal and realistic conditions, as well as the error patterns in the realistic condition indicate that further development of our system should focus on improving periodicity detection for separating sonorant and non-sonorant sounds, as well as finding better acoustic parameters and decision criteria that could increase the accuracy of fricative classification, particularly, alveolar and palatal fricative separation.

4.2.2. Consonant identification—The purpose of this experiment was to evaluate the perceptual benefit of our frequency-lowering system on consonant identification. Particularly, we examined the effect of our frequency-lowering and place-of-articulation enhancement method on the perception of different classes of consonants and different consonant features.

A. Methods

Subject: Four of the five subjects in the fricative identification experiment also participated in this experiment.

Stimuli: Twenty-two consonants in a/VCV/context with the vowel/a/from the test set were used. The stimuli underwent three types of signal processing: (1) LP filtering with 800-Hz cutoff frequency and 50 dB/octave attenuation (same filtering parameters as in the fricative identification task), (2) Posen et al.'s frequency-lowering system followed by LP filtering (800-Hz cutoff, 50 dB/octave rolloff), and (3) our modified frequency-lowering system (including both Stage 1 and Stage 2 processing) followed by LP filtering (800-Hz cutoff, 50 dB/octave rolloff).

Procedures: One repetition spoken by each speaker was used for practice and the other two repetitions were used in the test set. Each subject received a minimum of 5 practice blocks until performance appeared to level off, up to a maximum of 10 training blocks. After practice, each subject was tested with 10 blocks of 264 test trials. Two subjects were tested with stimuli processed with the modified system first, followed by LP stimuli and then the Posen et al.'s system. The remaining two subjects had the reverse order. Stimuli within each block were presented in random order. A list of 22 consonants was displayed on a computer screen and subjects responded by clicking a button corresponding to the consonant they heard. During practice, subjects were first given visual feedback for each trial, immediately followed by auditory feedback if the subject gave an incorrect response (see fricative

identification experiment for details). Subjects were given a break every 50 trials. No feedback was provided during the test sessions. All stimuli were presented at an averaged RMS level of 72 dBA via Sennheiser HD 265 headphones.

B. Results: A repeated-measures ANOVA showed a significant processing method main effect on overall consonant identification [$F(2,6) = 76.4$, $p < 0.005$]. Three planned pairwise comparisons with Bonferroni correction ($p < 0.05/3$) were performed to compare performance between the Posinet al.'s system and LP filtering, the modified system and LP filtering, and the modified system and the Posen et al.'s system for each of the performance measures (i.e., overall % correct scores, percent information transmission for different consonant features). All subjects showed significantly higher overall consonant scores with the modified system than with the simple LP filtering [$t(3) = 12.0$, $p < 0.005$] and with the Posen et al.'s system [$t(3) = 8.7$, $p < 0.005$]. Across subjects, the overall percentage of consonants identified correctly using the modified system was 72%, an increase of 6 percentage points over LP filtering and an increase percent correct scores and scores for various groupings of consonants (semivowels-nasals, stops-fricatives-affricates) for the group data. As predicted, the perception of nasals and semivowels under the modified and Posen et al.'s systems was similar to that observed under the LP filtering (about 78% correct for the three processing conditions) ($p > 0.05$), presumably attributed to the selective processing achieved by the conditional frequency-lowering rules. On the other hand, overall performance for the stop-fricative-affricate consonants improved by 9 and 6 percentage points with the modified system compared to the LP filtering [$t(3) = 18.3$, $p < 0.001$] and Posen et al.'s system [$t(3) = 19.1$, $p < 0.001$], respectively. Information transmission analysis (Miller and Nicely, 1955) was performed on the group data for seven features: voicing, nasality, frication, affrication, vocalic, duration, and place (Fig. 6 right panel). The perception of each feature related to nasals and semivowels (voicing, nasality, and vocalic) under the modified system and Posen et al.'s system was as good as that observed with LP filtering ($p > 0.05$). The modified system showed a significant improvement compared to LP filtering in the amount of information transferred for frication (10 points) [$t(3) = 8.5$, $p < 0.005$], duration (27 points) [$t(3) = 19.4$, $p < 0.001$], and affrication (5 points) [$t(3) = 10.1$, $p < 0.005$], features that related to fricative and affricate consonants. In addition, the modified system achieved a substantial gain in the amount of information transferred for the place feature by as much as 10 and 7 percentage points compared to LP filtering [$t(3) = 11.3$, $p < 0.005$] and to the Posen et al.'s system [$t(3) = 9.4$, $p < 0.005$], respectively. It is noted that patterns of results with the Posen et al.'s system in this study were similar to those reported in Posen et al. (1993), in which the Posen et al's provided significant benefit for frication [$t(3) = 6.5$, $p < 0.01$] and duration [$t(3) = 4.9$, $p < 0.0167$] features compared to LP filtering.

5. General Discussion and Conclusions

5.1. Classification of speech sounds

Acoustic parameters for the classification of fricative consonants have been investigated previously (Jongman et al., 2000; Ali et al., 2001; Fox and Nissen, 2005; Nissen and Fox, 2005; Maniwa et al., 2009). From the automatic speech recognition standpoint, Ali et al. (2001) identified a combination of five features (five processing steps) that produced 91% accuracy for place-of-articulation classification for fricative consonants, better than of 82% reported in this study. However, some of the features employed in Ali et al.'s algorithm are considered to be dynamic parameters, which we decided not to consider due to our limitations regarding real-time application. Among the parameters used in Ali et al., the maximum normalized spectral slope (MNSS) parameter took into account the maximum output level of neighboring sounds (i.e., the relative amplitude). MNSS was defined as "the maximum value (over the whole spectrum at a certain instant) of the smoothed differences

between neighboring filters, normalized with respect to the maximum energy of the utterance." (p. 2225). As pointed out in Ali et al, classification performance dropped significantly when the MNSS parameter was excluded from the analysis.

Our study distinguished three classes of fricative consonants using only two static parameters. The results were encouraging considering that we achieved a high level of accuracy for fricative classification. Overall accuracy was similar for fricative classification for both the design (79% correct) and test (82% correct) sets. For classification errors concerning sibilant/non-sibilant distinction in our system, non-sibilant perception in the frequency-lowered speech by our listeners was not affected by it. The high accuracy (95% correct) for alveolar and palatal discrimination contributed to the improved place-of-articulation perception.

Our speech samples included consonants produced under different vowel contexts and by 12 different speakers including adults and children from both genders. Ali et al. (2001) used a rather large speech sample from 60 speakers, but all of them were adults. Further research is needed to include speech samples from more speakers and children from different age groups, as well as speech samples from more naturalistic settings (e.g., continuous speech).

5.2. Perceptual benefit with frequency-lowered speech

Our system is the first to artificially enhance the spectral distinction among different classes of fricative consonants in frequency-lowered speech. We have shown positive results with our modified frequency-lowering system on normal-hearing listeners with simulated severe-to-profound high-frequency hearing loss above 800 Hz. Both Posen et al.'s system and our modified system improved the perception of consonant classes of fricatives and affricates compared to LP filtering. More importantly, our modified system, which incorporated decision rules for conditional frequency-lowering and separation of place of articulation for fricatives, also performed significantly better than the Posen et al.'s system for place-of-articulation perception among fricatives. Compared with LP filtering, our modified system performed equally well in handling nasals and semivowels and did not affect the perception of voicing information. The superior performance with the modified system over the extended BW condition in the fricative identification task indicates that the improved place-of-articulation distinction was largely due to the artificial feature enhancement in our modified system, not from the effect of the extended BW of the analysis filters in the signal processing.

In our system, high-frequency frication sounds, including stops, fricatives, and affricates underwent feature-enhancement processing and frequency lowering. The feature-enhancement processing was based on the classification of fricatives differing in place of articulation. Under this processing, the long-duration aspiration of voiceless stops – bilabial/p/, alveolar/t/, and velar/k/stops were classified as Group 1 (i.e., the high-frequency analysis filters and low-frequency synthesis filters were monotonically related) 74%, 57%, and 67% of the time, respectively. Although the stop consonants were involved in the place-of-articulation enhancement processing, overall percent correct identification for stop consonants was not significantly different [$t(3) = 3.0$, $p > 0.05$] between our modified system (67%) and LP filtering (69%). Given that percent information transferred for the stop feature was not 100%, estimations for place of articulation for only the stop consonants could not be performed. We then calculated the overall percent correct identification scores for bilabial, alveolar, and velar stop consonants separately for each subject and compared the results among the three processing conditions. A repeated measures ANOVA showed a non-significant difference ($p > 0.05$) for the processing condition main effect for each group (bilabial, alveolar, and velar) of stop consonants. Pairwise comparisons between the LP filtering and our modified system, and between Posen et al.'s system and our system were

performed to gain further insight on how our feature-enhancement processing affected the perception of place-of-articulation for each group of stop consonants. For bilabial stops, overall identification with LP filtering was six percentage points higher [$t(3) = 5.7$, $p < 0.05$] (86% correct) than our modified system (80% correct). However, the overall identification for bilabial stops was not significantly different [$t(3) = 2.5$, $p > 0.05$] between our modified system (80% correct) and the Posen et al.'s system (86% correct). Taken together the classification results for bilabial stops reported above (74% of bilabial stops were classified as Group 1), these findings suggest that the difference in performance between LP filtering and our modified system may not be attributed to our feature-enhancement method. For alveolar stops, the overall identification performance was not significantly different ($p > 0.05$) between LP filtering (60% correct) and our modified system (59% correct), and between our modified system and the Posen et al.'s system (57% correct). For velar stops, the overall identification performance was also not significantly different ($p > 0.05$) between LP filtering (61% correct) and our modified system (63% correct), and between our modified system and Posen et al.'s system (56% correct). These results suggest that the effect of our feature-enhancement method on place-of-articulation perception for stop consonants is minimal. Listeners could have placed greater perceptual weight on other robust cues (e.g., voice-onset time and formant transition) than on the spectral contrast in the bursts or aspirations for place-of-articulation perception for stop consonants.

5.3. Future direction

Future work will be concerned with (1) exploring other parameters in our frequency-lowering system, including the region and the bandwidth of the low-frequency synthesis bands for individuals with different degrees and slopes of hearing loss (Robinson et al., 2007; Füllgrabe et al., 2010); and (2) improving our classification accuracy for speech sounds by investigating other classification techniques and acoustic features, particularly for speech recognition in noise. Although the acoustic features and the decision criteria chosen for conditional frequency lowering and classification of fricatives produced high levels of accuracy in quiet, we have yet to evaluate how these features and decision criteria perform with the presence of competing noise. Classification of speech sounds in noise is challenging and we expect that overall accuracy would be affected in the presence of everyday noise (e.g., speech of one or more competing talkers). For the detection of non-sonorant sounds, determined by the aperiodic signals at high frequencies, we would expect fewer non-sonorant sounds would be identified correctly when the non-sonorant sounds of the target speech are mixed with sonorant sounds of masker speech because the sonorant sounds (e.g., vowels) generally have greater energy compared to non-sonorant sounds (e.g., non-sibilant fricatives and voiced stops). As Robinson et al. (2007) pointed out in their frequency-lowering system, also using a conditional frequency-lowering rule which compared the power in the high-frequency region to the power in the low-frequency region, the decrease in non-sonorant sound detection in noise in our system will cause the frequency-lowering processing to"fail gracefully", by simply becoming less active when in the presence of competing talkers. As for the classification of fricative consonants, we are currently investigating state-of-the-art feature-extraction and classification techniques that could reliably categorize different classes of speech sounds (e.g., stop vs. fricative consonants, difference in place of articulation for fricative and stop consonants) in quiet and in noise.

The effectiveness of our modified system for speech recognition will also be evaluated in hearing-impaired listeners with moderate to profound high-frequency hearing loss, and cochlear-implant users with low-frequency residual hearing. McDermott and Henshall (2010) tested a group of cochlear-implant users who used a cochlear implant on one side and a hearing aid in the contralateral ear (bimodal stimulation) on speech recognition tasks and

found that the use of frequency-compression processing in the hearing aid did not result in greater bimodal benefit compared to conventional amplification (i.e., without frequency lowering). Recent work in our laboratory (Kong and Braida, 2011) showed that bimodal cochlear-implant users receive somewhat redundant information (mainly voicing and manner of articulation) for consonant identification from their cochlear implant and conventional hearing aid devices. In this group of listeners, cochlear implants and hearing aids generally transmitted very limited information for the place-of-articulation feature. The advantage obtained from enhancing place-of-articulation perception with our modified frequency-lowering system could potentially provide improvement in speech recognition in bimodal listeners.

Acknowledgments

We would like to thank Professor Louis Braida for helpful suggestions and Ikaro Silva for technical support. We also thank Dr. Qian-Jie Fu for allowing us to use his Matlab programs for performing information transmission analysis. This work was supported by NIH/NIDCD (R03 DC009684-03 and ARRA supplement, PI: YYK).

Abbreviations

LP low-pass

References

Ali AM, Van der Spiegel J, Mueller P. Acoustic-phonetic features for the automatic classification of fricatives. J Acoust Soc Amer. 2001; 109:2217–2235. [PubMed: 11386573]

Baer T, Moore BC, Kluk K. Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. J Acoust Soc Amer. 2002; 112:1133–1144. [PubMed: 12243160]

Behrens S, Blumstein SE. Acoustic characteristics of English voiceless fricatives: a descriptive analysis. J Phonetics. 1988a; 16:295–298.

Behrens S, Blumstein SE. On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. J Acoust Soc Amer. 1988b; 84:861–867. [PubMed: 3183204]

Beasley DS, Mosher NL, Orchik DJ. Use of frequency-shifted/time-compressed speech with hearing-impaired children. Audiology. 1976; 15:395–406. [PubMed: 938345]

Boersma, P.; Weenink, D. Praat: doing phonetics by computer (Version 5.1.05). 2009.

Dudley H. Remaking speech. J Acoust Soc Amer. 1939; 11:165.

Fox RA, Nissen SL. Sex-related acoustic changes in voiceless English fricatives. J Speech, Language, Hearing Res. 2005; 48:753–765.

Füllgrabe C, Baer T, Moore BCJ. Effect of linear and warped spectral transposition on consonant identification by normal-hearing listeners with a simulated dead region. Int J Audiol. 2010; 49:420–433. [PubMed: 20180628]

Glista D, Scollie S, Bagatto M, Seewald R, Parsa V, Johnson A. Evaluation of nonlinear frequency compression: clinical outcomes. Int J Audiol. 2009; 48:632–644. [PubMed: 19504379]

Harris KS. Cues for the discrimination of American English fricatives in spoken syllables. Lang Speech. 1958; 1:1–17.

Henry BA, McDermott HJ, McKay CM, James CJ, Clark GM. A frequency importance function for a new monosyllabic word test. Austral J Audiol. 1998; 20:79–86.

Hogan CA, Turner CW. High-frequency audibility: Benefits for hearing-impaired listeners. J Acoust Soc Amer. 1998; 104:432–441. [PubMed: 9670535]

Hughes GW, Halle M. Spectral properties of fricative consonants. J Acoust Soc Amer. 1956; 28:303–310.

Jongman A, Wayland R, Wong S. Acoustic characteristics of English fricatives. J Acoust Soc Amer. 2000; 108:1252–1263. [PubMed: 11008825]

Kong YY, Braida LD. Cross-frequency integration for consonant and vowel identification in bimodal hearing. J Speech, Language, Hearing Res. 2011; 54:959–980.

Korhonen P, Kuk F. Use of linear frequency transposition in simulated hearing loss. J Amer Acad Audiol. 2008; 19:639–650. [PubMed: 19323355]

Kuk F, Keenan D, Korhonen P, Lau CC. Efficacy of linear frequency transposition on consonant identification in quiet and in noise. J Amer Acad Audiol. 2009; 20:465–479. [PubMed: 19764167]

Lippmann RP. Perception of frequency lowered speech. J Acoust Soc Amer. 1980; 67:S78.

Maniwa K, Jongman A, Wade T. Acoustic characteristics of clearly spoken English fricatives. J Acoust Soc Amer. 2009; 125:3962–3973. [PubMed: 19507978]

McDermott HJ, Dean MR. Speech perception with steeply sloping hearing loss: effects of frequency transposition. Br J Audiol. 2000; 34:353–361. [PubMed: 11201322]

McDermott H, Henshall K. The use of frequency compression by cochlear implant recipients with postoperative acoustic hearing. J Amer Acad Audiol. 2010; 21:380–389. [PubMed: 20701835]

Miller GA, Nicely PE. An analysis of perceptual confusion among some English consonants. J Acoust Soc Amer. 1955; 27:338–352.

Moore, FR. Elements of Computer Music. Engle-wood Cliffs, NJ: Prentice-Hall; 1990.

Nissen SL, Fox RA. Acoustic and spectral characteristics of young children's fricative productions: a developmental perspective. J Acoust Soc Amer. 2005; 118:2570–2578. [PubMed: 16266177]

Nittrouer S, Studdert-Kennedy M, McGowan RS. The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. J Speech Hearing Res. 1989; 32:120–132. [PubMed: 2704187]

Onaka, A.; Watson, CI. Acoustic comparison of child and adult fricatives. 8th Aust. Int. Conf. Speech Sci. & Tech.; 2000. p. 134-139.

Posen MP, Reed CM, Braida LD. Intelligibility of frequency-lowered speech produced by a channel vocoder. J Rehabil Res and Dev. 1993; 30:26–38. [PubMed: 8263827]

Reed CM, Hicks BL, Braida LD, Durlach NI. Discrimination of speech processed by low-pass filtering and pitch-invariant frequency lowering. J Acoust Soc Amer. 1983; 74:409–419.

Robinson JD, Baer T, Moore BC. Using transposition to improve consonant discrimination and detection for listeners with severe high-frequency hearing loss. Int J Audiol. 2007; 46:293–308. [PubMed: 17530514]

Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. J Acoust Soc Amer. 1999; 106:L71–74. [PubMed: 10615713]

Simpson A. Frequency-lowering devices for managing high-frequency hearing loss: A review. Trends Amplif. 2009; 13:87–106. [PubMed: 19447764]

Simpson A, Hersbach AA, McDermott HJ. Improvements in speech perception with an experimental nonlinear frequency compression hearing device. Int J Audiol. 2005; 44:281–292. [PubMed: 16028791]

Simpson A, Hersbach AA, McDermott HJ. Frequency-compression outcomes in listeners with steeply sloping audiograms. Int J Audiol. 2006; 45:619–629. [PubMed: 17118905]

Stelmachowicz PG, Pittman AL, Hoover BM, Lewis DE, Moeller MP. The importance of high-frequency audibility in the speech and language development of children with hearing loss. Arch Otolaryngol Head Neck Surg. 2004; 130:556–562. [PubMed: 15148176]

Turner CW, Hurtig RR. Proportional frequency compression of speech for listeners with sensorineural hearing loss. J Acoust Soc Amer. 1999; 106:877–886. [PubMed: 10462793]

Velmans M. Speech imitation in simulated deafness, using visual cues and 'recoded' auditory information. Lang Speech. 1973; 16:224–236. [PubMed: 4768191]

Velmans M. The design of speech recoding devices for the deaf. Br J Audiol. 1974; 8:1–5.

Wolfe J, John A, Schafer E, Nyffeler M, Boretzki M, Caraway T. Evaluation of nonlinear frequency compression for school-age children with moderate to moderately severe hearing loss. J Am Acad Audiol. 2010; 21:618–628. [PubMed: 21376003]

Wolfe J, John A, Schafer E, Nyffeler M, Boretzki M, Caraway T, Hudson M. Long-term effects of non-linear frequency compression for children with moderate hearing loss. Int J Audiol. 2011; 50:396–404. [PubMed: 21599615]

- We developed a vocoder-based frequency-lowering system for hearing devices.
- Novel features of our system include: conditional lowering and place-of-articulation enhancement.
- Perception of fricative and affricate sounds, and place distinction improved under this system.

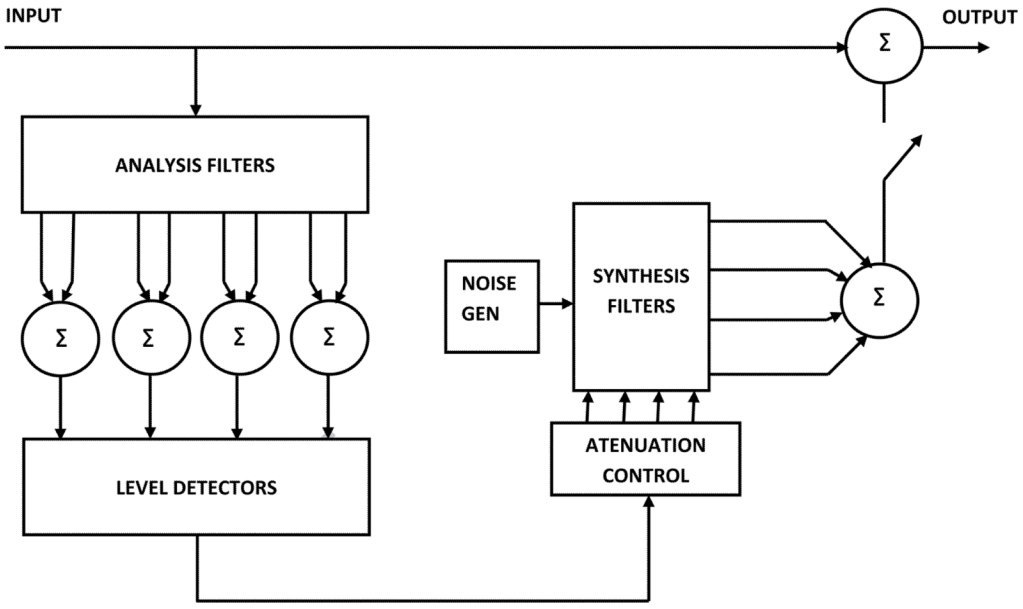

Figure 1.
Block diagram of the vocoder-based frequency-lowering system employed in Posen et al. (1993).

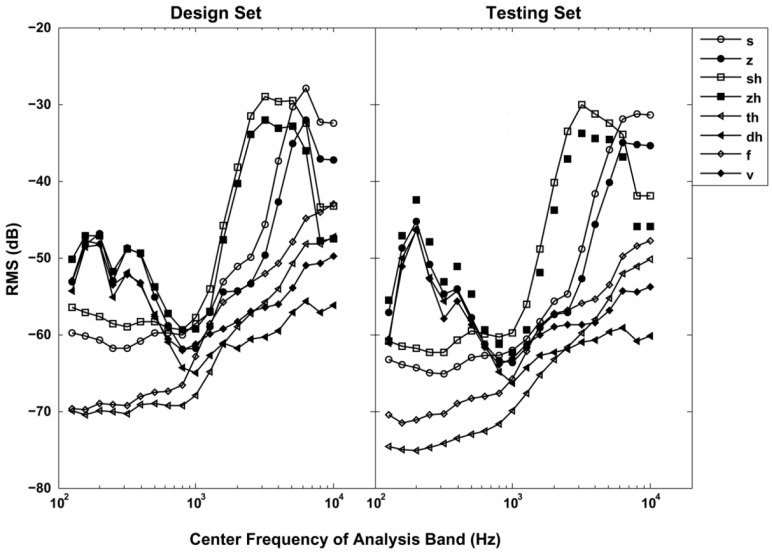

Figure 2.

Average spectra for different fricative consonants in the design set (left panel) and in the testing set (right panel). Output levels were averaged across all tokens and speakers for each fricative. The amplitude is plotted on a dB scale.

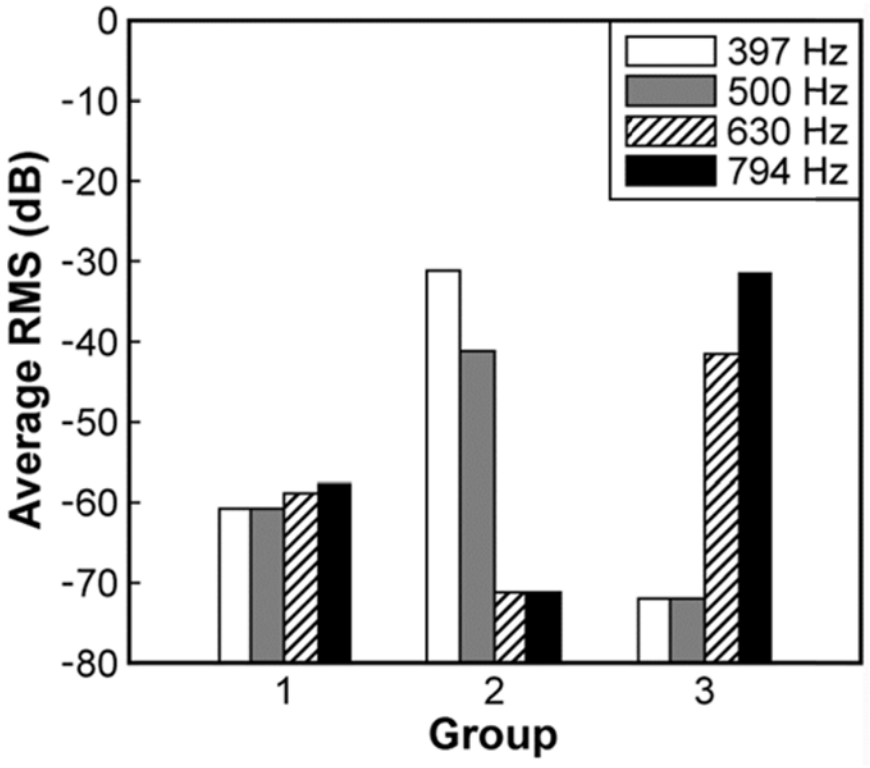

Figure 3.
The noise level at each of the four low-frequency synthesis bands with center frequencies of 397, 500, 630, and 749 Hz for three groups of frication sounds averaged across speakers and tokens in the design set. The amplitude is plotted on a dB scale.

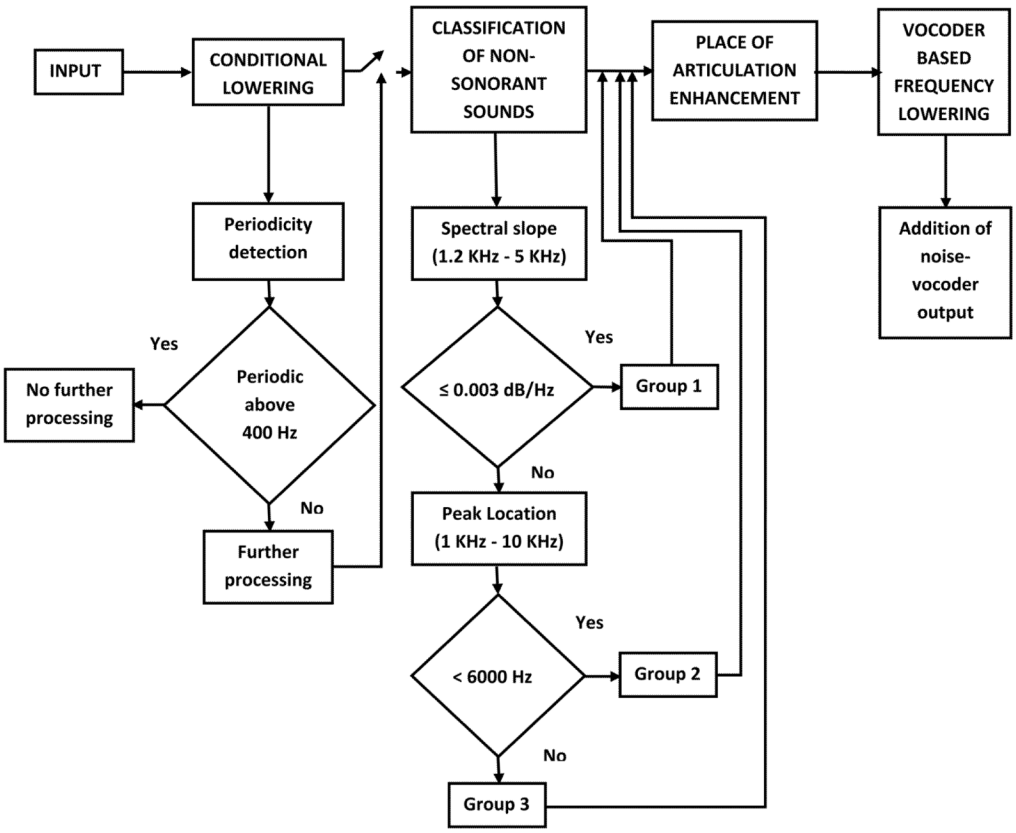

Figure 4.
Block diagram of our modified frequency-lowering system and the consonant classification algorithms.

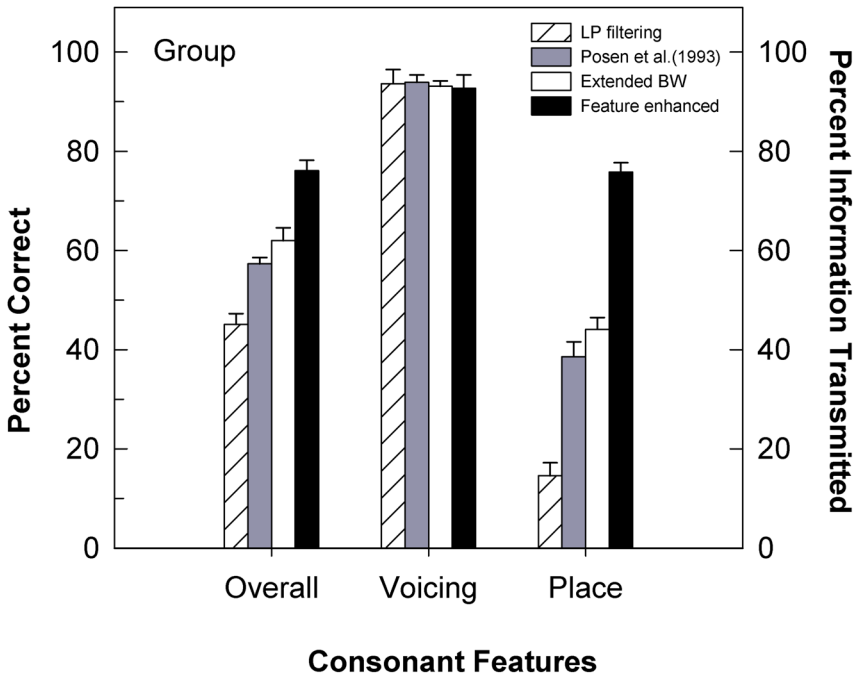

Figure 5.
Average percent correct fricative identification scores and percentage of information transmission on the voicing and place features for the four processing conditions. Error bars represent standard error around the mean.

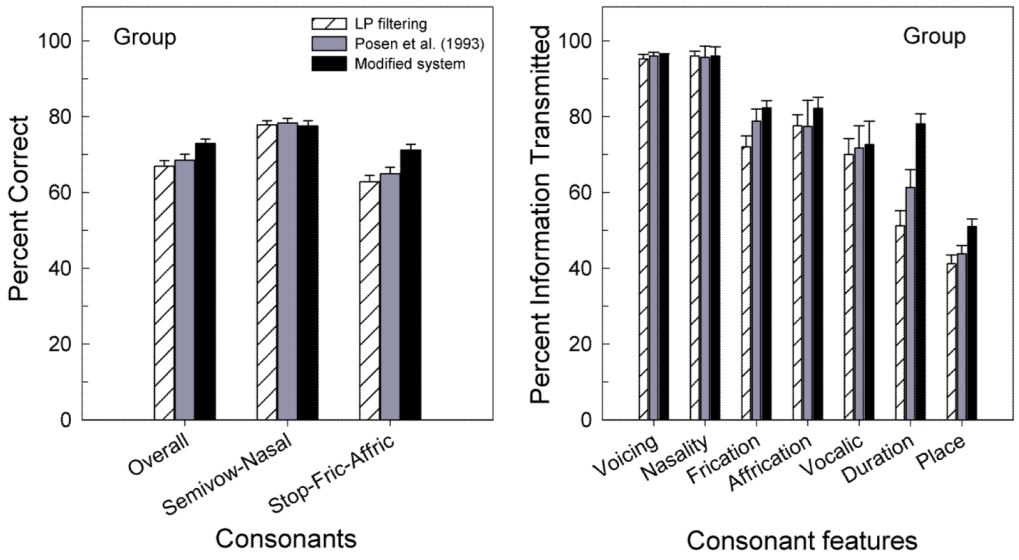

Figure 6.

Average percent correct scores for the overall set of 22 consonants and for two subgroups (semivow-nasal and stop-fric-affric) of consonants (left panel), and percentage of information transmission on each of the seven features (right panel) for the three processing conditions. Error bars represent standard error around the mean.

Table I

Confusion matrix for place-of-articulation classification for 432 fricative consonant tokens in the test set. Overall accuracy is 82% correct.

	Detected as dental	Detected as alveolar	Detected as palatal
Dentals/f, θ, v, ð/	72%	26%	2%
Alveolars/s, z/	7%	83%	10%
Palatals/ʃ, ʒ/	9%	1%	90%